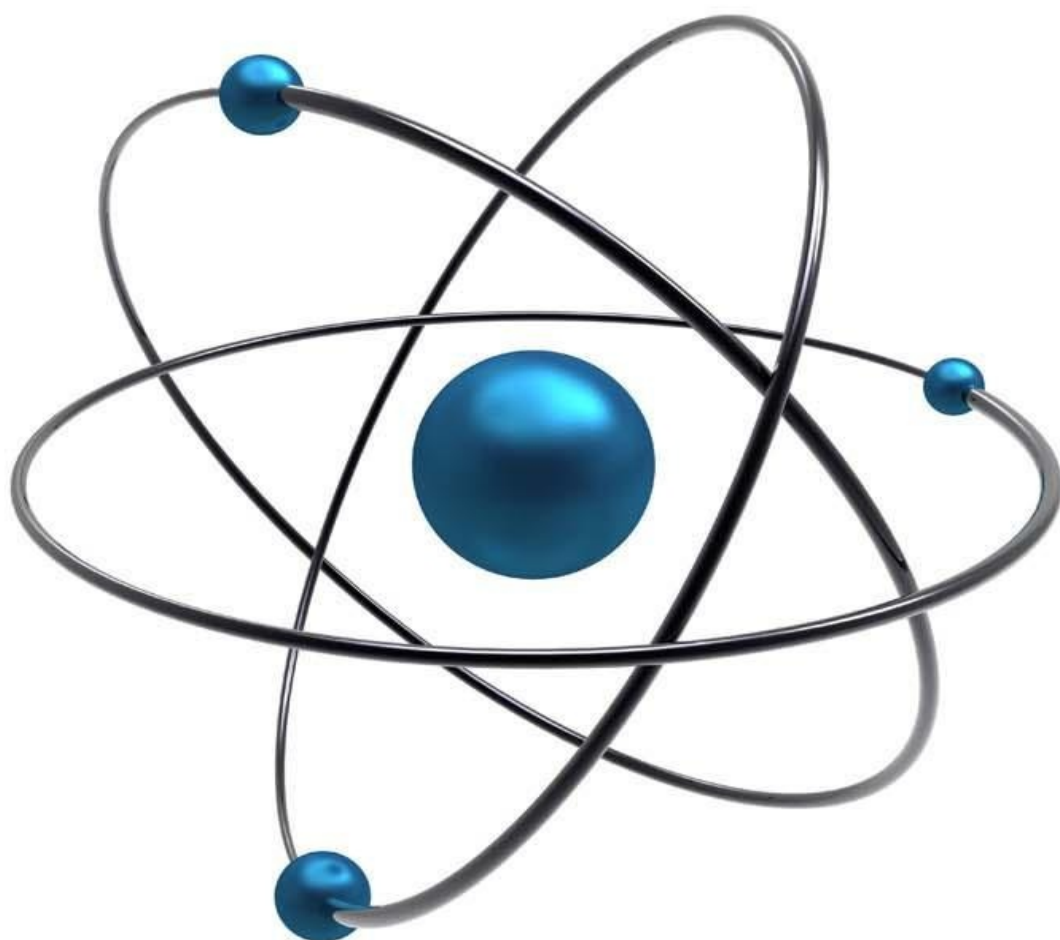


Introducción a la ciencia

Una guía para todos (o casi)



John Gribbin

Lectulandia

Al contrario de quienes han seguido el camino de una especialización que lleva a saber cada vez más de muy poco, aquí el autor nos presenta un muy vasto panorama de los grandes productos de la ciencia. Hoy, la visión científica del mundo tiene dos componentes principales que son muy útiles para su entendimiento: primero, que la ciencia moderna se ha desarrollado en tan sólo cuatrocientos años, el segundo y más importante, que todo esto lo puede comprender una mente humana. De esa forma, aquí están expuestos de manera bastante asequible los resultados de la investigación científica a través de los modelos, especialmente los de la física, que es la ciencia que permite una mejor explicación de las reglas que rigen el universo. Este volumen es, como dice Gribbin, “una guía, no tanto para los aficionados a la ciencia y los conocedores, sino más bien una guía para perplejos, es decir, para quienes son vagamente conscientes de que la ciencia es importante y podría ser incluso interesante, pero a quienes les espantan los detalles técnicos.”

Lectulandia

John Gribbin

Introducción a la ciencia

Una guía para todos (o casi)

ePUB v1.0

Miope 24.08.12

más libros en lectulandia.com

Título original: *Introducción a la ciencia. Una guía para todos (o casi)*
John Gribbin, 1998.

Editor original: Miope (v1.0 a v1.0)
ePub base v2.0

La preocupación por el hombre mismo y por su destino debe ser siempre el interés prioritario en todos los esfuerzos por lograr avances técnicos... con el fin de que todas las creaciones de nuestra mente sean una bendición, y no una maldición, para la humanidad. Nunca olvidéis esto cuando estéis inmersos en vuestros diagramas y ecuaciones.

ALBERT EINSTEIN

Caltech, 1931

Notación científica

Cuando es preciso manejar números muy grandes o muy pequeños, como nos sucederá en este libro, es conveniente utilizar expresiones científicas abreviadas para evitar escribir largas series de ceros. En esta notación científica estándar 10^2 significa 100 (un uno seguido de dos ceros), 10^3 significa 1000, y así sucesivamente. 10^{-1} significa 0'1, 10^{-2} significa 0'01, y así siempre. Esta notación es la más adecuada cuando estamos manejando números tales como el Número de Avogadro (véase el capítulo 1), que se escribe como 6×10^{23} en notación científica, una expresión abreviada para 600.000.000.000.000.000.000.000.

Hay que tener cuidado con el significado de cambios aparentemente pequeños en una potencia de diez determinada. 10^{24} , por ejemplo, es diez veces mayor que 10^{23} , y 10^6 no es la mitad de 10^{12} , sino una millonésima de 10^{12} (10^6 veces 10^{12}).

También seguimos el convenio científico según el cual un billón es mil millones, o 10^9 .

En países europeos no anglosajones, siempre utilizamos un billón para expresar un millón de millones, es decir 10^{12} , pero en este texto se utilizará el valor anglosajón de 10^9 .

Introducción

Si no concuerda con el experimento es falso

El destino al que está abocado cualquier especialista de cualquier área de la ciencia es ceñirse cada vez más estrechamente al tema de su especialidad, aprendiendo cada vez más sobre cada vez menos materia, hasta acabar finalmente sabiéndolo todo sobre nada.

Precisamente con el fin de esquivar este destino, hace muchos años, opté por convertirme en un escritor de temas científicos, en vez de ser un investigador. La oportunidad que esto me dio de plantear a los auténticos científicos preguntas relativas a su trabajo, para luego relatar mis hallazgos en una serie de libros y artículos, me ofreció la posibilidad de saber cada vez menos sobre más y más cosas, aunque por ahora no he alcanzado la fase de saber nada sobre todo. Después de treinta años haciendo esto y de muchos libros centrados en aspectos científicos específicos, me pareció una buena idea escribir un libro general, que ofreciera una amplia visión panorámica de la ciencia, mientras estoy aún en la fase de saber un poco sobre la mayoría de los temas científicos.

Habitualmente, cuando escribo un libro, la audiencia a la que me dirijo soy yo mismo. Es decir, escribo sobre, por ejemplo, la física cuántica o la evolución, precisamente el libro que me gustaría que otro hubiera escrito para mí, de tal forma que me hubiera evitado las molestias de tener que ir averiguando las cosas por mí mismo. Esta vez escribo para todos los demás, esperando que casi todo el mundo encuentre aquí algo que le haga disfrutar. Si usted sabe un poco de física cuántica (o incluso mucho), puede encontrar aquí algo que no sabía sobre la evolución; si usted ya sabe cosas sobre la evolución, puede encontrar algo que le resulte nuevo sobre el Big Bang y otras cuestiones, según los casos.

Así pues, aunque percibo el fantasma de Isaac Asimov que mira por encima de mi hombro (espero que con aprobación) observando este proyecto de tan amplio alcance, este libro no es una «Guía de John Gribbin por el mundo de la ciencia», sino una guía para casi todos los demás. Una guía, no tanto para los aficionados a la ciencia y los conocedores, sino más bien una guía para perplejos, es decir, para quienes son vagamente conscientes de que la ciencia es importante y podría ser incluso interesante, pero a quienes les espantan los detalles técnicos.

No encontrarán tales tecnicismos en este libro (mi coautora los ha suprimido todos, encargándose de mantener a raya todas mis salvajes extravagancias científicas,

para garantizar que lo que queda es inteligible para el profano en la materia). Lo que el lector hallará es una visión personal de la situación en que se encuentra la ciencia a finales del siglo XX y cómo encajan las distintas piezas para configurar una imagen amplia y coherente del universo y de todo lo que éste contiene.

El hecho de que las piezas encajen unas con otras de esta manera es algo que podría pasar inadvertido cuando uno se centra en un aspecto muy concreto de la ciencia, como puede ser el Big Bang o la evolución, pero resulta ser una característica extraordinariamente importante de la ciencia. Tanto la evolución como el Big Bang (y todo lo demás) se basan en los mismos principios, siendo posible seleccionar y tomar aquellos fragmentos del relato científico que uno está dispuesto a aceptar.

A menudo recibo correspondencia o mensajes de personas que, por una razón u otra, no pueden aceptar la teoría especial de la relatividad, la cual nos dice que los relojes se atrasan y las reglas de dibujo se encogen cuando se desplazan. A veces estas personas luchan desesperadamente por encontrar un modo de soslayar esto, aunque acepten todas las demás teorías científicas. Pero no es posible funcionar de esta manera. La teoría especial de la relatividad no es algo aislado, no es una teoría que se refiera exclusivamente a relojes y reglas en movimiento, sino que aparece cuando queremos comprender, por ejemplo, el modo en que la masa se convierte en energía para hacer que el Sol siga brillando, o cómo se comportan los electrones en el interior de los átomos. Si rechazamos los aspectos de la teoría que parecen ir en contra del sentido común, nos quedaremos sin una explicación de por qué el Sol brilla o de qué es la tabla periódica de elementos. Y esto no es más que un ejemplo.

La forma científica de ver el mundo tiene dos importantes características, que están además interrelacionadas, pero que a menudo se pasan por alto y vale la pena mencionarlas. En primer lugar, que todo este asunto se ha desarrollado en tan sólo cuatrocientos años (comenzando en la época de Galileo, que parece ser un momento tan bueno como otro cualquiera para empezar a datar a partir de ahí el inicio de la moderna investigación científica). Y, por otra parte, que todo esto lo puede comprender una mente humana. Es posible que no todos podamos entender todos los aspectos de la forma científica de ver el mundo, pero hay bastantes seres humanos que sí pueden hacerlo, a pesar de que las personas tengan un tiempo de vida tan limitado. Además, aunque haya que ser un genio para dar con una idea como la teoría de la evolución por selección natural, una vez que esta teoría ha quedado formulada, es posible explicársela a personas de una inteligencia mediana, lo que provoca a menudo la respuesta inicial: «Pero, si es obvio. ¡Qué tonto soy! Mira que no ocurrírseme a mí». (Así fue, por ejemplo, la reacción de Thomas Henry Huxley la primera vez que leyó *El origen de las especies*, de Charles Darwin.) Como Albert Einstein dijo en 1936: «El eterno misterio del mundo es su comprensibilidad».

La razón por la que el universo es comprensible para las mentes de los mortales

es que está gobernado por un pequeño conjunto de reglas muy sencillas. Ernest Rutherford, el físico que nos proporcionó el modelo nuclear del átomo a principios del siglo xx, dijo una vez que «la ciencia está dividida en dos categorías: la física y la filatelia». No lo dijo sólo por hacer una gracia, aunque sentía un desdén auténtico por el resto de las ciencias, lo cual hizo sumamente inapropiado el hecho de que le concedieran el Premio Nobel de Química (en 1908, por sus trabajos sobre la radiactividad). La física es la más fundamental de todas las ciencias, en primer lugar porque es la que maneja más directamente las reglas sencillas que gobiernan el universo y las simples partículas de las que está hecho todo lo que hay en él, y en segundo lugar porque los métodos físicos proporcionan el arquetipo utilizado en las otras ciencias para desarrollar sus propios aspectos de la configuración del mundo.

El más importante de dichos métodos es el uso de lo que los físicos denominan modelos. Sin embargo, incluso algunos físicos no llegan a veces a apreciar todo el alcance real de su utilización de los modelos, cosa que es importante precisar antes de que apliquemos esta técnica.

Para un físico, un modelo es una combinación de una imagen mental del aspecto que puede tener un ente fundamental (o no tan fundamental), y de un conjunto de fórmulas matemáticas que describen su comportamiento. Por ejemplo, un modelo del aire que llena la habitación en la que ahora estoy escribiendo estas palabras consideraría cada molécula de gas del aire como una diminuta bola dura. Existen, paralelamente a esto, fórmulas que describen, a un nivel, cómo chocan entre sí estas pequeñas bolas y rebotan una contra otra y también contra las paredes y, a otro nivel, cómo el comportamiento medio de una gran cantidad de ellas produce la presión del aire en la habitación.

El lector no ha de preocuparse por las fórmulas, ya que las ignoraré ampliamente en todo este libro. Sin embargo, deberá recordar que los buenos modelos siempre incluyen las fórmulas correspondientes, ya que se trabaja con ellas para realizar predicciones sobre el modo en que los objetos se comportan —para calcular, por ejemplo, el modo en que la presión del aire de mi habitación cambiaría si la temperatura subiera diez grados Celsius, permaneciendo todo lo demás igual—. La manera de distinguir un buen modelo de otro que es malo consiste en comprobarlo mediante un experimento —en este ejemplo, se calienta la habitación haciendo que su temperatura suba diez grados y se ve si la nueva presión que medimos coincide con la presión que predice el modelo. Si no coincide, el modelo, en el mejor de los casos, necesita alguna modificación y, en el peor caso, se debe descartar en su totalidad.

Richard Feynman, el más grande de los físicos del siglo xx, en una conferencia que pronunció en 1964, dio cuenta detallada de este método científico utilizando la palabra «ley», pero realizando unas observaciones contundentes que se pueden aplicar igualmente a los modelos:

En general, para buscar una nueva ley seguimos el proceso que detallaré a continuación. En primer lugar, hacemos una suposición sobre dicha ley. Luego calculamos las consecuencias de dicha suposición para ver qué implicaría esta ley si lo que hemos supuesto fuera correcto. A continuación, comparamos los resultados del cálculo con lo que se produce en la naturaleza, mediante un experimento o a través de la experiencia, es decir, lo comparamos directamente con lo que se observa, para ver si funciona. Si no concuerda con el experimento, entonces es falso. En esta afirmación tan sencilla está la clave de la ciencia. No importa lo maravilloso que nos parezca aquello que hemos supuesto. Tampoco importa lo ingeniosos que seamos, ni quién realizó la suposición, ni cómo se llama el que la formuló: si no concuerda con el experimento, es falso.

En esto consisten la ciencia y los modelos científicos. *Si no concuerda con el experimento, es falso*. Sin embargo, hay una cuestión más sutil. Incluso *si concuerda* con el experimento, esto no quiere decir que el modelo sea «correcto» en el sentido de ser algo eterno, esto es, la verdad absoluta y última sobre la naturaleza de las cosas que se están estudiando. Aunque las moléculas se puedan tratar como pequeñas bolas duras con el fin de calcular la presión del gas en la habitación, esto no significa que las moléculas *sean* en realidad pequeñas bolas duras. Sólo significa que, en determinadas circunstancias, se comportan *como* si lo fueran. Los modelos funcionan dentro de unos límites generalmente bien definidos, y fuera de dichos límites puede que tengan que ser reemplazados por otros modelos.

Para clarificar todo esto, veamos con otra perspectiva la imagen de las moléculas de gas que están en el aire en mi habitación. Algunas de esas moléculas serán de vapor de agua, y las moléculas de agua, como sabe cualquier niño que va a la escuela, están constituidas por tres átomos —dos de hidrógeno y uno de oxígeno, con la fórmula H_2O . En algunos casos, para determinados fines, un modelo conveniente de la molécula de agua es el que se realiza mediante dos bolas duras más pequeñas (los átomos de hidrógeno) unidas a una única bola dura de mayor tamaño (el átomo de oxígeno), todo ello en forma de V, de tal manera que el oxígeno está en el vértice de la V.

En este caso, los enlaces entre los átomos se pueden considerar como pequeños muelles rígidos, de tal modo que los átomos de la molécula pueden moverse vibrando hacia un lado y hacia el otro. Este tipo de vibración va asociado a una determinada longitud de onda de una radiación, ya que los átomos son portadores de carga eléctrica (esto lo ampliaremos más adelante). Además, si se ven obligados a vibrar de esta manera, emitirán una radiación de microondas y, a la inversa, si se envía a estas moléculas el tipo adecuado de radiación de microondas, las moléculas vibrarán en

simpatía.

Esto es exactamente lo que sucede en un horno microondas. Las microondas sintonizan con las longitudes de onda que hacen vibrar las moléculas de agua, llenan el horno y hacen que vibren las moléculas de agua que contienen los alimentos, de tal manera que absorben energía y calientan dichos alimentos.

Este comportamiento no sólo se ve en la cocina o en el laboratorio. Estudiando la radiación de microondas procedente de las nubes de gas que estaban en el espacio fue como los astrónomos detectaron allí la presencia de moléculas de agua, junto con otras muchas moléculas.

Por lo tanto, para un radioastrónomo que busque moléculas de agua en el espacio, o para un ingeniero eléctrico que diseñe un horno microondas, el modelo de molécula de agua hecha de bolas y varillas es un buen modelo, con tal de que las varillas que unen los átomos tenga un poco de la elasticidad de los muelles. En estos casos ya no se considera la totalidad de la molécula de agua como una sola esfera dura, sino que se consideran los átomos individualmente como si cada uno de ellos (por ejemplo, los átomos de oxígeno) fuera una esfera dura.

Un químico que analizara la composición de una sustancia lo haría, sin embargo, con otra perspectiva. Si se desea saber qué tipos de átomos están presentes en una sustancia, un modo de averiguarlo es estudiar la luz que dichos átomos irradian cuando se calientan. Las distintas clases de átomos irradian colores diferentes, unas líneas muy nítidamente definidas en el espectro del arco iris de la luz; uno de los ejemplos más conocidos es el brillante color naranja amarillento del alumbrado público cuando contiene compuestos de sodio. Son los átomos de sodio (en este caso, excitados por una corriente eléctrica, en vez de por el calor) los que irradian este color especial de la luz.

El modelo utilizado para describir cómo se produce esta luz no considera al átomo como una sola esfera dura, sino como un diminuto núcleo central rodeado por una nube de diminutas partículas con carga eléctrica, llamadas electrones. El núcleo central tiene carga eléctrica positiva, mientras que cada uno de los electrones tiene carga eléctrica negativa, por lo que el átomo en conjunto tiene carga eléctrica cero. Las líneas brillantes del espectro que van asociadas a un tipo particular de átomo se explican considerando el modo en que los electrones se mueven en la parte externa del átomo. Lo que hace que se distinga un tipo de átomo de otro es, hablando en términos químicos, el número de electrones (8 para el oxígeno, sólo 1 para el hidrógeno y 11 para el sodio). Debido a que cada tipo de átomo tiene su propia y única disposición de los electrones, cada tipo de átomo produce un único patrón de líneas de color en el espectro.

Podría continuar, pero la cuestión ya está clara. El modelo que considera las moléculas de aire como pequeñas bolas duras es un buen modelo, porque funciona

cuando se utiliza para calcular las variaciones que experimenta la presión cuando cambia la temperatura. El modelo que representa las moléculas como si estuvieran formadas por esferas rígidas de menor tamaño (átomos) que se mantienen unidas como las uvas en los racimos es también un buen modelo, porque funciona cuando se utiliza para calcular el modo en que las moléculas, cuando vibran, emiten ondas de radio. Y el modelo que representa a los átomos, no como esferas duras indivisibles, sino como diminutos núcleos rodeados de nubes de electrones también es bueno, porque funciona cuando se utiliza para predecir el color de la luz asociada a una clase concreta de átomo.

Ninguno de los modelos constituye la verdad absoluta y última, pero todos representan un papel en esta función teatral. Son herramientas que usamos para ayudar a nuestra mente a crear una imagen de lo que está sucediendo y a calcular cosas que podemos comprobar directamente realizando mediciones, como es el caso de la presión del aire en una habitación o el color de la luz que irradia una sustancia caliente.

Del mismo modo que un carpintero no utilizaría un escoplo para realizar el mismo trabajo que haría con un mazo, también un científico debe elegir el modelo adecuado para el trabajo que tiene entre manos. Cuando Feynman dice: «Si no concuerda con el experimento, es falso», lo que quiere decir es: «Si no concuerda con el experimento adecuado». El modelo de una molécula de vapor de agua como una única esfera dura no permite la posibilidad del tipo de vibración asociado con las microondas, es decir, predice que el vapor de agua no emitirá microondas. Esto significa que es un modelo erróneo para utilizarlo cuando lo que nos interese sean las microondas. Pero no significa que el modelo sea erróneo para utilizarlo si lo que nos interesa es cómo se ve afectada la presión del aire en una habitación cuando la temperatura aumenta.

En la ciencia todo consiste en modelos y predicciones, en hallar la manera de conseguir crear dentro de nuestras mentes una imagen de cómo funciona el universo y en encontrar el modo de efectuar cálculos que predigan lo que sucederá en determinadas circunstancias. Cuanto más nos alejemos del mundo ordinario de la vida cotidiana, ya sea hacia una escala muy pequeña, o hacia una escala muy grande, más tendremos que confiar en las analogías: un átomo es, en determinadas circunstancias, «como» una bola de billar; un agujero negro es, en cierto sentido, «como» una abolladura en un trampolín.

Resultaría tedioso seguir de esta manera precisando las calificaciones relativas al uso de los distintos modelos, por lo que, ahora que ya me he desahogado, me abstendré de hacerlo y confiaré en que el lector recuerde la precisión que he realizado en el sentido de que el mejor modelo sólo es bueno en su propio contexto, y que los escoplos nunca se deben usar para hacer el trabajo que tienen que hacer los mazos.

Siempre que nos refiramos a algo calificándolo de «real», lo que queremos decir es que es el mejor modelo que se puede utilizar en las circunstancias del caso.

Con esta condición, partiendo de la escala de los átomos, invitaré al lector a que descienda al mundo de lo muy, muy pequeño y, más adelante, a que salga al universo a gran escala, ofreciéndole la mejor explicación moderna (el mejor modelo) de la naturaleza de las cosas dentro de cada escala. Todas serán ciertas, en el sentido de que concuerdan con los experimentos; todas van a encajar unas con otras, como las piezas de un rompecabezas, para dar una imagen coherente de cómo funciona el universo y todo lo que contiene; y todo lo podrá comprender, al menos en líneas generales, cualquier mente humana con una capacidad mediana. Existe otra característica de la ciencia, un punto de vista del que soy firme partidario y que ha configurado la estructura de este libro (y toda mi carrera), pero que no todos los científicos comparten necesariamente. Para mí, la ciencia es primordialmente la investigación sobre nuestro lugar en el universo, el lugar que ocupan los seres humanos en un mundo que se extiende desde las más diminutas partículas subatómicas hasta las extensiones más largas en el espacio y el tiempo. No existimos de manera aislada, por lo que la ciencia es una actividad humana cultural, no un mero esfuerzo desapasionado por alcanzar la verdad, al margen de lo duramente que nos esforcemos. Aquello de lo que se trata es siempre de dónde venimos y a dónde vamos. Por eso es la historia más emocionante que se ha relatado jamás.

Átomos y elementos

En 1962, en una serie de conferencias para estudiantes universitarios impartidas en Caltech, Richard Feynman situó el modelo del átomo como centro de la explicación científica del mundo. Dijo lo siguiente:

Si por un cataclismo resultaran destruidos todos los conocimientos científicos y sólo una frase pudiera pasar a las generaciones siguientes, ¿qué sentencia contendría el máximo de información en el mínimo de palabras? Yo creo que es la *hipótesis atómica* (o el *hecho atómico*, o como quieran llamarlo), según la cual *todas las cosas están hechas de átomos —pequeñas partículas que se encuentran en perpetuo movimiento y se atraen entre sí cuando se sitúan a una corta distancia, pero que se repelen si se intenta introducir a la una en la otra—*. Sólo con que se utilice un poco de imaginación y de reflexión, en esta única frase, como verán, está contenida una enorme cantidad de información sobre el mundo.

La cursiva es de Feynman y se puede encontrar el texto completo en su libro *Six Easy Pieces* (Seis piezas fáciles). Atendiendo a esta idea de Feynman, comenzamos nuestra guía de la ciencia hablando de los átomos. A menudo se señala que la idea de los átomos como últimas piezas indivisibles de las que está compuesta la materia se remonta a los tiempos de los antiguos griegos, cuando, en el siglo v a.C., Leucipo de Mileto y su alumno Demócrito de Abdera hablaban de estos entes fundamentales. De hecho, aunque Demócrito dio a los átomos este nombre (que significa «indivisible»), el asunto tiene algo de señuelo para despistar. Sus contemporáneos no se tomaron en serio esta idea, ni tampoco nadie durante más de dos mil años. El auténtico desarrollo del modelo del átomo data de finales del siglo XVIII, cuando los químicos comenzaron a trabajar en la moderna investigación de las propiedades de los elementos.

El concepto de elementos —sustancias fundamentales de las que está hecha toda la complejidad del mundo cotidiano— se remonta también a los primeros filósofos griegos, a los que se les ocurrió la idea de que todo está hecho de distintas mezclas de cuatro elementos: aire, tierra, fuego y agua. Aparte del nombre «elemento» y de la idea de que un elemento no se puede descomponer en formas químicas más sencillas, no queda nada del concepto griego de elemento en la química moderna, que se ha construido a partir de la obra de Robert Boyle, desarrollada a mediados del siglo XVII.

Boyle fue la primera persona que enunció la definición de elemento como una sustancia que se podía combinar con otros elementos para formar compuestos, pero tal que ella misma no se podía descomponer en ninguna sustancia más simple. El agua, por ejemplo, es un compuesto que se puede descomponer químicamente en las

partes que lo componen, es decir, oxígeno e hidrógeno. Pero el oxígeno y el hidrógeno son elementos, porque no se pueden descomponer a su vez por medios químicos. No están hechos de otros elementos. El número de elementos conocidos fue aumentando a medida que los químicos desarrollaban nuevas técnicas para descomponer los compuestos; pero en el siglo XIX ya iba quedando claro qué sustancias eran realmente indivisibles.

El progreso definitivo en la comprensión del modo en que los elementos se combinan para dar compuestos llegó cuando John Dalton retomó el concepto de átomo a principios del siglo XIX. Basó su modelo en el descubrimiento de que, dado un compuesto concreto, independientemente de cómo haya sido preparado, la proporción entre los pesos de los diferentes elementos presentes en él es siempre la misma. Por ejemplo, en el agua la proporción de oxígeno respecto al hidrógeno es siempre 8:1 en cuanto al peso; en el carbonato cálcico (cal común) la proporción de calcio, carbono y oxígeno es siempre 10:3:12 en cuanto al peso.

La explicación de Dalton consistió en decir que cada clase de elemento está formado por una clase de átomos idénticos y que es la naturaleza de estos átomos la que determina las propiedades del elemento. Según esta idea, la característica diferenciadora clave para que sea posible decir que un elemento es distinto de otro es su peso. Cuando se combinan dos o más elementos, son realmente los átomos de los diferentes elementos los que se unen para formar lo que ahora se conoce como moléculas. Cada molécula de un compuesto contiene el mismo número de átomos que cualquier otra molécula del compuesto, teniendo cada una el mismo número de átomos de cada uno de los elementos que intervienen en cada molécula. Una molécula de agua está formada por dos átomos de hidrógeno y uno de oxígeno (H_2O); una molécula de carbonato cálcico está formada por un átomo de calcio, un átomo de carbono y tres átomos de oxígeno ($CaCO_3$). Además, actualmente sabemos que en el caso de algunos elementos los átomos pueden unirse entre sí para formar moléculas, sin que intervengan en esto otros elementos. El oxígeno que hay en el aire que respiramos, por ejemplo, está constituido por moléculas diatómicas, O_2 , las cuales no se consideran compuestos.

El modelo atómico de Dalton fue un éxito enorme dentro de la química, aunque a lo largo del siglo XIX algunos científicos lo consideraron sólo como una especie de truco práctico, un modo de calcular de qué manera se comportaban los elementos en las reacciones químicas, pero no una prueba de que los átomos existieran «realmente». Al mismo tiempo, otros científicos fueron encontrando pruebas cada vez más concluyentes de que los átomos se podían considerar como entidades reales, como pequeñas bolas duras que se atraían las unas a las otras cuando estaban separadas por una distancia, pero se repelían mutuamente al intentar ponerlas juntas.

Una nueva línea de ataque partió del trabajo de Amadeo Avogadro (que fue, por

cierto, quien mostró que la combinación de átomos existente en una molécula de agua es H_2O , no HO). En 1811 Avogadro publicó un trabajo en el que planteaba que volúmenes iguales de gas a la misma temperatura y presión contienen igual número de átomos. Esto fue antes de que se desarrollara el concepto de molécula, por lo que actualmente afirmaríamos que volúmenes iguales de gas a la misma temperatura y presión contienen igual número de moléculas. De cualquier modo, sin embargo, lo importante es que el modelo de Avogadro contemplaba un número igual de pequeñas esferas duras moviéndose y chocando unas con otras en un recipiente con gas de un determinado tamaño y en las condiciones mencionadas, tanto si este gas era oxígeno, como si era dióxido de carbono, o cualquier otro.

La idea que subyace detrás de todo esto es que en un recipiente de gas lo que hay es sobre todo espacio vacío con unas pequeñas bolas duras zumbando por todo el recinto dentro del recipiente, chocando unas con otras y con las paredes de dicho recipiente. No importa de qué estén hechas las pequeñas bolas; por lo que respecta a la presión ejercida sobre las paredes del recipiente, lo único que importa es la velocidad de las partículas y con qué frecuencia lo golpean. La velocidad está en función de la temperatura (a temperaturas más elevadas corresponden movimientos más rápidos) y el número de choques por segundo depende de cuántas bolitas duras haya en el recipiente. Así, a la misma temperatura, la misma presión y el mismo volumen, el número de partículas debe ser el mismo.

Este tipo de modelo también explica la diferencia entre gases, líquidos y sólidos. En un gas, como ya hemos dicho, hay sobre todo espacio vacío y las moléculas se precipitan a través de ese espacio, chocando unas con otras. En un líquido no hay espacio vacío y se puede considerar que las moléculas se tocan entre sí, pero manteniendo un movimiento constante, deslizándose una junto a otra en una masa amorfa. En un sólido el movimiento casi ha cesado y las moléculas están sujetas a un lugar determinado, salvo un ligero meneo, una especie de correr sin avanzar que practican como deporte las moléculas.

La idea de Avogadro no fue tomada muy en serio en su época (ni siquiera por Dalton). Sin embargo, a finales de la década de 1850 fue retomada por Stanislao Cannizzaro, que se dio cuenta de que esta idea proporcionaba un modo de medir los pesos atómicos y moleculares. Si podemos hallar el número de moléculas que hay en un cierto volumen de un gas determinado, a una temperatura y una presión dadas (las condiciones habituales suelen ser cero grados Celsius y una presión atmosférica estándar), sabremos el número de moléculas que están presentes en cualquier gas en tales condiciones estándar. Con el fin de hallar cuánto pesa cada molécula, basta con pesar el gas y dividir este dato entre el número de moléculas.

En estas condiciones estándar, podemos elegir un volumen de gas correspondiente a dos gramos de hidrógeno (dos gramos, no uno, porque cada

molécula de hidrógeno, H_2 , contiene dos átomos). Esta cantidad resulta ser poco más de trece litros de gas. El número de moléculas en un volumen como éste se llama número de Avogadro. El mismo volumen de oxígeno en las mismas condiciones pesa treinta y dos gramos y la evidencia química nos dice que hay dos átomos de oxígeno en cada molécula. Sin embargo, contiene el mismo número de moléculas que dos gramos de hidrógeno. Por lo tanto, sabemos inmediatamente que un átomo de oxígeno pesa dieciséis veces lo que pesa un átomo de hidrógeno. Ésta era una manera muy útil de calcular pesos atómicos y moleculares relativos; sin embargo, para la averiguación de los pesos reales se necesitaba conocer el número de Avogadro, y éste era más difícil de calcular.

Hay varias formas diferentes de abordar este problema, pero sólo con poner un ejemplo nos podemos hacer una idea del modo en que tendría que hacerse. Dicho ejemplo es una variación sobre el tema utilizada por Johann Loschmidt a mediados de la década de 1860. Recuérdese que en un gas hay mucho espacio vacío entre las moléculas, pero en un líquido están tocándose unas con otras. Loschmidt pudo calcular la presión de un gas en un recipiente (en condiciones estándar) a partir del número de Avogadro, y esta presión determina la distancia media que recorren las moléculas entre las colisiones (el llamado «recorrido libre medio»), y la fracción del volumen del gas que está ocupada realmente por las propias moléculas. Además, consiguió calcular cuánto espacio vacío había en el gas licuándolo y midiendo qué cantidad de líquido resultaba; o también, desde luego, utilizando las mediciones de la densidad del oxígeno líquido y la del nitrógeno líquido que otros científicos habían realizado. Dado que las partículas que se hallan en un líquido están tocándose unas a otras, restando el volumen del líquido del volumen del gas pudo averiguar cuánto espacio vacío había en el gas. Así, ajustando el valor del número de Avogadro en sus cálculos de la presión para que correspondiera a la presión medida, pudo averiguar la cantidad de moléculas que estaban presentes.

Dado que las densidades del nitrógeno líquido y el oxígeno líquido utilizados en estos cálculos no eran tan exactas como lo son en las mediciones modernas, las cifras que dio Loschmidt en 1866 para el número de Avogadro, $0'5 \times 10^{23}$, resultaban un poco inferiores para lo que en realidad debían ser. Utilizando una técnica diferente, Albert Einstein en 1911 calculó un valor de $6'6 \times 10^{23}$. El moderno valor óptimo de este número es $6'022045 \times 10^{23}$ o, en términos coloquiales, algo más de seiscientos mil trillones. Éste es el número de átomos que hay en un gramo de hidrógeno, dieciséis gramos de oxígeno o, en general, en el equivalente en gramos del peso atómico de cualquier elemento.

Así, cada átomo de hidrógeno pesa $0'17 \times 10^{-23}$ gramos. Cada molécula de aire tiene un diámetro de unas pocas cienmillonésimas de centímetro. A $0^\circ C$ de temperatura y una atmósfera de presión, un centímetro cúbico de aire contiene

4.5×10^9 moléculas; el recorrido libre medio de una molécula de aire es igual a trece millonésimas de metro, y una molécula de oxígeno del aire a la temperatura anteriormente indicada viaja a algo más de 461 metros por segundo (aproximadamente 17.000 km por hora). De este modo, cada molécula participa en más de 3.500 millones de colisiones cada segundo, produciendo en promedio una sensación de presión uniforme en nuestra piel o en las paredes de la habitación.

De hecho, en 1738 Daniel Bernoulli planteó por primera vez la teoría cinética de los gases. Se inspiró en el trabajo de Robert Boyle, realizado a mediados del siglo XVII. Boyle había descubierto que cuando se comprime un gas (por ejemplo, mediante un émbolo) el volumen del gas varía en proporción inversa a la presión: al duplicar la presión, el volumen se reduce a la mitad. Bernoulli explicaba esto en términos de teoría cinética, observando también que la relación entre la temperatura de un gas y su presión (cuando se calienta un gas, su presión aumenta, permaneciendo iguales otros factores) se podía explicar también en términos de energía cinética (energía del movimiento) de las pequeñas partículas contenidas en el gas: el calentamiento del gas hace que las partículas se muevan más rápido, por lo que producen impactos más fuertes en las paredes del recipiente. Pero Bernoulli se adelantó a su tiempo. En aquella época, la mayoría de la gente que reflexionaba sobre el calor pensaba que éste estaba relacionado con la presencia de una especie de fluido, llamado calórico, que se trasladaba de una sustancia a otra. La versión de Bernoulli de la teoría cinética no tuvo ninguna resonancia en el mundo científico de la época.

La teoría cinética se descubrió dos veces (primero fue John Herapath en 1820, luego la redescubrió John Waterston en 1845) y las dos veces fue ignorada, hasta que, finalmente, la mayoría de los científicos la aceptaron en la década de 1850, en gran medida debido al trabajo de James Joule. En la década de 1860 surgió una versión matemática completa de la teoría cinética (un modelo completo), gracias sobre todo al trabajo de Rudolf Clausius, James Clerk Maxwell y Ludwig Boltzmann. Este modelo se conoció con el nombre de estadística mecánica, porque trata del comportamiento medio estadístico de un número muy grande de partículas que participan en interacciones mutuas como si fueran diminutas bolas de billar, saltando por todas partes según las leyes de la mecánica de Newton.

Éste es un ejemplo impresionante del modo en que las leyes de la física se pueden aplicar en circunstancias bastante diferentes de aquellas en las que se investigaron cuando fueron descubiertas y subraya la importante diferencia que existe entre una ley y un modelo. Una ley, como la ley de la gravedad de Newton, es en realidad una verdad universal. Newton descubrió que todo objeto del universo atrae a cualquier otro objeto del universo con una fuerza que es proporcional a uno partido por el cuadrado de la distancia que separa a los dos objetos. Ésta es la famosa «ley de los cuadrados inversos». Se puede aplicar, como afirmó Newton, a una manzana que cae

de un árbol y a la Luna en su órbita, siendo en ambos casos la gravedad de la Tierra la que ejerce la atracción. Se aplica también a la fuerza que mantiene a la Tierra en su órbita alrededor del Sol y a la fuerza que está haciendo que la actual expansión del universo sea gradualmente cada vez más lenta. Sin embargo, aunque esta ley es una verdad absoluta, el propio Newton no tenía ni idea de cuál era su causa: no tenía un modelo de la gravedad.

En efecto, Newton dijo específicamente en este contexto «hypotheses non fingo» («No hago hipótesis») y no intentó explicar por qué la gravedad obedecía a una ley de cuadrados inversos. Por contra, la teoría general de la relatividad de Einstein ofrece un modelo que produce automáticamente una ley de cuadrados inversos para la gravedad. En vez de echar abajo las teorías de Newton sobre la gravedad, como sugieren algunas historias famosas, la teoría general realmente las refuerza, aportando un modelo que explica la ley de la gravedad (también va más allá de las teorías de Newton para describir el comportamiento de la gravedad en condiciones extremas; hablaremos sobre esto más tarde).

Con el fin de ser un buen modelo, cualquier modelo de la gravedad debe, por supuesto, «predecir» una ley de los cuadrados inversos, pero esto no significa que dicho modelo sea necesariamente la última palabra, por eso actualmente los físicos esperan confiados que algún día se desarrollará una teoría cuántica de la gravedad que irá aún más lejos que la teoría de Einstein. Sin embargo, aunque no sepamos si esto se llegará a hacer realmente, ni cuándo sucederá, sí que podemos estar seguros de una cosa: ese nuevo modelo también predecirá una ley de cuadrados inversos. Después de todo, sean cuales sean las nuevas teorías y los nuevos modelos que encuentren los físicos, las órbitas de los planetas alrededor del Sol seguirán siendo las mismas y las manzanas no empezarán a saltar hacia arriba desde los árboles.

La gravedad, en realidad, es una fuerza muy débil, salvo que entre en juego una gran cantidad de materia. Es necesaria la gravedad de toda la Tierra para que tire de una manzana, la arranque del árbol y la haga caer al suelo. Sin embargo, un niño de dos años puede coger la manzana del suelo, venciendo así la atracción gravitatoria. En cuanto a los átomos y las moléculas que están golpeteando en un recipiente de gas, las fuerzas de atracción gravitatoria entre las partículas es tan leve que se puede ignorar completamente. Lo que importa aquí, como observaron los que desarrollaron la mecánica estadística en el siglo XIX, son las otras leyes que descubrió Newton: las leyes de la mecánica.

Existen sólo tres leyes de la mecánica de Newton, y son tan conocidas actualmente que parecen obvias y de sentido común, pero constituyen los fundamentos de toda la física. La primera ley afirma que cualquier objeto se mantiene en reposo o se mueve a velocidad constante en línea recta a menos que sea empujado o atraído por una fuerza. Esto no es simplemente cuestión de sentido

común, porque si ponemos un objeto en movimiento aquí en la Tierra (por ejemplo, si le damos una patada a un balón) pronto cesa de moverse, debido a la fricción. Su propia perspicacia le hizo considerar a Newton cómo se comportan los objetos cuando no hay fricción, como cuando unas rocas se desplazan por el espacio o, incluso, cuando unos átomos están zumbando en un recipiente de gas (sin embargo, dicho sea de paso, aunque nunca desarrolló una teoría cinética de los gases, Newton era partidario del modelo del átomo y escribió sobre la materia que está formada por «primitivas partículas... incomparablemente más duras que cualquier cuerpo poroso constituido por ellas; incluso tan duras que nunca se desgastan ni se rompen en pedazos»).

La segunda ley de Newton dice que, cuando se aplica una fuerza a un cuerpo, éste se acelera y sigue acelerándose durante todo el tiempo que esté actuando sobre él la fuerza (aceleración significa un cambio en la velocidad de un cuerpo, o en la dirección en que se mueve, o en ambas cosas; así, la Luna se acelera en su recorrido alrededor de la Tierra, porque su dirección está cambiando constantemente, aunque su velocidad permanezca prácticamente invariable). La aceleración producida por una fuerza depende del valor de esta fuerza dividido entre la masa del cuerpo (dándole la vuelta, los físicos dicen a menudo que la fuerza es igual a la masa por la aceleración). Esto sí que concuerda con lo que dice el sentido común, ya que cuesta más esfuerzo empujar los objetos si tienen más masa. Finalmente, la tercera ley de Newton señala que cuando un objeto ejerce una fuerza sobre otro, hay una reacción igual y opuesta en el primer objeto. Cuando le doy una patada a un balón (o a una piedra, si estoy tan loco como para hacerlo), la fuerza que mi pie ejerce sobre el balón (o la piedra) hace que éste se desplace, pero la fuerza igual y opuesta que el objeto ejerce sobre mi pie se puede sentir claramente.

De un modo más sutil, cuando la Tierra tira de la Luna mediante la gravedad, existe una fuerza igual y opuesta que tira de la Tierra. Más que pensar que la Luna describe una órbita alrededor de la Tierra, deberíamos decir realmente que ambas recorren órbitas en torno a su centro de gravedad mutuo, pero como la Tierra tiene una masa mucho mayor que la de la Luna, lo que en realidad sucede es que este punto de equilibrio se encuentra bajo la superficie de la Tierra. También se puede decir, hablando en sentido estricto, que la igualdad entre la acción y la reacción significa que, cuando la manzana cae al suelo, toda la Tierra atraída por la manzana se desplaza una distancia infinitesimal «hacia» dicha manzana. La tercera ley de Newton es la que explica el retroceso de una escopeta cuando se dispara y el modo en que funciona un cohete, expulsando materia en un sentido y retrocediendo en sentido contrario.

Estas tres leyes se refieren al universo a gran escala, que es donde Newton las aplicó para explicar las órbitas de los planetas, y al entorno cotidiano, donde se

pueden investigar haciendo experimentos tales como hacer rodar bolas por planos inclinados y medir su velocidad, o rebotar dos pelotas la una contra la otra. Sin embargo, debido a que, desde luego, son leyes universales, también se cumplen a escala de los átomos y las moléculas, proporcionando así la mecánica que, como ya he mencionado, es la base de la mecánica estadística y de la moderna teoría cinética de los gases; aunque la teoría cinética moderna se desarrolló casi dos siglos después de que Newton propusiera sus tres leyes de la mecánica y él nunca aplicó sus leyes en ese sentido. En realidad, Newton no inventó ninguna de esas leyes, sino que son leyes del universo y actuaban de la misma manera antes de que él las pusiera por escrito, del mismo modo que actúan en situaciones que Newton nunca llegó a investigar.

Existen razones de peso que explican por qué la teoría cinética y la mecánica estadística fueron aceptadas finalmente por los científicos a mediados del siglo XIX. En aquella época, el terreno ya había sido preparado por el estudio de la termodinámica (literalmente, calor y movimiento), que tuvo una importancia práctica inmensa en los días en que la potencia del vapor estaba siendo impulsora de la revolución industrial en Europa.

Los principios de la termodinámica se pueden resumir también en tres leyes, que tienen gran importancia aplicable en amplios sectores por toda la ciencia (en realidad, aplicable a todo el universo) y no sólo al diseño y a la construcción de máquinas de vapor. La primera ley de la termodinámica se conoce también como ley de la conservación de la energía y dice que la energía total de un sistema cerrado permanece constante. El Sol no es un sistema cerrado, ya que está emitiendo energía al espacio; la Tierra tampoco es un sistema cerrado, ya que recibe energía del Sol. Sin embargo, en un proceso tal como una reacción química que tiene lugar en un tubo de ensayo aislado, o en los procesos correspondientes a la mecánica estadística donde pequeñas partículas duras saltan por todas partes dentro de un recipiente, la cantidad total de energía es fija. Si una partícula de movimiento rápido que lleva mucha energía cinética choca con una partícula de movimiento lento que tiene poca energía cinética, la primera partícula probablemente perderá energía y la segunda recibirá energía. No obstante, el total de energía transportada entre las dos partículas antes y después del choque permanecerá constante.

Desde que Einstein propuso la teoría especial de la relatividad a principios del siglo XX, sabemos que la masa es una forma de energía y que, en las circunstancias adecuadas (como en el interior de una central nuclear, o en el núcleo del Sol) la energía y la masa se pueden intercambiar. Por ello, actualmente, la primera ley de la termodinámica se llama ley de la conservación de la masa y la energía, y no sólo ley de la conservación de la energía.

Se puede decir que la segunda ley de la termodinámica es la ley más importante de toda la ciencia. Se trata de la ley que dice que las cosas se desgastan. En términos

del calor —la forma en que la ley se descubrió en los días de las máquinas de vapor — la segunda ley dice que éste no fluye de un lugar más frío a otro más caliente por su propia voluntad. Si ponemos un cubito de hielo en una taza de té caliente, el hielo se derrite mientras el té se refresca; nunca se ha visto una taza de té templado en la que el té se caliente mientras se forma un cubito de hielo en el centro de la taza, aunque tal proceso no violaría la ley de la conservación de la energía. Otra manifestación de esta segunda ley es el modo en que un muro de ladrillos, sin que nadie lo toque, envejecerá y se desmoronará, mientras que los ladrillos que están en un montón, sin que nadie los toque, nunca se unirán para formar un muro.

En la década de 1920 el astrofísico Arthur Eddington resumió la importancia de la segunda ley en su libro *The Nature of the Physical World* («La naturaleza del mundo físico»):

La segunda ley de la termodinámica ocupa, en mi opinión, la posición más importante entre las leyes de la naturaleza. Si alguien le dice a usted que su teoría favorita sobre el universo está en desacuerdo con las ecuaciones de Maxwell, pues peor para las ecuaciones de Maxwell. Si le dicen que entra en contradicción con lo observado, bueno, esos experimentalistas hacen chapuzas a veces. Pero si resulta que su teoría va en contra de la segunda ley de la termodinámica, no le queda a usted ninguna esperanza; su teoría se hundirá sin remedio en la más profunda de las humillaciones.

La segunda ley también está relacionada con el concepto denominado entropía, que mide la cantidad de desorden que hay en el universo, o en una parte cerrada del universo (como un tubo de ensayo que ha sido sellado en el laboratorio). La entropía en un sistema cerrado no puede disminuir, por lo que cualquier cambio en el sistema lo llevará a un estado de mayor entropía. El «sistema» de un cubito de hielo que flota en una taza de té tiene más orden (menos entropía) que una taza de té templado, y es por esto por lo que el sistema se mueve desde un estado de orden a un estado desordenado.

El universo considerado en su totalidad es un sistema cerrado, por lo que la entropía del universo entero debe estar aumentando. Pero la Tierra, como ya he indicado, no es un sistema cerrado, ya que recibe un aporte continuo de energía procedente del Sol. Es este suministro de energía desde el exterior lo que hace posible que creamos orden localmente a partir del desorden (por ejemplo, construyendo casas a partir de montones de ladrillos); la disminución de la entropía que va asociada con todos los procesos de la vida en la Tierra se ve más que compensada por el aumento de la entropía dentro del Sol como resultado de los procesos que producen la energía de la que nos alimentamos.

Si tiene dudas, piense que el mismo tipo de fenómeno se produce, a menor escala, cuando enfriamos el interior de un frigorífico y hacemos cubitos de hielo. Tenemos que utilizar energía para extraer calor del frigorífico, y este proceso aumenta la entropía del universo, con un incremento que es mayor que la disminución en entropía que se produce como consecuencia dentro del frigorífico. Si se coloca un frigorífico en una habitación cerrada, con la puerta del frigorífico abierta y el motor funcionando, la habitación se calentará, no se enfriará, porque la energía que consume el motor al calentarse será mayor que el efecto refrigerante del frigorífico abierto.

La tercera ley de la termodinámica tiene que ver con el concepto cotidiano y familiar de temperatura, algo que hemos dado por sabido en esta discusión hasta ahora. Aunque sólo he hablado de la relación entre calor y entropía en términos generales, existe, de hecho, una relación matemática entre las dos magnitudes, y ésta muestra que cuando los cuerpos se enfrían es cada vez más difícil obtener energía de ellos. Esto es bastante obvio en un contexto cotidiano; así, la auténtica razón por la cual las máquinas de vapor fueron tan importantes en la revolución industrial fue porque el vapor caliente podía realizar trabajo útil, subiendo y bajando los pistones, o haciendo girar las ruedas una y otra vez. Se podría, si se deseara realmente, hacer un tipo de motor que hiciera subir y bajar los pistones utilizando un gas mucho más frío (quizá dióxido de carbono), pero esto no sería muy efectivo.

En la década de 1840, William Thomson (quien después sería lord Kelvin) desarrolló estas teorías de la termodinámica en una escala absoluta de temperaturas, siendo el cero de esta escala la temperatura a la cual no se podía obtener más calor (o más energía) de un cuerpo. Este cero absoluto está determinado por las leyes de la termodinámica (por lo que Thomson pudo calcularlo matemáticamente, aunque nunca consiguió enfriar nada hasta esta temperatura) y se sitúa en -273°C . Se conoce como O K, en honor de Kelvin. En la escala Kelvin de temperaturas las unidades tienen el mismo valor que en la escala Celsius, por lo que el hielo se derrite a 273 K (no se escribe el «símbolo de grado» delante de la K). La tercera ley de la termodinámica dice que es imposible enfriar algo hasta los O K, aunque si nos esforzamos podemos (en principio) llegar tan cerca del cero absoluto como queramos. Un cuerpo a cero grados Kelvin estaría en el estado más bajo de energía que un cuerpo puede alcanzar y no se podría obtener de él nada de energía para realizar ningún trabajo.

Algún bromista resumió las tres leyes de la termodinámica en términos coloquiales de la siguiente manera:

1. No se puede ganar.
2. No se puede ni siquiera empatar.
3. No se puede abandonar el juego.

El éxito de la teoría cinética y de la mecánica estadística contribuyó a convencer a muchos físicos de que los átomos eran algo real. Sin embargo, justo hasta finales del siglo XIX muchos químicos siguieron considerando sospechoso el concepto de átomo. A nosotros esto nos puede parecer extraño, porque hacia el final de la década de 1860 (justo en la época en que la teoría cinética estaba produciendo tantos éxitos) se había descubierto una pauta en las propiedades de los distintos elementos, una pauta que ahora se explica totalmente en términos de las propiedades de los átomos.

En la década de 1820, Jöns Berzelius, aunque nunca aceptó la hipótesis de Avogadro, hizo un primer intento de clasificar los elementos según el orden de sus pesos atómicos (estableciendo como unidad el peso del hidrógeno, el más ligero de los elementos), pero esto no tuvo éxito. El auténtico progreso llegó después de 1860, cuando Cannizzaro retomó la idea de Avogadro y convenció a muchos de sus colegas de que funcionaba y de que el peso atómico era un concepto útil dentro de la química. Pero, incluso entonces, tuvo que pasar tiempo para que se comprendiera toda la importancia de este progreso. El gran descubrimiento fue que si los elementos se ordenaban según su peso atómico, aquellos que tenían propiedades químicas similares se encontraban a intervalos regulares en la lista; de este modo el elemento que tiene peso atómico 8 presenta propiedades similares a las del elemento cuyo peso atómico es 16 (y también con el elemento de peso atómico 24), y el elemento que tiene peso atómico 17 posee propiedades similares a aquel cuyo peso es 25, y así sucesivamente.

No hacía falta tener mucha imaginación para ir de este descubrimiento a la idea de escribir una lista de elementos en forma de tabla, disponiendo los elementos que tienen propiedades similares unos debajo de otro en un conjunto de columnas verticales. Al principio de la década de 1860, el químico francés Alexandre Beguyer de Chancourtois y el químico británico John Newlands, cada uno de ellos independientemente del otro, propusieron varias versiones de esta idea, pero su trabajo fue ignorado. Lo que es peor, los contemporáneos de Newlands se burlaron de su idea, diciendo que hacer una lista ordenada de los elementos no tenía más sentido que colocarlos en orden alfabético. Fue un extraordinario ejemplo de redomada (y arrogante) estupidez, ya que el alfabeto es una convención arbitraria inventada por los seres humanos, mientras que los pesos de los átomos constituyen una propiedad física fundamental. No obstante, la anécdota muestra lo lejos que estaban los físicos de aceptar la realidad de los átomos a mediados de la década de 1860.

Incluso cuando, por fin, esta idea empezó a ser aceptada, hubo un elemento de controversia. Al final de la década de 1860, el alemán Lothar Meyer y el ruso Dmitri Mendeleiev, cada uno de ellos de forma independiente —y sin tener ninguno de ellos conocimiento de la obra de Beguyer de Chancourtois y Newlands— plantearon la idea de representar los elementos en una tabla periódica (una red de cuadrículas

parecida a una tabla de ajedrez) en la que éstos se colocaban en orden según sus pesos atómicos, situando los elementos que tienen propiedades químicas similares uno debajo de otro. Sin embargo, el resultado se conoce actualmente como tabla periódica de Mendeleiev, por lo que Meyer quedó relegado a ser para la historia tan sólo una nota a pie de página, al igual que los otros dos pioneros de la idea, sólo porque Mendeleiev fue lo bastante sagaz como para reordenar los elementos en la tabla, cambiando un poco el orden, y así garantizar que los elementos que presentaban propiedades químicas similares quedaban dispuestos en la misma columna vertical, aunque esto significara revolver ligeramente el orden de los pesos atómicos.

Estos cambios fueron realmente insignificantes. Por ejemplo, el telurio tiene un peso atómico igual a 127'61, justo una pizca más que el yodo, cuyo peso atómico es 126'91. La inversión del orden de estos dos elementos en la tabla permitió a Mendeleiev colocar el yodo debajo del bromo, al que se parece muchísimo químicamente, y el telurio bajo el selenio, dado que ambos se parecen mucho en cuanto a propiedades químicas, en vez de disponer el telurio bajo el bromo y el yodo bajo el selenio. Actualmente sabemos que Mendeleiev tenía razón al efectuar estos cambios, porque el peso de un átomo se determina por el número combinado de protones y neutrones que hay en él, mientras que sus propiedades químicas se relacionan sólo con el número de protones —veremos más sobre esto en el próximo capítulo—; pero en el siglo XIX no se conocían ni los protones ni los neutrones, por lo que no había manera de que Mendeleiev pudiera explicar la razón física de su ligero reordenamiento de los elementos en una tabla que estaba basada en los pesos atómicos, y lo que hizo fue confiar en la evidencia química de las similitudes.

El paso más audaz de Mendeleiev, el que condujo finalmente a la aceptación generalizada de su tabla periódica como algo relacionado con las propiedades fundamentales de los elementos (y no una convención arbitraria como el alfabeto), fue su decisión de dejar huecos en aquellos espacios de la tabla en que no había elementos conocidos con propiedades que «correspondieran» a dichos espacios. Para el año 1871 Mendeleiev ya había realizado una tabla que contenía los sesenta y tres elementos conocidos en aquella época. Además, mostraba la marcada periodicidad en la cual las familias constituidas por elementos cuyos pesos atómicos eran múltiplos de ocho veces el peso atómico del hidrógeno tenían propiedades químicas similares. Sin embargo, incluso después de los pequeños ajustes como el intercambio de posiciones entre el telurio y el yodo, para conseguir que esta configuración funcionara tuvo que dejar tres huecos en la tabla, lo cual constituyó una audaz predicción de que se descubrirían nuevos elementos cuyas propiedades (que él también especificó) habrían de corresponder a los lugares de esos huecos de la tabla. Estos tres elementos se descubrieron durante los quince años siguientes, y tenían

exactamente las propiedades que él había predicho: el galio, descubierto en 1875; el escandio, en 1879; y el germanio, en 1886.

Siguiendo la tradición clásica de la ciencia («Si no concuerda con el experimento, es falso») Mendeleiev había realizado una predicción y se había comprobado que ésta era correcta. Este hecho persuadió a los científicos de que la tabla periódica era importante y, cuando se descubrieron nuevos elementos y se vio que cada uno de ellos encajaba en la tabla de Mendeleiev, la aceptación de estas ideas se convirtió en entusiasmo. Actualmente hay noventa y dos elementos de los que se sabe que existen de forma natural en la Tierra y más de veinte elementos pesados que se han creado artificialmente en aceleradores de partículas. Todos ellos encajan en la tabla de Mendeleiev y han permitido realizar algunas mejoras en su configuración. Estas mejoras se han llevado a cabo en el siglo xx con el fin de reflejar nuestro modo actual de comprender la estructura de los átomos.

Pero incluso el éxito de la tabla periódica de Mendeleiev en el último tercio del siglo xix no llegó a convencer a todos de la existencia real de los átomos. La aceptación final de la «hipótesis atómica», como la denominó Feynman, no llegó hasta la primera década del siglo xx y llegó en buena medida gracias al trabajo del hombre cuya imagen ocupa la cima de la ciencia del siglo xx: Albert Einstein.

Einstein seguramente habría aprobado los comentarios de Feynman sobre la importancia de la hipótesis atómica. Este físico de origen alemán dedicó conscientemente sus primeros esfuerzos como investigador científico a realizar varios intentos para probar la realidad de los átomos y las moléculas, principalmente en su tesis doctoral (finalizada en 1905) y luego en una serie de publicaciones científicas en las que enfocaba el rompecabezas de distintas maneras y donde consiguió hallar varios procedimientos diferentes para hallar un valor del número de Avogadro. El hecho de que un físico con la perspicacia de Einstein sintiera la necesidad de hacer esto a principios del siglo xx demuestra, tanto la importancia del concepto de átomo, como el constante rechazo de los científicos a asumir este concepto. No hay duda de que era esto lo que tenía en su mente cuando realizó esta enorme tarea. Como escribiría más tarde a Max Born, «mi propósito era principalmente... hallar hechos que atestiguaran la existencia de átomos de un tamaño determinado».

El primer modo de aproximación al problema por parte de Einstein (en su tesis doctoral) consistió en calcular la velocidad a la cual las moléculas de azúcar de una disolución en agua pasan a través de una barrera llamada membrana semipermeable, y comparar sus cálculos (que dependen del tamaño de las moléculas y de su recorrido libre medio) con los resultados de los experimentos llevados a cabo por otros científicos. Esto es conceptualmente muy similar al procedimiento mediante el cual Loschmidt consiguió acercarse al número de Avogadro, usando los recorridos libres medios y los tamaños moleculares de algunos gases. Sin embargo, una característica

clave de Einstein es que nunca fue un experimentador y confió en los resultados experimentales aportados por otros científicos. Así, en 1905, el mejor valor que consiguió para el número de Avogadro, usando esta técnica, fue 21×10^{23} , y no porque hubiera nada equivocado en sus cálculos, sino porque los resultados experimentales no eran lo suficientemente exactos. En 1911, fue esta técnica, utilizando datos obtenidos en experimentos más precisos, la que le dio un valor de 66×10^{23} .

Pero en 1905 Einstein ya tenía otro método para «confirmar la existencia de los átomos». Dicho método incluía un fenómeno conocido como movimiento browniano, llamado así en honor del botánico escocés Robert Brown. En 1827 Brown había observado que los granos de polen que se encontraban en suspensión dentro del agua se podían ver (con la ayuda del microscopio) moviéndose de forma errática, como danzando en zigzag. Al principio, algunos interpretaron este sorprendente descubrimiento como un signo de que los granos de polen estaban vivos y en acción; pero pronto quedó claro que el mismo movimiento danzante se daba en diminutas partículas de polvo que no podían de ningún modo estar vivas.

En la década de 1860, varios físicos especularon con la posibilidad de que la causa de este movimiento pudiera ser un zarandeo producido por las moléculas del líquido en el que estaban las partículas (el mismo tipo de movimiento se observa, por ejemplo, en las partículas de humo de un cigarrillo que están en suspensión en el aire). Sin embargo, esta idea no prosperó en aquella época, en gran medida porque se había supuesto erróneamente que cada sacudida del grano (cada zig o cada zag) debía estar causada por el impacto de una sola molécula, lo que significaría que las moléculas tendrían que ser enormes, comparables en tamaño a las partículas en suspensión, lo cual evidentemente, incluso en aquella época, resultaba incorrecto.

Einstein abordó el problema desde el otro extremo. Estaba convencido de la realidad de la existencia de átomos y moléculas y deseaba encontrar el modo de convencer a los demás. Se dio cuenta de que una pequeña partícula en suspensión dentro de un líquido tenía que estar siendo sacudida por las moléculas de ese líquido y calculó el tipo de sacudida que se produciría. Einstein no sabía gran cosa sobre la historia del estudio del movimiento browniano (a lo largo de toda su carrera, Einstein nunca leyó mucho sobre la historia de ningún tema que le interesara, ya que prefería averiguar las cosas por sí mismo a partir de principios básicos, un método excelente para hacer física, siempre que se sea tan listo como Einstein). Así, en su primera publicación sobre el tema, calculó el modo en que las partículas en suspensión en un líquido tendrían que moverse, y sólo dijo, precavidamente, que «es posible que los movimientos de que hablamos aquí sean idénticos a lo que se llama movimiento molecular browniano». Los colegas que leyeron esta publicación le confirmaron al momento que lo que había descrito en términos matemáticos era exactamente el

movimiento browniano observado; así, en cierto modo, Einstein predijo el movimiento browniano y los experimentos confirmaron sus predicciones. Estaba especialmente convencido de la idea, asimismo cierta, de que podemos ver directamente los movimientos responsables del calor con sólo observar a través del microscopio las pequeñas partículas en suspensión dentro de cualquier líquido. Como indicó en 1905, «según la teoría cinética molecular del calor, unos cuerpos en suspensión dentro de un líquido, y que por su tamaño sólo sean visibles mediante el microscopio, realizarán movimientos de tal magnitud que resultará fácil observarlos microscópicamente».

La idea en la que Einstein basó sus cálculos era que incluso un objeto tan pequeño como un grano de polen se ve sacudido por todos los lados en cualquier instante por un número muy grande de moléculas. Cuando este objeto se mueve a sacudidas en una dirección, esto no se debe a que haya recibido un único golpe fuerte que le impulsa en tal dirección, sino a que se ha producido un desequilibrio temporal en la sacudida, es decir, que en ese instante concreto le han golpeado por un lado unas cuantas moléculas más que por el otro. A continuación, Einstein aplicó hábilmente algunos instrumentos matemáticos para hallar la estadística de este tipo de sacudida y predecir, en consecuencia, el recorrido en zigzag que el grano de polen (o lo que fuese) iba a realizar, teniendo en cuenta que cada pequeño movimiento se producía en una dirección totalmente aleatoria. Resulta que la distancia que recorre la partícula desde su punto de partida aumenta en proporción a la raíz cuadrada del tiempo transcurrido. En cuatro segundos recorre el doble de la distancia que recorre en un segundo (2 es la raíz cuadrada de 4), tarda dieciséis segundos en recorrer cuatro veces la distancia que recorre en un segundo (4 es la raíz cuadrada de 16), y así sucesivamente. La dirección en la que la partícula acaba su recorrido está orientada aleatoriamente con respecto a su punto de partida, al margen de que se mueva durante cuatro segundos, dieciséis segundos o cualquier otro número de segundos. En la actualidad esto se conoce como «recorrido aleatorio», y el mismo tipo de estadística es trasladable a muchas otras áreas de la ciencia, incluido el comportamiento de los átomos radiactivos cuando se desintegran progresivamente.

La relación entre el movimiento browniano y el número de Avogadro era un hecho claro para Einstein, por lo que sugirió el modo en que se podían llevar a cabo los experimentos encaminados a estudiar el movimiento exacto de las partículas en suspensión dentro de cualquier líquido y cómo se podía calcular el número de Avogadro a partir de dichos estudios. Pero, como de costumbre, no realizó los experimentos él mismo. Esta vez fue Jean Perrin, en Francia, quien hizo los experimentos. Perrin estudió el modo en que forma capas la suspensión de partículas en un líquido y descubrió que la mayoría de las partículas se sitúan cerca del fondo y unas cuantas más arriba. Las pocas partículas que suben más arriba dentro del líquido

(a pesar de sufrir la atracción hacia abajo por efecto de la gravedad) lo hacen porque son empujadas hacia zonas más altas por el movimiento browniano y la altura que alcanzan depende del número de choques que reciban, lo cual depende, a su vez, del número de Avogadro.

En 1908 Perrin utilizó esta técnica con el fin de hallar un valor para el número de Avogadro muy próximo al valor hallado hasta entonces mediante varias técnicas diferentes. Sus experimentos (combinados con las predicciones de Einstein) se consideran en general como lo que marcó el momento (hace menos de un siglo) en que ya no se podía seguir dudando del concepto de átomo. Einstein escribió a Perrin en 1909 lo siguiente: «Creía que era imposible investigar el movimiento browniano con tanta precisión». El propio Perrin escribió el mismo año:

Creo que es imposible que una mente libre de todo prejuicio pueda meditar sobre la extrema diversidad de los fenómenos que convergen así en el mismo resultado sin experimentar una fuerte impresión, y también pienso que de ahora en adelante será difícil defender mediante argumentos racionales una actitud hostil a la hipótesis molecular.

Confiamos en que ustedes estén asimismo convencidos de que el modelo atómico es un buen modelo. Sin embargo, antes de que pasemos a mirar lo que sucede en el interior del átomo, nos gustaría compartir con ustedes uno más de los distintos fenómenos que señalan en la misma dirección: el azul del cielo.

Esta historia se remonta a la obra de John Tyndall en la década de 1860, pero culmina, una vez más, en un trabajo realizado por Albert Einstein. Tyndall constató que la razón por la cual el cielo es azul se debe a que la luz azul se dispersa por el cielo más fácilmente que la luz roja. La luz procedente del Sol contiene todos los colores del arco iris (o espectro), con el color rojo en un extremo del espectro, mientras la luz azul, la de color añil y la luz violeta están en el otro extremo. Todos los colores del espectro se mezclan para dar como resultado la luz blanca. La luz roja tiene longitudes de onda más largas que la luz azul, lo que significa (entre otras cosas) que, al reflejarse en pequeñas partículas, no se dispersa con tanta facilidad como el extremo azul del espectro. La idea de Tyndall era que la dispersión que es la causa de que el cielo se vea azul, haciendo que la luz azul salte de una partícula a otra por todo el cielo, está originada por partículas de polvo y pequeñas gotas de líquido que se encuentran en suspensión en el aire.

Esto no era muy acertado. Este tipo de dispersión explica por qué las puestas de Sol y los amaneceres son rojos —en el polvo o en la neblina de vapor que está cerca del horizonte la luz roja penetra mejor que la luz azul— pero han de ser verdaderamente muy pequeñas las «partículas» que se necesitan para dispersar la luz

azul por todo el cielo, de modo que parezca venir hacia nosotros desde todas las direcciones. A finales del siglo XIX y principios del XX, varios físicos sugirieron que la dispersión podría estar causada por las propias moléculas del aire. Pero fue Einstein el que realizó los cálculos definitivos y probó, en una publicación escrita en 1910, que el azul del cielo está originado en efecto por la dispersión de la luz desde las moléculas del aire. Y, una vez más, el número de Avogadro se puede deducir de estos cálculos. En consecuencia, ni siquiera se necesita un microscopio para comprobar que las moléculas y los átomos existen; todo lo que se necesita es mirar el cielo azul en un día claro.

Para situar en perspectiva algunos de los números relacionados con los átomos, recuérdese que el peso molecular de cualquier sustancia en gramos contiene un número de moléculas igual al número de Avogadro. Por ejemplo, treinta y dos gramos de oxígeno contienen algo más de 6×10^{23} moléculas de oxígeno. Más adelante en este libro discutiremos sobre la naturaleza del universo considerado a gran escala. Nuestro Sol y nuestro sistema solar forman parte de una galaxia de estrellas, la cual tiene forma de disco y se llama la Vía Láctea. Esta galaxia contiene unos pocos cientos de miles de millones de estrellas (unas pocas veces 10^{11}), cada una de ellas bastante parecida al Sol. En todo el universo existen varios cientos de miles de millones de galaxias visibles en principio mediante telescopios. Un proyecto de investigación en el que participé en la Universidad de Sussex demostró que nuestra galaxia es ligeramente menor que la media de las galaxias con forma de disco. Así pues, a todo esto, ¿cuántas estrellas hay en total? Si multiplicamos los números, unas pocas veces varias vendría a ser aproximadamente 10 , y $10^{11} \times 10^{11}$ es 10^{22} , con lo que tenemos 10×10^{22} , o 10^{23} estrellas en total. Redondeando, el número de estrellas resulta ser un poco menos que el número de Avogadro; esto significa que hay unas cuantas veces más moléculas de oxígeno en sólo treinta y dos gramos de este gas (trece litros en condiciones estándar de presión y temperatura) que estrellas hay en todo el universo visible.

Para ver esto desde una perspectiva humana, digamos que la capacidad máxima de los pulmones del hombre es de alrededor de seis litros; por lo tanto, si tomamos aire hasta el máximo de capacidad pulmonar, tendremos más moléculas de aire en los pulmones que estrellas hay en el universo visible.

Si tienen que caber tantas moléculas en una cantidad tan pequeña de materia, cada molécula (y cada átomo) ha de ser de un tamaño muy pequeño. Hay muchas maneras de calcular los tamaños de los átomos y de las moléculas; el más sencillo consiste en tomar una cantidad de un líquido o un sólido, de tal modo que su volumen contenga un número de partículas igual al número de Avogadro (por ejemplo, treinta y dos gramos de oxígeno líquido), y dividir el volumen entre el número de Avogadro para averiguar el tamaño que tiene cada partícula —una idea que se remonta a los trabajos

de Cannizzaro pero que se puede llevar a la práctica actualmente con mucha más precisión—. Al hacer esto, resulta que todos los átomos son aproximadamente del mismo tamaño, siendo el mayor (el del cesio) de 0'0000005 mm de diámetro. Esto significa que serían necesarios diez millones de átomos, juntos en hilera uno tras otro, para recorrer el espacio que hay entre dos de las puntas del borde en sierra de un sello de correos.

A principios del siglo xx, ya se iba aceptando plenamente la idea de la existencia de estas diminutas entidades. Sin embargo, a partir de entonces y durante unas cuantas décadas, muchos de los grandes logros de la física partieron, no sólo del estudio del comportamiento de los átomos, sino de comprobar la estructura interna de los mismos, descendiendo a escalas de una diezmilésima del tamaño del átomo para estudiar su núcleo, y posteriormente a escalas aún menores para estudiar las partículas fundamentales de la naturaleza (al menos así se consideran actualmente, a finales del siglo xx). Podemos empezar a hacernos una idea de lo que sucede en el interior del átomo observando el modo en que están dispuestos los electrones en la parte más externa de los átomos y cómo ejercen interacciones con la luz.

En el interior del átomo

Aunque al principio no fueron conscientes de que era eso lo que estaban haciendo, los físicos habían comenzado ya a investigar en el interior del átomo en la segunda mitad del siglo XIX, antes de que el concepto de átomo fuera aceptado plenamente. Todo esto llegó a través del estudio de la electricidad.

Los físicos que estaban estudiando las propiedades de la electricidad deseaban investigarla en el estado más puro posible. La electricidad que llega a través de un cable está afectada por las propiedades del cable; incluso las descargas eléctricas que chispean en el aire, como la iluminación artificial, a través de un espacio entre dos placas cargadas están influidas por las propiedades del aire. Lo que los físicos necesitaban era un modo de estudiar en el vacío las descargas eléctricas que saltan atravesando el espacio existente entre dos electrodos cargados, y esto no llegó a ser posible hasta mediados de la década de 1880, cuando Johann Geissler inventó una bomba de vacío lo bastante potente como para extraer el aire de un recipiente de cristal hasta conseguir que la presión fuera sólo unas pocas diezmilésimas de la presión del aire al nivel del mar.

Varios científicos empezaron a realizar experimentos en los que se utilizaba la electricidad que pasaba a través de tubos de los que se había extraído el aire. El principio básico eran dos cables (los electrodos) que se sellaban dentro del tubo de cristal, uno en cada extremo, con un espacio entre ellos. Uno de los cables (llamado ánodo) se empalmaba a una fuente de electricidad positiva (por ejemplo, la terminal positiva de una batería) y el otro (el cátodo) se empalmaba a una fuente de electricidad negativa (por ejemplo, la terminal negativa de la misma batería). Pronto quedó claro que, cuando una corriente eléctrica fluía en el circuito, había algo que emergía del cátodo y se desplazaba cruzando el espacio vacío hasta el ánodo. Ese «algo» hacía patente su aparición creando una incandescencia en la pequeñísima cantidad residual de aire que había en el tubo (un precursor del moderno tubo de neón), o bien chocando con la pared de cristal del tubo y poniéndola incandescente.

Este fenómeno fue investigado con todo detalle por William Crookes, quien llegó a la conclusión de que estos rayos catódicos, que así era como se llamaban, eran moléculas gaseosas del aire que aún quedaba en el tubo. Según Crookes, estas moléculas habían tomado del cátodo carga eléctrica negativa, siendo después repelidas por el cátodo y atraídas hacia el ánodo (las cargas del mismo signo se repelen entre sí; las de signo opuesto se atraen). Pero estaba equivocado. Un simple cálculo no tardó en demostrar que, incluso a la baja presión del interior de los tubos, el recorrido libre medio de una molécula de aire (el promedio de distancia que podría recorrer sin chocar con otra molécula) era sólo alrededor de medio centímetro; pero resulta que los rayos catódicos recorrían casi un metro en línea recta en algunos de

los tubos más largos que se construyeron en aquellos tiempos.

Sin embargo, los rayos catódicos parecían ser un chorro de partículas. En 1895, Jean Perrin (al que ya conocimos en el capítulo 1) demostró que los misteriosos rayos eran desviados lateralmente por un campo magnético, exactamente de la misma manera que lo sería un chorro de partículas con carga negativa, y descubrió que, cuando los rayos del cátodo chocaban con una placa de metal, esta placa se cargaba negativamente. Entonces comenzó a idear experimentos para averiguar cuáles eran las propiedades de las partículas que se encontraban en un haz de rayos catódicos, pero perdió su derecho preferente sobre la investigación, ya que J. J. Thomson se le adelantó con sus trabajos en Inglaterra, siendo el primero en realizar los descubrimientos.

Thomson ideó un brillante experimento para abordar la cuestión de la naturaleza de los rayos catódicos; y en este caso «ideó» es el término correcto, porque J. J., como se le llamaba, era notoriamente desmañado y los experimentos con los que soñaba tenían que ser realizados por personas más diestras que tuvieran unas manos más hábiles. Lanzando un haz de rayos catódicos, una vez con un campo magnético y otra con un campo eléctrico, el equipo de Thomson consiguió hallar la fuerza de cada tipo de campo que se necesitaba para que los rayos catódicos se movieran en línea recta. A partir de esto este físico pudo averiguar el valor del cociente entre la carga eléctrica de una sola partícula del haz y su masa —expresado brevemente e/m —. Este valor tenía que ser el mismo para cada partícula del haz, puesto que todas las partículas se desplazaban juntas en el campo eléctrico y en el magnético.

Puede que esto no parezca realmente un avance. Lo que de verdad se desea es hallar e y m de forma separada. Pero la cuestión crucial fue que, en 1897, Thomson descubrió que, para las partículas de los rayos catódicos, e/m sólo era alrededor de una milésima del valor de e/m para el llamado átomo de hidrógeno ionizado, la partícula de carga más ligera que se conocía en aquella época. Un átomo de hidrógeno ionizado es un átomo de hidrógeno que ha perdido una unidad de carga eléctrica (actualmente ya sabemos que esto se debe a que ha perdido un electrón). Dando un atrevido salto (no enteramente justificado en aquellos tiempos) Thomson supuso que la cantidad de carga eléctrica negativa de cada una de las partículas de los rayos catódicos era igual en magnitud (aunque de signo opuesto) que la cantidad de carga positiva de un ion de hidrógeno. En ese caso, la masa de cada una de las partículas de los rayos catódicos debía ser sólo una milésima de la masa de un átomo de hidrógeno.

Thomson comunicó sus resultados a la Royal Institution de Londres en 1897, comentando que «la suposición de un estado de la materia más finamente subdividido que el átomo de un elemento no deja de ser algo sorprendente». De hecho, resultó tan sorprendente que algunos de los miembros de la audiencia pensaron que les estaba

tomando el pelo, y tuvieron que pasar dos años, hasta 1899, para que se dispusiera ya de resultados experimentales que aportaran una prueba convincente de la existencia real de estas pequeñas partículas cargadas negativamente, que pronto llegaron a ser conocidas como electrones. Sin embargo, en su ansia por celebrar aniversarios, la mayoría de los físicos considera actualmente el año 1897 como la fecha del descubrimiento del electrón y, por consiguiente, se celebró el centenario de dicho descubrimiento en 1997.

Así pues, a principios del siglo xx, mientras algunos científicos (incluido Einstein) todavía estaban dando los toques finales a las distintas demostraciones de la existencia del átomo, otros físicos ya estaban intentando averiguar cuál debía ser la estructura interna del átomo y cómo podían separarse los electrones, o eliminarse, de lo que previamente se había considerado como una entidad indivisible.

En varios experimentos ideados por Thomson se siguió una línea de ataque que incluía el estudio del comportamiento de iones cargados con carga positiva. Fue Thomson (o, más bien, su equipo) quien demostró que, de hecho, estos iones se comportan exactamente como si fueran átomos de los cuales se ha quitado alguna carga eléctrica negativa, dejándolos con una carga resultante positiva; exactamente lo que sabemos que sucedería si de los átomos se hubieran extraído electrones, que tienen carga negativa.

Durante la segunda década del siglo xx, esta investigación condujo al descubrimiento de que no todos los átomos de un mismo elemento tienen necesariamente la misma masa. El equipo de Thomson midió el valor del cociente e/m en iones que tenían la misma cantidad de carga eléctrica y utilizaron esto para determinar sus masas. Hallaron que, por ejemplo, el neón aparece en dos variedades, una de las cuales tiene átomos cuya masa es veinte veces la de un solo átomo de hidrógeno, y la otra tiene átomos cuya masa es veintidós veces la masa del átomo de hidrógeno. Estas variedades diferentes del mismo elemento se conocen actualmente como isótopos. Es precisamente la existencia de isótopos diferentes del mismo elemento la que es responsable del hecho de que el peso atómico medio de un elemento pueda no ser un múltiplo entero del peso atómico del hidrógeno, y lo que explica por qué algunos de los elementos de la tabla periódica de Mendeleiev parecen estar fuera de la hilera que les corresponde por razón de sus propiedades químicas, si se procede a disponerlos según sus pesos atómicos medios. Sin embargo, para cualquier isótopo puro el peso atómico es siempre un múltiplo exacto del peso de un átomo de hidrógeno. Esto constituyó un importante descubrimiento a la hora de determinar la naturaleza de la estructura interna de los átomos.

La otra línea principal de ataque provenía de los teóricos que intentaban construir un modelo que explicara el descubrimiento de Thomson, según el cual los átomos están formados por electrones junto con alguna otra cosa y, además, dichos electrones

pueden ser extraídos de sus contenedores. El primero de estos modelos fue idea de lord Kelvin en 1902 (Kelvin se llamaba en realidad William Thomson, pero no tenía parentesco con J. J.). Este modelo del átomo lo describía como una diminuta esfera cuyo diámetro era de un décimo de una milmillonésima de metro ($0,1 \times 10^{-9}$ m o 0,1 nanómetros), con una carga positiva repartida uniformemente por la esfera y electrones empotrados en ella como guindas en un pastel. Sin embargo, en aquellos tiempos el descubrimiento de la radiactividad había aportado a los físicos los medios necesarios para comprobar la estructura del átomo, y, si no concordaba con el experimento, entonces era falso.

La primera evidencia de que dentro del átomo se producían procesos energéticos se produjo en 1895, cuando Wilhelm Röntgen descubrió los rayos X. Como muchos de los físicos contemporáneos suyos, por aquel entonces Röntgen estaba experimentando con los rayos catódicos, y observó que cuando el haz de rayos catódicos choca con un objeto (incluso si se trata de la pared de cristal del tubo de rayos catódicos) puede hacer que dicho objeto emita otro tipo de radiación. Dio la casualidad de que Röntgen tenía depositado sobre un banco del laboratorio cerca del experimento con rayos catódicos un artilugio llamado pantalla fluorescente y observó unos destellos luminosos procedentes de dicha pantalla que aparecían cuando el experimento de rayos catódicos estaba en marcha. La causa de estos destellos era la misteriosa radiación X, llamada así porque en matemáticas tradicionalmente X es la magnitud desconocida.

Hoy día sabemos que los rayos X son precisamente como la luz, una forma de radiación electromagnética, pero con longitudes de onda mucho más cortas que las de la luz visible. Sin embargo, lo que resulta interesante para la comprensión de la estructura del átomo es que el descubrimiento de los rayos X en 1895 estimuló a otros científicos para buscar otras formas de radiación procedentes de los átomos y, como consecuencia, en 1896, Henri Becquerel descubrió que los átomos de uranio producen espontáneamente otro tipo de radiación. Dos años más tarde, Ernest Rutherford demostró que existen en realidad dos clases de radiación atómica de este tipo, a las que llamó rayos alfa y rayos beta. Una tercera clase de radiación, los rayos gamma, fue descubierta posteriormente. Se constató que los rayos beta eran electrones de movimiento muy rápido, y los rayos gamma resultaron ser una radiación electromagnética, como los rayos X, pero con unas longitudes de onda aún más cortas. Sin embargo, eran los rayos alfa los que encerraban el secreto que sería en lo sucesivo objeto de la investigación sobre la estructura del átomo.

A principios del siglo xx nadie sabía qué eran los rayos alfa, salvo que eran partículas que llevaban dos unidades de carga positiva. Rutherford demostró que tenían la misma masa que un átomo de helio (en términos estrictos y con una terminología moderna, tenían la masa de un tipo determinado de átomo de helio: el

isótopo conocido como helio-4). Son como átomos de helio a los que se les ha privado de dos electrones, lo que actualmente se conoce como núcleos de helio. Sin embargo, lo que a Rutherford le interesaba era su carácter de partículas de rápido movimiento, con las que se podía bombardear los átomos para comprobar su estructura.

En 1909, dos físicos que trabajaban bajo la dirección de Rutherford, Hans Geiger y Ernest Marsden, realizaron el asombroso descubrimiento: cuando se disparaba un haz de partículas alfa contra una fina lámina de oro, la mayoría de las partículas pasaba directamente a través de la lámina, pero unas pocas rebotaban casi en la misma dirección en que habían venido. Esto, dijo Rutherford más tarde, fue «prácticamente el hecho más increíble que me ha sucedido jamás... Casi tan increíble como si al disparar un proyectil de 15 pulgadas contra un trozo de papel de seda, el proyectil se diera la vuelta y te impactara».

No había ningún modo de explicar este hecho mediante el modelo de pastel de guindas que se había realizado para el átomo. Según este modelo, los átomos estarían apretados unos junto a otros en la lámina de oro, con una distribución uniforme de la densidad por toda la lámina. Las partículas alfa tendrían que disminuir su velocidad al pasar a través de las apretadas hileras de átomos —como una bala de rifle que frena su velocidad si se dispara a través de un depósito de agua—, pero no debería haber ningún tipo de zonas duras donde las partículas alfa pudieran rebotar. Por lo tanto, Rutherford tuvo que realizar otro modelo de átomo mejor, uno que concordara con los experimentos. Lo anunció en 1911 y se trataba del modelo que se estudia hoy en la escuela.

En el modelo del átomo de Rutherford, casi toda su masa está concentrada en un diminuto núcleo central, que tiene carga positiva, y alrededor de este núcleo describen sus órbitas los electrones, cargados negativamente, de una manera algo parecida (en una aproximación muy burda) al modo en que los planetas describen sus órbitas alrededor del Sol. Además, del mismo modo que la mayor parte del sistema solar es espacio vacío, también lo es la mayor parte del volumen que ocupa un átomo. La mayoría de las partículas alfa disparadas contra cualquier objeto «sólido» pasan alrededor del núcleo, a través de las nubes de electrones, casi sin ser obstaculizadas; pero, ocasionalmente, una partícula alfa puede chocar de frente contra un núcleo y ser desviada.

A partir de la frecuencia con que se producen esas desviaciones, Rutherford pudo averiguar el tamaño del núcleo en comparación con el tamaño del átomo. Un átomo tiene un diámetro de aproximadamente 10^{-8} cm; pero el diámetro del núcleo es sólo 10^{-13} cm. Es una cienmilésima del diámetro de todo el átomo, equivalente al tamaño de una cabeza de alfiler comparada con la cúpula de la catedral de San Pablo en Londres. Puesto que el volumen es proporcional al cubo del radio (o del diámetro),

esto significa que la proporción entre la materia sólida y el espacio vacío en un átomo no es de 10^{-5} , sino de 10^{-15} . Solamente una milésima de una billonésima de átomo es núcleo sólido. Así pues, dado que todo lo que existe en la Tierra está hecho de átomos, eso significa que nuestro propio cuerpo, y la silla en la que nos sentamos, están constituidos por mil billones de veces más espacio vacío que materia sólida. La única razón por la cual nuestro propio cuerpo, y la silla en la que está sentado, parecen sólidos e impenetrables es que esos diminutos puntos de materia se mantienen juntos por el efecto de fuerzas eléctricas que operan entre las partículas cargadas, es decir, los núcleos y los electrones.

Uniendo todo esto, tenemos un modelo de átomo que funciona muy bien, dentro de ciertos límites. En 1919 Rutherford había descubierto que a veces, cuando una partícula alfa acelerada choca contra un núcleo de nitrógeno, éste cambia convirtiéndose en un núcleo de oxígeno y se emite un núcleo de hidrógeno. Esto sugería la idea de que los núcleos de los átomos contienen partículas equivalentes a núcleos de hidrógeno, y hubo otros experimentos que lo confirmaban. A estas partículas se les bautizó con el nombre de protones. Como cada protón tiene una carga positiva exactamente igual en magnitud a la carga negativa de un electrón y los átomos son en conjunto eléctricamente neutros, resulta que, con el fin de que las cargas se equilibren, todos los átomos tienen un número de protones en su núcleo igual al número de electrones que hay en la nube exterior al núcleo. Por primera vez llegó a ser posible explicar la química en términos de la estructura interna del átomo (véase el capítulo 4).

Sin embargo, quedaba sin resolver una cuestión obvia en relación con el modelo de Rutherford y era cómo toda la carga positiva del núcleo podía permanecer allí sin separarse (repeliéndose las partículas entre sí); al menos, en el modelo del pastel de guindas las cargas positivas y negativas estaban mezcladas y podían mantenerse unas con otras en su sitio. La respuesta evidente era que podría haber algún tipo de partículas neutras en el núcleo que actuara como una especie de lastre para mantener unida a la materia positiva. Estas partículas, llamadas neutrones, no fueron identificadas hasta 1932.

Además de contribuir a mantener unida la materia de los núcleos, la presencia de los neutrones en dichos núcleos explicaba la existencia de isótopos. Los neutrones tienen una masa casi igual a la de los protones, por lo que realizan una gran contribución al peso del átomo. El comportamiento químico de un elemento depende casi por completo, como veremos en el capítulo 4, del número de electrones asociados con cada átomo de este elemento. Dicho número de electrones es igual al número de protones que hay en el núcleo. Sin embargo, el peso atómico de un elemento depende de la combinación del número de protones y de neutrones (su suma) que hay en cada núcleo, por lo que los átomos de diferentes isótopos, si tienen

distinto número de neutrones en sus núcleos, pero el mismo número de protones, pueden tener pesos diferentes, aunque sus propiedades químicas sean las mismas.

La radiación beta se produce cuando un núcleo emite un electrón, con lo que se gana una unidad de carga positiva en este proceso. Entrando en más detalles, podemos decir que uno de los neutrones del núcleo ha emitido un electrón, transformándose él mismo en un protón (éste es el proceso que se conoce como desintegración beta). Debido a que esto aumenta el número de protones del núcleo, éste cambia convirtiéndose en el núcleo de otro elemento. De manera similar, en la desintegración alfa un núcleo inestable emite un paquete de dos protones y dos neutrones, todo ello junto formando una partícula alfa. Una vez más, el núcleo que queda se ha transformado así en un núcleo de otro elemento.

Incluso con la presencia de neutrones para mantener unidas las componentes del núcleo, sigue habiendo mucha cantidad de carga positiva que intenta romper y separar dicho núcleo. En consecuencia, los físicos eran conscientes de que tenía que haber un tipo de fuerza hasta entonces desconocida, pero conocida ya actualmente como fuerza nuclear, que mantuviera unido todo lo que hay en el núcleo. La potente fuerza nuclear es bastante diferente de las dos fuerzas que percibimos en el entorno cotidiano: la gravedad y el electromagnetismo. Ambas son unas fuerzas de largo alcance —cualquiera que haya jugado con un imán sabe cómo su fuerza parece prolongarse a través del espacio y las puertas cerradas hasta actuar sobre un trozo de metal situado en las proximidades—, mientras que la gravedad desde la Tierra consigue mantener a la Luna en su órbita (¡y también más allá!). Pero la potente fuerza nuclear sólo tiene un alcance muy corto. Se trata de una fuerza de atracción que opera tanto en los neutrones como en los protones, empujándolos unos hacia otros. A distancias muy cortas (del orden del volumen de un núcleo) es casi cien veces más fuerte que la fuerza eléctrica, por lo que supera la repulsión entre los protones cargados positivamente. Pero, a diferencia de la fuerza eléctrica, la fuerza nuclear reduce su efecto rápidamente con la distancia.

Los protones y los neutrones (que colectivamente se suelen denominar nucleones) sólo sienten esta fuerza si se encuentran prácticamente en contacto unos con otros. Así, por ejemplo, si (por cualquier razón) una partícula alfa cargada positivamente consigue alejarse un poco del núcleo, la poderosa fuerza nuclear efectivamente desaparece y la partícula es repelida por la carga positiva de lo que queda del núcleo. Se puede hacer una analogía con un muelle rígido que apretamos entre las manos, comprimiéndolo y venciendo la fuerza que intenta hacer que se expanda; en cuanto lo soltamos, el muelle salta con toda su fuerza. En el próximo capítulo tendremos más cosas que decir sobre las fuerzas de la naturaleza. Sin embargo, vale la pena mencionar de paso que también se considera que la desintegración beta está causada por una fuerza, que se llama interacción nuclear débil.

La imagen que empezó a surgir a partir de los trabajos de Rutherford era la de un núcleo diminuto y con carga positiva rodeado de electrones que se movían «en órbitas» alrededor de él. Hubo que esperar hasta la década de 1930 para llegar a identificar el neutrón como partícula individual y a estudiar sus propiedades, pero en el lenguaje actual se puede decir que un átomo de hidrógeno corriente se compone de un único protón y un único electrón que describe una órbita a su alrededor. Existe un isótopo del hidrógeno, llamado deuterio o hidrógeno pesado, en el que cada núcleo contiene un protón y un neutrón, pero la parte exterior de este átomo sigue teniendo únicamente un solo electrón. El siguiente elemento es el helio, que tiene dos protones en su núcleo. Sin un neutrón que los mantuviera unidos, la fuerza nuclear no sería efectiva y los dos protones por sí mismos se repelerían mutuamente debido a la carga positiva de cada uno de ellos. Así pues, incluso el átomo más sencillo de helio tiene dos protones y un neutrón en su núcleo, y dos electrones en el exterior de éste. Se le conoce, por razones obvias, como helio-3. La forma más común del helio (helio-4) tiene en realidad dos protones y dos neutrones en su núcleo, y cuenta con dos electrones en el exterior. El núcleo del helio-4 es lo mismo que una partícula alfa y se trata de una configuración especialmente estable.

Así se construyen los elementos. El isótopo más común del carbono, el carbono-12, tiene seis neutrones y seis protones en su núcleo, y seis electrones fuera de él. Gracias a esta agradable simetría, y porque resulta más fácil trabajar con el carbono que con el helio, los pesos atómicos se definen actualmente de tal forma que el peso del carbono-12 sea exactamente igual a 12; sin embargo, cualquiera que no esté preocupado por la precisión en los detalles puede seguir pensando que los pesos atómicos se miden en unidades de tal forma que el hidrógeno tiene una masa igual a la unidad. La diferencia es pequeña, pero real, porque un átomo de hidrógeno común sólo tiene en su núcleo un protón y ningún neutrón, siendo las masas de los protones y los neutrones ligeramente diferentes unas de otras.

En los elementos que son sucesivamente más pesados, existe una tendencia a tener más neutrones que protones en el núcleo, porque necesitan mantener unida una cantidad de carga positiva que es cada vez mayor. El hierro-56, por ejemplo, tiene 26 protones y 30 neutrones en su núcleo, mientras que el uranio-238 tiene 92 protones, pero nada menos que 146 neutrones en cada núcleo. En efecto, al llegar al uranio el núcleo se está haciendo tan grande, con tanta carga positiva en él, que la potente fuerza nuclear apenas es suficiente para mantener unido el núcleo, incluso con el pegamento extra que aportan los neutrones. Cada nucleón sólo nota el efecto de la fuerza nuclear de atracción proveniente de sus vecinos inmediatos, pero cada protón siente la fuerza eléctrica de repulsión de los otros 91 protones que hay en el núcleo, razón por la cual el número de los elementos que existen de forma natural sólo llega hasta 92. Los físicos pueden fabricar núcleos más pesados en los aceleradores de

partículas, pero estos núcleos son inestables y se rompen por sí solos en poco tiempo.

Sin embargo, toda esta discusión es sobre los núcleos de los elementos. Pero queda pendiente un gran enigma en relación con el modelo atómico de Rutherford, un enigma que se aclaró incluso antes de que se supiera de la existencia de los neutrones en el núcleo. ¿Cómo podían todos los electrones (o simplemente el único electrón del átomo de hidrógeno) mantenerse en órbita alrededor del núcleo?

El problema es que cuando una carga eléctrica es acelerada irradia energía, una radiación electromagnética, que es como la luz. Como ya mencionamos en el capítulo 1, incluso un movimiento circular a velocidad constante tiene una aceleración, porque la dirección de la partícula que gira está cambiando constantemente. Es este tipo de aceleración la que nos lanza hacia un lado cuando vamos en un coche que toma una curva a gran velocidad. Los experimentos realizados en el laboratorio confirman que, si se hace que una carga eléctrica gire en círculo, ésta irradia energía. Pero recordemos la ley de la conservación de la energía. Si un electrón irradia energía al recorrer una órbita dentro de un átomo, debe perder energía. La única fuente de energía que tiene para seguir moviéndose es la energía cinética que le mantiene en su órbita alrededor del núcleo. Por lo tanto, un electrón que describe una órbita alrededor de un núcleo debe irradiar energía y caer hacia abajo, describiendo espirales hasta que cae en el propio núcleo. Según las leyes de la física, tal como se entendían a principios del siglo xx (en lo que habitualmente recibe el nombre de física «clásica»), todo átomo debería colapsarse en un instante, dejando sólo una ráfaga de radiación.

La única forma de resolver este problema era invocar un nuevo tipo de física, una nueva clase de modelo para el átomo. Como siempre, sin embargo, esto no significa que se tenga que abandonar el viejo modelo. Debe suceder algo diferente a escala atómica que impida que los átomos se destruyan de esta manera; no obstante, la física clásica sigue siendo una descripción perfectamente válida de lo que le ocurre a una carga eléctrica a la que se le hace girar en el laboratorio.

La persona que dio el paso crucial hacia adelante para encontrar el modo de estabilizar el átomo de Rutherford fue Niels Bohr, un físico danés que trabajó durante un tiempo con Rutherford en Inglaterra. Introdujo el concepto de cuanto en la física atómica.

El modelo de Bohr para el átomo fue publicado en 1913, después de su retorno a Dinamarca. Este modelo era una descarada mezcla de ideas clásicas (sobre todo el concepto de órbita) y de las nuevas ideas cuánticas, con la gran ventaja de que funcionaba. Además, no sólo aportaba una explicación sobre la razón por la cual los electrones no se caen en el núcleo, sino que también explicaba la pauta de las líneas producidas por cada elemento y observadas en el espectro luminoso; en particular, la pauta asociada con el hidrógeno, el elemento más sencillo que existe. Todo esto se

explicó más tarde, en la década de los años veinte, con detalles aún más satisfactorios aportados a través de una descripción completa de los átomos y de la radiación en términos de mecánica cuántica. Sin embargo, el modelo de Bohr resulta tan fácil de entender que es el que la mayoría de nosotros aprendió en la escuela.

La idea básica de la física cuántica se remonta a los trabajos de Max Planck, que fueron hechos públicos por él mismo en Berlín en 1900, aunque se tardó mucho tiempo en profundizar en todas sus implicaciones.

Antes de 1900, el modelo que se había aceptado para la luz la consideraba como una forma de onda electromagnética que se movía a través del espacio. Varios descubrimientos realizados en el siglo XIX habían demostrado que un campo eléctrico variable produce un campo magnético, y que un campo magnético variable produce un campo eléctrico. Esta teoría de los campos es tan importante que volveremos a ella en el próximo capítulo, pero la idea resulta familiar después de haber jugado con juegos que tienen imanes de juguete; así, el campo del imán es aquella región en torno a él donde se ejerce su influencia magnética.

James Clerk Maxwell descubrió las ecuaciones que describen un par de campos variables que se desplazan juntos por el espacio; el campo eléctrico variable produce un campo magnético variable, y el campo magnético variable produce un campo eléctrico variable, y marchan juntos al paso. Ambos a la vez configuran una onda electromagnética. La luz, la radio y otras formas de ondas electromagnéticas están todas ellas descritas mediante las ecuaciones de Maxwell. La energía de la onda es suministrada por la energía introducida en el sistema al principio, por ejemplo, la corriente eléctrica que fluye por un cable para producir la incandescencia caliente de una bombilla eléctrica.

Todo esto era de lo más fino y elegante, y las ecuaciones de Maxwell constituyeron uno de los mayores triunfos de la física del siglo XIX. Sin embargo, había un problema. El considerar la luz sencillamente como una clase de onda no explicaba el tipo de radiación que realmente producen los objetos calientes. Creamos la radiación electromagnética, en última instancia, moviendo cargas eléctricas (de hecho, acelerando electrones). Pero si aplicamos el mismo tipo de reglas estadísticas que se aplican al modo en que las ondas se comportan en el mundo clásico —por ejemplo, las ondas sonoras que se obtienen al pulsar una cuerda de guitarra— resulta que un electrón acelerado tendría que producir un enorme número de ondas electromagnéticas con longitudes de onda muy cortas y difícilmente alguna con larga longitud de onda.

Planck encontró un modo de soslayar el problema tratando la radiación como si sólo pudiera ser emitida en paquetes de un tamaño determinado: los cuantos. La cantidad de energía que hay en cada cuanto de radiación es inversamente proporcional a la longitud de onda de dicha radiación, por lo tanto, según esto, un

cuanto correspondiente a longitudes de onda cortas precisa mucha más energía para formarse que un cuanto correspondiente a longitudes de onda largas. La relación entre la energía que un cuanto transporta y su longitud de onda depende de un número llamado actualmente constante de Planck, que se puede determinar mediante experimentos. En el caso de un cuanto concreto, la energía que transporta es igual a la constante de Planck dividida por la longitud de onda.

Si hacemos que haya un movimiento de muchos electrones a la vez, calentando la materia de la que forman parte, sólo unos pocos de ellos tendrán la energía suficiente para crear cuantos de alta energía y longitudes de onda cortas, por lo que solamente se emitirá una pequeña radiación de alta energía. Habrá muchos más electrones con suficiente energía para hacer cuantos de tamaño mediano con ondas de longitud mediana y, en consecuencia, se emitirá mucha radiación de energía mediana. Sin embargo, aunque haya muchos electrones con energía suficiente para hacer cuantos de baja energía y largas longitudes de onda, cada cuanto aporta sólo una pequeñísima cantidad, con lo que únicamente se emite una pequeña cantidad de radiación de alta energía.

Uniendo todas las contribuciones en la forma estadística correcta, Planck descubrió que un objeto caliente debería irradiar la mayor parte de su energía en la parte central de una banda de distintas longitudes de onda, siendo menor la energía emitida tanto en las longitudes de onda más cortas como en las más largas. Además, a medida que el objeto se calienta más, habrá más electrones con energía suficiente para producir una radiación de longitud de onda más corta, por lo que el balance se inclina hacia longitudes de onda más cortas y el pico de la emisión de energía se desplaza en el mismo sentido. Esto es exactamente lo que vemos en la vida real: un atizador al rojo vivo está menos caliente que un atizador incandescente con color anaranjado, y la luz roja tiene longitudes de onda más largas que la luz anaranjada.

La idea de Planck funcionaba perfectamente para explicar la naturaleza de la radiación electromagnética emitida por objetos calientes y cómo el color del objeto cambia a medida que se calienta más.[1] Pero la idea de la luz como una forma de onda electromagnética estaba tan bien arraigada que al principio nadie aceptó la idea de que cada cuanto de luz era una entidad real por propio derecho (lo que actualmente llamamos un fotón). La idea consistía en afirmar que había algo en relación con los átomos y los electrones que les impedía irradiar luz si no era en cantidades de un determinado tamaño, no que la luz sólo existiera en cantidades de un determinado tamaño.

Mi analogía favorita es la del cajero automático del banco. La máquina sólo facilita dinero en múltiplos de £10. Puedo sacar £20 o £60, o cualquier otro múltiplo de diez (con tal de que haya suficiente en mi cuenta y en el cajero automático). Pero nunca podré sacar, por ejemplo, £27'43, aunque tenga esta cantidad en mi cuenta y

aunque, desde luego, esta cantidad de dinero existe. La causa es que la máquina no funciona de esta manera. Del mismo modo, a principios del siglo xx los físicos pensaban que la radiación electromagnética existía en ondas de cualquier tipo de longitud, con cualquier cantidad de energía, pero que los átomos sólo tenían «permitido» emitir luz en paquetes de una cierta cantidad de energía para cada longitud de onda en particular.

Durante los veinticinco años siguientes, poco a poco se fue viendo cada vez más claro que en algunas circunstancias los cuantos de luz —los fotones— tenían que ser tratados como entidades reales y que la idea de onda electromagnética simplemente no era algo que funcionara en todas las circunstancias. Extendiendo la analogía del cajero automático, el dinero de hecho está cuantificado; así, la unidad de moneda básica es el penique y nunca se puede tener en efectivo una cantidad de, por ejemplo, 241'378 peniques, sólo 241 peniques o 242 peniques. Lo que importa es que la unidad monetaria (el «cuanto» que en este caso es el penique) es tan pequeña que es posible tener cualquier cantidad de dinero en una aproximación bastante buena. Siguiendo este ejemplo, a muy pocos adultos les preocuparía, o pocos sabrían, si tienen en sus bolsillos realmente £2'41 o £2'42.

En el contexto de los fotones, la luz también es así. Hay tantos fotones, contribuyendo cada uno de ellos al haz de luz con su pequeña cantidad de energía, que se unen todos y consiguen dar la apariencia de un flujo de radiación uniforme y continuo. ¿Que cuántos son? Alrededor de un billón (10^{12}) de fotones de luz solar caen sobre una cabeza de alfiler cada segundo en un día soleado; cuando estamos mirando a una estrella pálida, el ojo recibe unos pocos cientos de fotones de la estrella cada segundo.

El descubrimiento de que la luz se comporta como si fuera un chorro de partículas produjo una enorme confusión durante cierto tiempo, y además desconcierto entre los físicos, que poco a poco aprendieron a pensar que entidades tales como la luz eran al mismo tiempo partículas y ondas, que unas veces mostraban una de las dos caras y otras veces la otra. Una entidad cuántica, como es un fotón, no es ni una partícula ni una onda. Es algo que no podemos entender si pensamos en alguna cosa relacionada con nuestra experiencia cotidiana, excepto si explicamos su comportamiento en algunas circunstancias utilizando la analogía de la onda para describir su carácter, y en otras circunstancias utilizando la analogía de la partícula para describir su comportamiento. Hagamos lo que hagamos, es mejor que no perdamos el tiempo tratando de comprender qué «es realmente» una entidad cuántica como el fotón, ya que nadie sabe qué es, solamente a qué se parece. Citando de nuevo a Feynman:

Si lo puede evitar, no siga diciéndose a sí mismo: «Pero ¿cómo es posible que sea así?», porque «se irá por el desagüe» para aterrizar en un callejón sin

salida del que nadie ha podido aún escapar. Nadie sabe cómo es posible que sea así.[2]

Sin embargo, todo esto está yendo un poco por delante de la versión del modelo del átomo de Bohr dada en 1913, que sólo necesitaba completarse con la idea de la partición en cuantos que existe en la emisión y absorción de la radiación (no necesariamente de la partición en cuantos de la propia radiación) para resolver el misterio de por qué los electrones de los átomos no caen dentro del núcleo.

Bohr desarrolló el modelo planetario de Rutherford para el átomo diciendo que los electrones sólo pueden ocupar ciertas «órbitas estables» alrededor del núcleo, cada una de ellas correspondiente a una cierta cantidad fija de energía, un múltiplo del cuanto básico. Sin embargo, lo que no podía haber eran órbitas intercaladas, porque corresponderían a cantidades fraccionarias de energía. Un electrón podría saltar de una órbita a otra, emitiendo un cuanto de energía si la nueva órbita estaba más cerca del núcleo o absorbiendo un cuanto de energía si el salto le llevaba más lejos del núcleo. Pero no podía describir espirales continuamente en dirección al núcleo.

Entonces ¿por qué no saltaban todos los electrones con su carga negativa directamente sobre el núcleo, atraídos por su carga eléctrica positiva? Bohr añadió otro ingrediente a su modelo, argumentando que cada órbita estable alrededor del núcleo, en cierto sentido, sólo tenía espacio para un número limitado de electrones. Si la órbita estaba llena, entonces no importaba cuántos electrones podría haber en órbitas de más energía, porque no podrían desprenderse de ese exceso de energía y bajar de un salto a la órbita que ya estaba ocupada. Del mismo modo, estaba sencillamente prohibido a los electrones de la órbita de energía mínima dar el salto final hasta el núcleo. Sin embargo, si una órbita inferior tenía espacio para él, un electrón de una órbita con más energía podía dar un salto y bajar a esa órbita inferior, radiando un cuanto de energía (lo que hoy llamaríamos un fotón correspondiente a una determinada longitud de onda de luz) al hacerlo. Para saltar de una órbita de menos energía a otra de más energía, un electrón tendría que absorber precisamente un cuanto con la cantidad adecuada de energía, la misma cantidad de energía que irradiaría si cayera de nuevo a una órbita inferior.

Se podría ilustrar con una comparación diciendo que es casi como si cada electrón estuviera sentado en un escalón de una escalera y sólo pudiera subir o bajar saltando un número entero de escalones, porque no hay escalones intermedios donde pueda quedarse; no obstante, la comparación es complicada porque los escalones no están todos a la misma altura.

Parecía natural suponer que cada átomo de un elemento concreto se comporta de la misma manera, teniendo sus electrones el mismo espacio entre los escalones de la

escalera de niveles energéticos, ya que los átomos de un mismo elemento (hablando con propiedad, de los isótopos de un mismo elemento) son idénticos uno a otro. Por lo tanto, lo que realmente vemos al mirar a la luz que irradia un objeto caliente es el efecto combinado de todos los pequeños paquetes de luz, producido por los electrones que saltan entre los niveles de una amplia gama de escalones con idéntica energía.

Lo que se ve realmente en el espectro de la luz que emite un objeto caliente es una serie de líneas brillantes, con longitudes de onda claramente determinadas. Estas longitudes de onda se podrían definir como las correspondientes al cuanto de energía irradiado cuando un electrón hace su salto hacia abajo en la escalera de energías (una transición específica desde un nivel de energía a otro inferior). La pauta de estas líneas del espectro es diferente para cada elemento y

actúa como una huella dactilar única que muestra la presencia de un elemento concreto en el objeto caliente. Por ejemplo, el gas hidrógeno caliente produce su propia huella espectral distintiva, diferente de la huella de cualquier otro elemento, incluido su pariente atómico más próximo, el helio. También si dirigimos la luz a través de un gas frío, se observa la misma pauta de líneas oscuras en el espectro allí donde la energía ha sido absorbida mientras la luz pasaba a través del gas, siendo utilizada esta energía para hacer que los electrones salten de un nivel de energía a otro superior.[3]

El poder de la espectroscopia para identificar la presencia de distintos elementos se descubrió en el siglo XIX y fue desarrollado por investigadores tales como Joseph Fraunhofer, Gustav Kirchhoff y Robert Bunsen. El famoso «mechero de Bunsen» (que realmente no fue inventado por Bunsen, aunque éste hizo un buen uso de él) fue un instrumento clave para su trabajo. Cuando una sustancia se calienta en la llama clara del mechero (por ejemplo, cuando un cable se baña en una sustancia en polvo o en un líquido y luego se pone sobre la llama) produce una luz con un color característico, producido por las líneas brillantes del espectro en las longitudes de onda características —es decir, colores— del elemento que se está estudiando.

Uno de los ejemplos cotidianos que resultan más familiares en relación con este tipo de procesos es el típico color naranja amarillento de muchas farolas del alumbrado urbano. Dicho color tiene su causa en la presencia de sodio. En este caso, es la energía eléctrica que pasa a través del gas en la luz la que mueve los electrones en los átomos de sodio llevándolos a un nivel de energía superior. Cuando vuelven a caer en la escala de energías, emiten una luz que tiene unas longitudes de ondas definidas con mucha precisión y que forma dos líneas brillantes en la parte amarilla del espectro. El característico amarillo de sodio se ve si, por ejemplo, la sal ordinaria (cloruro de sodio) se calienta en la llama de un mechero de Bunsen, o sencillamente si la sal se arroja sobre el fuego.

Cada elemento tiene su propia pauta de líneas espectrales y en cada caso la pauta se mantiene igual (aunque la intensidad de las líneas se modificará) incluso si cambia la temperatura. Comparando sus estudios de laboratorio sobre los espectros con las líneas observadas en la luz procedente del Sol y de las estrellas, los técnicos en espectroscopia pudieron explicar la mayoría de estas líneas hablando de la presencia en el Sol y en las estrellas de elementos conocidos en la Tierra (esta clase de estudios comenzó en 1859, cuando Kirchoff descubrió que existe sodio en la atmósfera del Sol).

En un famoso procedimiento inverso al mencionado, el astrónomo británico Norman Lockyer explicó las líneas del espectro solar que no correspondían a ningún elemento conocido en la Tierra, atribuyéndolas a un elemento desconocido, al que llamó helio (según el nombre griego del Sol, helios). El helio se descubrió posteriormente en la Tierra y se vio que tenía exactamente el espectro requerido para coincidir con esas líneas solares específicas.

El espectro del hidrógeno es especialmente sencillo; hoy en día sabemos que esto se debe a que el átomo de hidrógeno está formado por un único protón asociado a un solo electrón. Las líneas del espectro que constituyen la huella única del hidrógeno llevan el nombre de Johann Balmer, un maestro suizo que desarrolló en 1884 una fórmula matemática para describir esta pauta y la publicó en 1885. La fórmula de Balmer era tan sencilla que contenía claramente una profunda verdad en relación con la estructura del átomo de hidrógeno. Pero nadie lo supo, hasta que Bohr entró en escena.

Aunque había hecho algo de espectroscopia antes de terminar la carrera, Bohr no era un experto en esta técnica, por lo que, cuando empezó a trabajar en el enigma de la estructura del átomo de hidrógeno, no se le ocurrió inmediatamente que la serie de Balmer era la clave con la que podía desentrañar el misterio. No llegó a apreciar su importancia hasta que un colega le indicó lo sencilla que era realmente la fórmula de Balmer. Esto sucedió en 1913 y fue lo que condujo directamente al modelo de Bohr del átomo de hidrógeno, en el que hay un solo electrón que puede saltar de un nivel de energía a otro. El espacio existente entre los niveles de energía depende del valor de la constante de Planck, y ésta a su vez determina el espacio entre las líneas de la serie de Balmer. Por lo tanto, utilizando el espaciamiento observado entre las líneas de la serie, Bohr pudo calcular el espacio existente entre los niveles de energía, y así la fórmula de Balmer se pudo reescribir, de un modo muy natural, para que incluyera la constante de Planck. Fue este trabajo de Bohr, inspirado en la física del siglo XIX, el que demostró de manera definitiva que el átomo de hidrógeno contiene exactamente un electrón.

El modelo atómico de Bohr explicaba tanto la aparición de líneas brillantes en el espectro de emisión de un elemento, como la presencia de líneas oscuras en las

mismas longitudes de onda en su espectro de absorción. Desde luego, este modelo funcionó. Con cada salto de una órbita a otra correspondiente a un nivel específico de energía, y por consiguiente a una longitud de onda de luz determinada, este modelo explicaba el espectro de luz emitido por el átomo más sencillo, el hidrógeno. Pero la constante de Planck es muy pequeña. En el mismo sistema de unidades en que las masas se miden en gramos, el valor de la constante de Planck es 6.55×10^{-27} . Los efectos cuánticos sólo llegan a ser importantes para aquellas partículas cuyas masas en gramos son aproximadamente iguales (o menores) que este número. La masa de un electrón es 9×10^{-31} gramos, por lo que los efectos cuánticos son enormemente importantes para los electrones. Estos efectos se hacen cada vez menos relevantes a medida que los objetos aumentan de tamaño, por lo que cualquier objeto mucho mayor que un átomo no estará profundamente afectado por procesos cuánticos (excepto, por supuesto, en el sentido de que está constituido por átomos y éstos sí que están afectados por procesos cuánticos).

Es difícil hacerse una idea de lo pequeño que es el mundo de los cuantos; en cierto sentido, de lo lejos que está de nosotros. Pero vamos a intentar lo siguiente: si un objeto tuviera un diámetro de 10^{27} cm, harían falta 10^{27} de estos objetos para completar un solo centímetro, que viene a ser la arista de un terrón de azúcar de forma cúbica. Este número, 10^{27} , es igual a aproximadamente 10^{16} veces el número de estrellas brillantes que hay en la Vía Láctea. Es el número que obtendríamos si tomáramos el número de estrellas de esta galaxia (10^{11}), multiplicado por sí mismo (daría 10^{22}) y luego multiplicáramos este cuadrado por 100.000. Si tomamos 10^{27} terrones de azúcar y los colocamos en una hilera, tocándose entre sí, recorrerían una distancia de mil millones de años luz, aproximadamente un décimo del tamaño del universo conocido. En números redondos, la escala en la que predominan los procesos cuánticos es mucho menor que un terrón de azúcar, en una proporción aún menor que aquella en que un terrón de azúcar es menor que el universo. En este tipo de aproximación unas pocas potencias de diez no suponen gran cosa, por lo que podemos afirmar que tanto un terrón de azúcar como un ser humano se quedan más o menos en la mitad de la escala de tamaños, a medio camino entre la escala de los cuantos y la escala del universo, si se mide de esta manera logarítmica («potencias de diez»).

Así pues, aunque el modelo de Bohr funcionaba (especialmente como explicación del origen de las líneas espectrales), no sorprenderá saber que en los primeros intentos de abordar las propiedades del mundo cuántico pronto se necesitó mejorar dicho modelo, a medida que iban surgiendo nuevos descubrimientos. El más importante de éstos se refería a la naturaleza de los electrones, que, como ya hemos explicado, son las entidades cuánticas arquetípicas, ya que tienen justo la masa adecuada para que les afecten profundamente aquellos procesos en los que hay que

tener en cuenta la constante de Planck.

Durante el primer cuarto del siglo xx surgieron algunos descubrimientos asombrosos relativos al comportamiento de los electrones. Se pueden resumir diciendo que afectan al comportamiento de los electrones dentro de los átomos. El modelo de Rutherford-Bohr para el átomo, planteado como un sistema solar en miniatura, en el que los electrones actúan como diminutas bolas que zumban por allí dando vueltas al modo de los planetas, funcionaba muy bien por lo que respecta a explicar la espectroscopia (siempre que se tuviera en cuenta el comportamiento cuántico de los electrones saltando entre los distintos niveles de energía). Pero los propios átomos seguían comportándose como esferas rígidas y lisas cuando, por ejemplo, funcionan en conjunto para producir la presión que un gas ejerce sobre las paredes de un recipiente. Es difícil ver cómo unos pocos electrones que describen órbitas en torno al núcleo se podrían combinar para darle al átomo la apariencia de una superficie lisa y rígida.

Dándole la vuelta a la analogía, si una estrella que tiene otro sistema planetario alrededor de ella se aproxima a nuestro propio sistema solar, no rebotaría cuando los planetas más exteriores de ambos sistemas se pusieran en contacto. La estrella visitante surcaría directamente el espacio a través del sistema solar, lanzando a los planetas fuera de su camino y, si llevara la trayectoria adecuada para ello, acabaría chocando con el propio Sol. Del mismo modo, podemos imaginarnos uno o dos, o incluso una docena, de pequeños electrones que describen órbitas alrededor del núcleo de un átomo, creando una barrera eficaz cuando los átomos chocan. Esto podría funcionar en el caso del uranio, que tiene 92 electrones moviéndose sin parar. Sin embargo, incluso el hidrógeno se comporta como si su núcleo estuviera totalmente rodeado por su único electrón, existiendo una distribución uniforme de carga eléctrica negativa que hace todo lo que puede para que la carga positiva quede confinada en el núcleo. ¿Cómo puede ser esto?

La respuesta comenzó a surgir a principios de los años veinte, cuando el físico francés Louis de Broglie observó que se podía invertir la ecuación que utiliza la constante de Planck para relacionar la longitud de onda de la luz con una energía equivalente para una partícula (el fotón). En esencia se podía utilizar la misma ecuación para tomar la energía de una partícula (tal como el electrón) y obtener una longitud de onda equivalente para la entidad cuántica que correspondiera a dicha partícula. De Broglie sugirió en su tesis doctoral que los electrones podrían tener que tratarse como si fueran ondas, en determinadas circunstancias, del mismo modo que la luz ha de ser tratada como un chorro de partículas en algunas circunstancias concretas.

El supervisor de la tesis de De Broglie, Paul Langevin, no fue capaz de decidir si esto era la obra de un genio o una auténtica basura, por lo que envió la tesis a Albert

Einstein para que éste la valorara. Acertó plenamente. Einstein había estado precisamente participando en el trabajo definitivo que estableció que la luz tiene que ser tratada realmente como fotones, y que la razón por la cual los átomos sólo pueden emitir o absorber luz en ciertas cantidades se debe a que la luz, de hecho, sólo existe en paquetes de un determinado tamaño, los cuantos. La luz aparece en cuantos, de una forma análoga al modo en que el dinero en efectivo se cuantifica en peniques, y resulta que un electrón y un átomo sólo pueden emitir o absorber un fotón cada vez. De manera que Einstein ya estaba trabajando en una dirección que llevaba a la idea de que todas las entidades cuánticas podrían ser tanto partículas como ondas, por lo que la respuesta que envió a Langevin fue que la tesis de De Broglie tenía que tomarse muy en serio.

La idea de De Broglie, combinada con la creciente evidencia de la realidad de los fotones, llevó a los físicos en la segunda mitad de los años veinte a desarrollar una completa teoría de la física cuántica (denominada a menudo también mecánica cuántica) basada en el concepto de la dualidad onda-partícula, lo cual significa que todas las entidades cuánticas comparten la misma esquizofrenia que tiene la luz. En principio esta dualidad se aplica a todo; así, existen, por ejemplo, características ondulatorias asociadas a la mesa ante la cual estoy sentado, o a mi propio cuerpo. Sin embargo, debido a que el peso (hablando más estrictamente, la masa) de la mesa, o de mi cuerpo en gramos es muchísimo mayor que 6×10^{27} , la longitud de onda correspondiente es tan pequeña que se desvanece y se puede despreciar. No le veo a mi mesa ningún borde rizado producido por ningún fenómeno ondulatorio. Solamente para cosas tales como electrones y fotones sucede que las dos facetas de la realidad tienen la misma importancia.

Entre otras cosas, esto resuelve el enigma de cómo un átomo que contiene unos pocos electrones puede mostrar al mundo una superficie lisa y comportarse como una pequeña esfera rígida. Todo electrón de la nube que rodea al núcleo tiene que ser considerado, no como una pequeña bola rígida, sino como una onda que se propaga alrededor de todo el núcleo. El único electrón del átomo de hidrógeno forma en realidad el solo toda una nube esférica alrededor del núcleo. Los electrones que se encuentran en niveles superiores de energía pueden constituir nubes con formas más complejas, algunos de ellos como lóbulos situados a ambos lados del núcleo y que recordarían a unas pesas de halterofilia. Pero el efecto general siempre es que el núcleo se encuentra rodeado completamente por su nube de electrones.

En un mismo átomo las nubes de electrones correspondientes a distintos niveles de energía pueden, hasta cierto punto, interpenetrarse, por lo que la mejor imagen del átomo no es algo como un sistema solar con los electrones en órbitas individuales, sino más bien algo parecido a una serie de capas como la cebolla, nubes de carga eléctrica que rodean completamente al núcleo y que están situadas bastante cerca las

unas de las otras. Entonces el cambio en los niveles de energía cuando un átomo absorbe o emite un fotón puede imaginarse, no tanto como una pelota botando arriba y abajo por una escalera, sino más bien como las diferentes notas que se pueden obtener de una sola cuerda de guitarra haciéndola vibrar en diferentes armónicos.

En 1927 se proclamó la nueva física cuántica, cuando dos equipos de investigadores, uno de ellos encabezado por George Thomson (el hijo de J. J.), llevaron a cabo unos experimentos que mostraban sin ambigüedades unos electrones cuyo comportamiento al saltar fuera de los átomos en un cristal era como el de las ondas. Thomson había recibido el Premio Nobel por descubrir el electrón y por demostrar que es una partícula. En 1937, George Thomson recibió, junto con otros, un Premio Nobel por demostrar que los electrones son ondas. Ambos tenían razón y ambos premios estuvieron bien merecidos. No hay nada que demuestre mejor las peculiaridades del mundo cuántico y la necesidad de recordar que los modelos son sólo ayudas para la imaginación, no la verdad última. El modo más claro de pensar sobre esta dualidad onda-partícula es decir que una entidad cuántica tal como un fotón o un electrón viaja como una onda, pero llega como una partícula. Cuando se les deja actuar por sus propios mecanismos, los objetos tales como los electrones se propagan como ondas; sin embargo, cuando se mide su posición, se concentran en un instante en un lugar determinado, en un fenómeno que se denomina «el colapso de la función de ondas». Luego comienzan de nuevo a propagarse hasta que tienen una interacción con alguna cosa.

La nueva física cuántica, construida sobre la idea de dualidad onda-partícula, se centra en torno a uno de los más famosos (pero no siempre uno de los mejor comprendidos) conceptos de la física: el principio de incertidumbre, descubierto por Werner Heisenberg a finales de 1926.

Una característica de las ondas es que son algo que se propaga y nunca se pueden concentrar literalmente «en un punto». Lo mismo se puede decir de los electrones. Se pueden situar bastante ajustadamente —por ejemplo, cuando uno de los electrones del haz que dibuja la imagen en la pantalla de nuestro televisor choca con dicha pantalla y forma una diminuta salpicadura de luz— pero nunca en lo que los matemáticos llaman un punto, un lugar de dimensión cero. El carácter de onda de un electrón significa que siempre hay una cierta incertidumbre acerca de su posición, y esto mismo se aplica a todas las entidades cuánticas. Heisenberg descubrió una relación matemática (relacionada, una vez más, con la constante de Planck) que expresa esta incertidumbre. Resulta que para una entidad cuántica, tal como un electrón, cuanto más se delimita su posición (cuanto más ajustadamente se fija su onda), menos certeza existe sobre dónde va a ir a continuación (de hecho, sobre su velocidad). Nos podemos imaginar esto un poco como cuando apretamos un muelle entre el índice y el pulgar: después saltará en alguna dirección impredecible.

Recíprocamente, cuanto mayor sea la precisión con que se determina la velocidad de una entidad cuántica, menos certeza existe sobre su posición: se hace más «ondulante». (El muelle que salta de nuestros dedos se mueve en una dirección determinada, pero se ha expandido y se ha estirado hasta su máxima longitud, no se concentra en un punto.)

Lo que es extremadamente importante en relación con todo esto es que no se debe a ninguna deficiencia en los aparatos de medición, y tampoco es que los seres humanos seamos demasiado ineptos para poder medir exactamente la posición y la velocidad de un electrón al mismo tiempo. La incertidumbre es una propiedad inherente a la propia naturaleza de los objetos cuánticos. Un electrón no tiene al mismo tiempo una posición exactamente determinada y una velocidad exactamente definida; no puede, por sí mismo, «saber» simultáneamente dónde está y adónde va a ir.

Esto tiene unas implicaciones muy intrigantes en el mundo de la física de partículas, que discutiremos en el próximo capítulo. Pero en el mundo de los átomos representa una implicación de la máxima importancia, que coloca la última pieza en el rompecabezas que es el modelo atómico de Rutherford-Bohr. La incertidumbre cuántica afirma que el «paquete ondulatorio» que define a un electrón siempre debe superar un determinado tamaño (en cierto modo, igual que existe un límite en relación con cuánto se puede comprimir un muelle). Resulta que, en las condiciones habituales, este tamaño es mayor que el del núcleo de un átomo. Ésta es la razón por la cual ninguno de los electrones de un átomo puede caer dentro de su núcleo: no existe allí espacio suficiente para que quepa su función de ondas. Dos electrones pueden caber en el nivel de energía más próximo al núcleo, ocho en el nivel siguiente de energía contando hacia el exterior, ocho más en el siguiente nivel, y así sucesivamente (esto tiene unas implicaciones en la química que comentaremos en el capítulo 4).

En el otro extremo del átomo, en su parte exterior, el comportamiento de los electrones que acabamos de describir claramente explica aquellas características del átomo que Feynman consideró tan importantes y tan reveladoras. El núcleo es una bola de carga eléctrica positiva y está rodeada por una nube de carga eléctrica negativa que, globalmente, equilibra y anula con exactitud la carga positiva del núcleo.

Ahora imaginemos cómo lo ve (o cómo lo siente) otro átomo idéntico que se sitúa a un lado del primer átomo. La nube de electrones del exterior del segundo átomo sólo puede ver la mitad de la nube de electrones del exterior del primer átomo, que sirve de escudo a la mitad de la carga positiva que hay en el interior. En consecuencia, la nube percibe que la mitad de la carga del núcleo del primer átomo ejerce de ese modo una fuerza de atracción sobre ella (y, por lo tanto, sobre la

totalidad del segundo átomo). Del mismo modo, la nube de electrones del primer átomo percibe una atracción ejercida por el núcleo parcialmente protegido del segundo átomo (los átomos se atraen entre sí cuando se encuentran a poca distancia). Sin embargo, cuando los dos átomos se acercan tanto que sus nubes de electrones están de hecho tocándose, la carga negativa de estas nubes, el mismo tipo de carga en cada nube, producirá una repulsión entre ambos (los átomos se repelen cuando se sitúan apretados uno contra el otro). No obstante, en todo este proceso los dos núcleos no perciben nada el uno del otro, porque están protegidos mutuamente por una doble capa de electrones.

En palabras de la vida cotidiana, lo que sucede en zonas profundas del interior del átomo, dentro del propio núcleo, no tiene en la práctica ninguna importancia. El modo en que funciona la atracción entre átomos y cómo la disposición de sus electrones favorece que algunos de ellos se unan formando moléculas tiene una importancia decisiva en la vida cotidiana, y explica cómo hemos llegado nosotros a estar aquí en primera fila. Pero esto lo explicaremos más adelante, en este mismo libro. Desde luego, si usted no tiene un interés especial por saber cómo funcionan las cosas a niveles tan profundos, ya sabe bastante como para saltar al capítulo 4 y averiguar cómo se unen los átomos para formar moléculas, incluidas las moléculas de los seres vivos. Sin embargo, una vez que se ha profundizado hasta el nivel en que suceden las cosas en el interior de los átomos, y después de la introducción a los conceptos de la física cuántica, sería una lástima marcharse sin llegar a vislumbrar lo que está pasando en los niveles más profundos estudiados hasta la fecha por los físicos: allí donde gobiernan las partículas, los campos y los efectos cuánticos.

Partículas y campos

Cuando los físicos quieren describir el modo en que el mundo funciona en su Cuando más fundamental, la componente más importante de su caja de herramientas es el concepto de campo. Este concepto es también una de las ideas que resultan más familiares dentro de la física que estudiábamos cuando íbamos a la escuela, cuando todos vimos cómo se puede hacer visible el campo de fuerzas alrededor de una simple barra imantada. Se sostiene el imán directamente bajo una hoja de papel en la que se han diseminado limaduras de hierro o algún otro material magnético. Dándole un golpecito suave a la hoja, las limaduras de hierro se colocan en arcos que unen los polos norte y sur del imán. Fue precisamente el estudio del magnetismo (y de la electricidad) lo que llevó al pionero del siglo XIX Michael Faraday a introducir el concepto de campo en la física.

La historia de Faraday suena a ciencia ficción. Era hijo de un pobre herrero y sólo recibió la instrucción más elemental, pero se hizo aprendiz en el taller de un encuadernador de Londres y se quedó enganchado a las ciencias cuando leyó la entrada correspondiente a la palabra «ciencia» en la Encyclopedia Britannica que estaba reencuadernando para un cliente. Sus loables esfuerzos de autodidacta se vieron recompensados en 1813, cuando a los veintiún años de edad obtuvo un puesto como ayudante de Humphry Davy en la entonces recién creada Royal Institution. Éste era el escalón más bajo que se podía ocupar en la escalera de la ciencia e incluía, entre otras cosas, tareas tales como lavar botellas. Pero Faraday ascendió, llegando a suceder a Davy como director de dicha institución, a declinar la oferta de la reina Victoria de concederle una orden nobiliaria y a rechazar dos veces la presidencia de la Royal Society. Era un hombre modesto, miembro de una estricta secta religiosa y no aprobaba los honores personales de este tipo.

Aunque comenzó como químico, a Faraday le intrigó la naturaleza de la electricidad y del magnetismo, que a principios de la década de 1820 empezaban a ser reconocidos como diferentes aspectos de un solo fenómeno. El experimento clave en este terreno lo realizó Hans Christian Oersted, en Dinamarca, y descubrió que, cuando se coloca una brújula magnética cerca de un cable que transporta una corriente eléctrica, la aguja de la brújula, que es un pequeño imán de barra, se desvía siempre, apuntando en una dirección que forma ángulo recto con el cable. De esta manera, se observa que una corriente eléctrica produce una fuerza magnética que influye en la aguja de la brújula.

El avance era lento y se vio obstaculizado por varias iniciativas erróneas. A finales de 1831 fue Faraday quien aclaró la naturaleza del vínculo existente entre electricidad y magnetismo. Una corriente eléctrica que fluye por un cable está causada por cargas eléctricas (sabemos actualmente que son los propios electrones)

que se desplazan a lo largo del cable. Una carga eléctrica en movimiento siempre crea una fuerza magnética (la que, en el caso mencionado, desvía la aguja de la brújula). Faraday descubrió que, de un modo similar, cuando movemos un imán por delante de un cable, o lo introducimos en el interior de una bobina de cable, mientras el imán se mueve se crea una corriente eléctrica en el cable. Un imán en movimiento siempre crea una fuerza eléctrica (que, en este caso, hace que los electrones que hay dentro del cable se muevan).

Estos dos descubrimientos constituyeron la base del generador eléctrico, o dinamo (donde la electricidad se produce haciendo girar imanes que se mueven a gran velocidad ante bobinas de cable) y el motor eléctrico (en el que una corriente eléctrica que pasa por bobinas de cable hace que rote un conjunto de imanes unidos al árbol del motor).

En 1831, poco después de que Faraday hubiera descubierto el efecto dinamo, el primer ministro Robert Peel visitó la Royal Institution y vio una demostración del fenómeno. En uno de los diálogos más ampliamente citados de la historia de la ciencia, el primer ministro preguntó a Faraday para qué servía su descubrimiento. Faraday respondió: «No lo sé, pero estoy seguro de que un día su gobierno le aplicará impuestos».

No es necesario explicar aquí la importancia práctica de la obra de Faraday, que desde luego ha aportado una rica fuente de ingresos a los gobiernos. A los cincuenta años de hacer Faraday ese comentario, los trenes eléctricos ya circulaban por Alemania, Gran Bretaña y Estados Unidos. Aunque no hubiera hecho nada más, Faraday (que tenía cuarenta años en 1831) habría sido recordado como uno de los científicos más importantes del siglo XIX. Sin embargo, después de 1831, en vez de desarrollar las aplicaciones prácticas de sus descubrimientos (tarea que dejó a otros), Faraday se interesó profundamente por la naturaleza del vínculo existente entre la electricidad y el magnetismo, y por cómo estas fuerzas podían atravesar el espacio consiguiendo influir en objetos con los que no tenían contacto directo. Desde el principio observó la similitud de este alcance con el modo en que actúa la gravedad, por cuyo efecto el Sol es capaz de llegar a través del espacio a mantener los planetas en sus órbitas.

En el experimento de las limaduras de hierro dispersas en una hoja de papel, las líneas que dibujan las limaduras muestran el camino que seguiría un diminuto polo magnético si fuera libre para recorrer una trayectoria que va desde un polo del gran imán hasta el otro. Esto indujo a Faraday a pensar en términos de «líneas de fuerza» (un término que inventó en 1831) que se propagaban desde cada polo del imán y, de forma similar, desde cada partícula con carga eléctrica. Según este modelo, si un cable se mueve en relación con un imán, el cable corta las líneas de fuerza, y es este corte lo que hace que fluya la corriente eléctrica.

Hasta ese momento, los científicos se imaginaban que fuerzas como la gravedad operaban como una «acción a distancia». No se consideraba que la fuerza que el Sol ejerce sobre la Tierra y los otros planetas tardara tiempo en llegar a través del espacio, sino que se pensaba que actuaba de forma instantánea y a distancia, para mantener su atracción sobre los planetas. Sin embargo, Faraday reflexionó sobre cómo podrían propagarse las líneas de fuerza y construyó lo que llegó a llamarse un campo de fuerzas, que es lo que se produce cuando una corriente se pone a circular por una bobina de cable.

Según su razonamiento estaba claro que esto llevaría tiempo. Una bobina de cable que esté dispuesta adecuadamente y transporte una corriente eléctrica produce un campo magnético que, a cierta distancia, no se puede distinguir del campo que produce un imán de barra. Pero, cuando no hay ninguna corriente que circule por la bobina, tampoco hay campo magnético. Faraday estaba convencido de que, cuando se pusiera en marcha la corriente en la bobina, el campo magnético se propagaría hacia el exterior a partir de dicha bobina, de tal modo que se desviarían en primer lugar las agujas de las brújulas que estuvieran cerca de ella, y después, cuando el campo magnético se propagara hacia el exterior, las agujas de las brújulas que estuvieran más alejadas. Sabemos actualmente que esta influencia se propaga a la velocidad de la luz, demasiado rápido como para que se pudiera medir el intervalo de tiempo con los medios de un laboratorio del siglo XIX. Sin embargo, Faraday seguía una línea de razonamiento correcta.

No obstante, la idea resultaba tan atrevida en aquella época, que Faraday no la hizo pública al principio. En vez de eso, en 1832 dejó una nota sellada en una caja fuerte de la Royal Society, con el encargo de que no se abriera hasta después de su muerte. Entre otras cosas, la nota exponía la teoría de Faraday de que cuando un imán actúa sobre otro imán o sobre un trozo de hierro que se encuentran a cierta distancia, la causa influenciante (que, por el momento, voy a llamar magnetismo) opera gradualmente desde los cuerpos magnéticos y requiere un tiempo para su transmisión ... Estoy inclinado a pensar que la difusión de las fuerzas magnéticas desde un polo magnético se puede comparar con las vibraciones que se producen en la superficie del agua cuando hay una perturbación, o con las del aire en los fenómenos relacionados con el sonido: es decir, estoy inclinado a pensar que la teoría vibratoria es aplicable a estos fenómenos, como lo es al sonido y, con toda probabilidad, también a la luz.

Doce años más tarde Faraday hizo públicas estas teorías, que prepararon el camino para los trabajos que James Clerk Maxwell llevó a cabo en la década de 1860 sobre la naturaleza del electromagnetismo y la luz.

Maxwell halló un conjunto de ecuaciones que describen cómo los campos eléctrico y magnético producen interacciones mutuas. Existen en total cuatro ecuaciones, que juntas pueden decirnos todo lo que podríamos necesitar por lo que se

refiere al electromagnetismo, salvo que nos decidamos a aventurarnos en el terreno de la física cuántica. Todo fenómeno en el que intervengan la electricidad y el magnetismo se puede resolver a nivel de la física clásica utilizando las ecuaciones de Maxwell; además, éstas llegaron con un plus incluido.

Entre otras cosas, las ecuaciones de Maxwell explican cómo se desplaza una onda electromagnética. Imaginemos una onda eléctrica que se desplaza por el espacio, dotada de energía, quizá, por un electrón que baila arriba y abajo en un cable. Cuando la onda eléctrica oscila subiendo y bajando, está experimentando variaciones. Por lo tanto tiene que crear una onda magnética, que se desplaza junto a la onda eléctrica. Pero esta onda magnética también experimenta variaciones, porque la onda eléctrica varía. En consecuencia la onda magnética debe crear una onda eléctrica que se desplazará junto a la onda magnética. El efecto global, descrito con precisión por las ecuaciones de Maxwell, es el de una onda electromagnética combinada que se desplaza a través del espacio de tal modo que las componentes eléctrica y magnética van al paso la una con la otra. El plus que llevan las ecuaciones de Maxwell es un número que especifica la velocidad a la que se desplaza la onda. Este número se puede calcular mediante experimentos en los que se miden las propiedades de campos eléctricos y magnéticos aislados. Y resulta que el número que se obtiene en estos experimentos es la velocidad de la luz, a pesar de que, como dijo Maxwell, «lo único para lo que se utilizó la luz en el experimento fue para poder ver los instrumentos».

Maxwell descubrió que las ondas electromagnéticas se desplazan a la velocidad de la luz y se dio cuenta de que la luz tenía que ser alguna clase de onda electromagnética. Esta conclusión se hizo inevitable cuando Heinrich Hertz, en la década de 1880, descubrió las ondas de radio, que Maxwell ya había predicho y que se desplazan también a la velocidad de la luz. Ésta es la razón por la que los físicos se sorprendieron tanto cuando, a principios del siglo xx, descubrieron que en el mundo cuántico la luz ha de ser tratada a veces como un chorro de diminutas partículas. Sin embargo, es posible que Faraday no se hubiera sorprendido tanto.

La primera vez que Faraday expresó en público sus teorías acerca de las líneas de fuerza y los campos fue en dos conferencias que dio en la Royal Institution en 1844 y 1846. Sus teorías resultaron avanzadas para su tiempo, ya que en aquella época todavía se aceptaba en amplios sectores que el «espacio vacío» estaba lleno de una misteriosa sustancia llamada éter, que transportaba los movimientos de las ondas luminosas, del mismo modo que las ondulaciones que se producen en una charca son transportadas por el agua de la charca. En su conferencia de 1846, hablando más o menos del modelo que acabamos de esbozar, pero sin las matemáticas de Maxwell, Faraday dijo explícitamente que su propósito era «descartar el éter, pero no las vibraciones», sugiriendo en cambio que las vibraciones estaban asociadas con las

líneas de fuerza, como cuerdas vibrantes de una guitarra. Indicó que la propagación de la luz tarda un tiempo, lo cual encaja con la idea de que una ondulación se desplaza a lo largo de una línea de fuerza, y formuló la conjetura de que la gravedad debe operar de un modo similar. Sin embargo, en algunos aspectos lo que dijo en la conferencia anterior, el 19 de enero de 1844, es aún más fascinante.

Además de descartar el concepto de éter, Faraday deseaba suprimir la idea de átomo (esto sucedía antes de que la idea de átomo hubiera sido aceptada por completo dentro del mundo de la ciencia). Tal deseo suena a locura, desde la perspectiva del éxito que ha tenido el modelo atómico que describíamos en el capítulo anterior. Sin embargo, desde la perspectiva de la teoría cuántica de campos desarrollada en el siglo xx, sería perfectamente razonable, si reemplazamos el significado decimonónico de la palabra «átomo» por el término moderno equivalente: «partícula». Esto constituye otro ejemplo de la necesidad de utilizar modelos diferentes cuando describimos la misma cosa en distintas circunstancias.

Faraday argumentaba que podía no existir una diferencia real entre el llamado espacio y los átomos («partículas») que se encuentran en el mismo. Decía que los átomos (partículas) deberían considerarse sencillamente como concentraciones de fuerzas y que, en vez de pensar en un átomo (partícula) como el origen de un tejido de fuerzas (el campo) que se despliegan alrededor de él, lo que deberíamos entender como realidad fundamental era el propio campo de fuerzas, considerando las partículas como simples concentraciones en las líneas de fuerza, es decir, nudos del tejido.

Para lograr sus propósitos, Faraday pidió a la audiencia que se imaginara un «experimento mental» en el que el Sol estaba a su libre albedrío en el espacio. ¿Qué le sucedería a la Tierra si la colocáramos de repente, como por arte de magia, en su posición, a la distancia del Sol que de hecho le corresponde? ¿Tendría que viajar alguna fuerza desde el Sol hacia el exterior y agarrar firmemente a la Tierra para que ésta empezara a sentir la gravedad del Sol? ¿O la Tierra sentiría el tirón gravitatorio del Sol inmediatamente?

Faraday dijo que, aunque el Sol estuviera solo en el espacio, su influencia gravitatoria se extendería por todas partes, incluso a través del lugar donde la Tierra iba a hacer su aparición. Nuestro planeta no aparecería en un «espacio vacío», sino en un tejido de fuerzas (el campo), y respondería instantáneamente a la naturaleza del campo y a la posición de la propia Tierra. Lo que a la Tierra le afecta es su posición y la naturaleza del campo en dicha posición, no la naturaleza de la fuente del campo (en este caso, el Sol) —siempre que, por supuesto, la idea de que el campo tiene una fuente fuera correcta—. Para Faraday el campo era la realidad, y la materia no era más que una zona donde el campo se concentraba, o hacía un nudo. En el transcurso de estas dos conferencias, descartó tanto la idea de existencia del éter, como la idea

de partículas con una realidad material, dejando una imagen del universo en la que éste era nada más (o nada menos) que un tejido de campos con interacciones mutuas, anudados aquí y allá. Es casi exactamente el modo en que un moderno experto en la teoría cuántica de campos describiría el universo.

El ejemplo clásico, desde luego, es el propio fotón. Un fotón puede considerarse como un nudo en un campo electromagnético, una pequeña maraña de ondas electromagnéticas. Pero recordemos que los físicos descubrieron en el siglo xx que incluso aquello que solemos considerar como partículas materiales, por ejemplo electrones, en su propia naturaleza manifiestan también un carácter ondulatorio. En la teoría cuántica de campos, cada tipo de partícula tiene su propio campo; así, por ejemplo, hay un campo de electrones que llena el universo. Lo que nosotros percibimos como partículas (electrones u otros) son nudos en el campo correspondiente, del mismo modo que los fotones son nudos del campo electromagnético.

Se hace extraño pensar en la existencia de un campo de electrones que llena el universo y que nosotros, de forma natural, no podemos percibir. Es aún más extraño pensar en la existencia de una clase distinta de campo para cada clase de partícula, así como en la existencia de los campos eléctrico y magnético, que nos resultan algo más familiares, y otros. Suena como si el universo estuviera atestado de campos en un confuso enmarañamiento (como de hecho lo está, según este modelo) y tuviéramos que ser conscientes de ello (y no lo somos). Sin embargo, recordemos que no notamos los campos electromagnéticos que impregnan el entorno; en este momento hay señales de docenas de emisoras de radio y televisión que pasan justo a través de nosotros.

Lo que sucede con todos estos campos de materia, como se les denomina, se parece bastante al modo en que percibimos —o, más bien, no percibimos—la atmósfera de la Tierra. En un día en que el tiempo es apacible no notamos que estamos inmersos en un mar de aire. Dado que es igual por todas partes, no hay forma de precisar que está ahí. El único caso en que percibimos el aire es cuando está en movimiento, y ese movimiento está producido por irregularidades que oscilan desde pequeños remolinos hasta grandes tormentas y huracanes. Las partículas materiales que percibimos —electrones y otras— son el equivalente a los remolinos, pero en los campos de materia y, al igual que éstos, su influencia se propaga a poca distancia en su entorno. Sólo percibimos las irregularidades, no la uniformidad lisa y suave de los campos que llenan el resto del universo.

Este modelo resulta muy práctico para comprobar los estratos más profundos del mundo cuántico; sin embargo, para lo que nosotros nos proponemos basta con saber que éste es el modo en que ven el mundo los expertos en teoría cuántica de campos, y para quedarnos impresionados por la perspicacia de Faraday. Si el propósito es

conseguir un esbozo de lo que sucede en lo más profundo de los átomos, podemos seguir trabajando con un modelo que es una mezcla de partículas y campos, y explica el comportamiento de estas entidades por analogía con el modo en que los electrones ejercen interacciones con los campos electromagnéticos, y mutuamente entre los propios electrones. El concepto de partículas sigue siendo práctico, tanto si se piensa en ellas como nudos de un campo como si se contemplan de otra manera.

Cuando dos electrones se acercan el uno al otro, se repelen entre sí, porque cada uno de ellos tiene la misma carga eléctrica (en este caso, negativa). Pero ¿cómo funciona esta repulsión? Siguiendo con la idea de los electrones como partículas y aplicando la teoría cuántica a los campos electromagnéticos que rodean a estas partículas, podemos conseguir una imagen clara de lo que está sucediendo. En la interacción entre los dos electrones participa un chorro de fotones (encargados de transportar la fuerza electromagnética) que se desplazan de un electrón a otro (y, desde luego, del segundo electrón al primero). Pensemos en estos fotones como en una ráfaga de balas de metralleta. Cada electrón emite una ráfaga de partículas y retrocede, mientras que, al mismo tiempo, cada electrón es alcanzado por una ráfaga de partículas que lo rechazan con sus golpes. No es de extrañar que los dos electrones se repelan mutuamente.

Es más difícil plantearse por qué las partículas de carga opuesta (por ejemplo, un electrón y un protón) han de atraerse unas a otras, pero así es precisamente como sucede. Una analogía que ayuda algo es pensar en un grupo de atletas que participan en un ejercicio de entrenamiento en el que van corriendo mientras se pasan del uno al otro un pesado balón de fisioterapia. Se tienen que mantener muy cerca unos de otros, porque no pueden arrojar este balón a mucha distancia. Sin embargo, a veces esta atracción funciona, en el mundo de las partículas, incluso cuando se intercambian partículas muy ligeras. Un chorro de fotones procedente de un electrón y que llega a un protón, no produce ningún empuje para alejar a este protón, sino que tira de él hacia el electrón, y viceversa.

Todo esto puede suceder a una distancia considerable, porque es fácil producir fotones. No tienen absolutamente nada de masa, por lo que, aunque transportan algo de energía, no existe energía másica implicada en el proceso, y así las partículas cargadas pueden producir chorros de fotones sin perder mucha energía. Una vez que han sido emitidos, los fotones pueden desplazarse, en principio, eternamente por el espacio. Por lo tanto, el electromagnetismo es una fuerza de muy largo alcance y puede, si nada lo impide, llegar a todas partes atravesando el universo (igual que lo hace la gravedad y por la misma razón: el gravitón tiene también masa cero). De hecho, el alcance de las fuerzas eléctricas y magnéticas está limitado por su tendencia a ser neutralizadas a una escala menor. Debido a que por cada protón hay un electrón en el átomo, no existe una carga eléctrica global que pueda ejercer su influencia en el

mundo de una manera extensa.

Las cosas difieren significativamente cuando descendemos a la escala del núcleo del átomo. Allí los protones y los neutrones se mantienen juntos por el efecto de otra fuerza, la poderosa fuerza nuclear, que impide que se separe toda la carga positiva concentrada en el núcleo. Sin embargo, esta fuerza no está relacionada directamente con los neutrones y los protones, sino con un estrato estructural más profundo que se encuentra dentro de ellos, al nivel de los quarks.

Existe todo tipo de pruebas de la existencia de los quarks, pero las mejores y más directas provienen de experimentos llevados a cabo a finales de la década de los años sesenta y durante los setenta, cuando se realizó el bombardeo de núcleos atómicos mediante ráfagas de electrones dotadas de gran cantidad de energía, para estudiar el modo en que los electrones rebotaban (se dispersaban) rechazando a las partículas de los núcleos. Por un procedimiento que recuerda sorprendentemente la forma en que el equipo de Ernest Rutherford había descubierto la estructura interna del átomo, dispersando partículas alfa que rebotaban en el núcleo, estos experimentos, de los que fueron pioneros los investigadores de la Universidad Stanford de California, mostraron la estructura interna de los protones y los neutrones, dispersando electrones que rebotaban en las partículas del núcleo.

Los protones y los neutrones están formados por quarks; en cada caso, tres quarks. Sin embargo, sólo se necesitan dos tipos diferentes de quarks para explicar la estructura de los protones y los neutrones. Se les han dado nombres elegidos arbitrariamente, cuya única función es hacer de etiquetas: un tipo de quark se llama «superior» y el otro recibe el nombre de «inferior». Pero esto no significa, en ningún caso, que apunten hacia direcciones diferentes, ni que se encuentren a distintas alturas. Sus etiquetas podrían haber sido igualmente José y Rita.

Cada quark superior lleva una cantidad de carga eléctrica que es positiva e igual a dos tercios de la carga de un electrón. Cada quark inferior contiene una cantidad de carga eléctrica que es negativa e igual a un tercio de la carga de un electrón. Un protón está formado por dos quarks superiores y uno inferior que juntos dan una unidad de carga positiva ($2/3 + 2/3 - 1/3 = 1$), mientras que cada neutrón está formado por dos quarks inferiores y uno superior, por lo que globalmente no tiene ninguna carga eléctrica ($2/3 - 1/3 - 1/3 = 0$).

La característica más intrigante que poseen los quarks es que nunca se encuentran aislados, sino solamente en tríos o parejas. Las parejas de quarks (de hecho, siempre una pareja quark-antiquark)[4] reciben el nombre de mesones, y son precisamente estos mesones los que mantienen unidos los núcleos de los átomos, siendo intercambiados constantemente entre los protones y los neutrones que constituyen el núcleo. Los mesones transportan una fuerza que se ejerce entre las partículas componentes del núcleo (los nucleones) de una manera análoga al modo en que los

fotones transportan la fuerza electromagnética. Sin embargo, a diferencia de los fotones, los mesones poseen masa. Por consiguiente, hace falta mucha energía para fabricar un mesón. La única razón por la que los mesones se pueden fabricar dentro de un núcleo viene dada por la entrada en escena de la incertidumbre cuántica.

Esta vez se trata de una importante incertidumbre que afecta a la cantidad de energía asociada con cada uno de los nucleones. Del mismo modo que no se pueden determinar con exactitud la posición y el momento de una entidad cuántica, tampoco se puede determinar con precisión la cantidad de energía asociada a la misma, y no por la imprecisión de los aparatos de medición, sino porque el propio universo no «conoce» con exactitud cuánta energía hay allí en un instante cualquiera. Si se considera un tiempo suficientemente breve (un tiempo relacionado con la constante de Planck, por lo tanto muy breve), la energía puede surgir de la nada, siempre que vuelva a desaparecer dentro del tiempo asignado. Cuanta más energía surja, menos tiempo puede permanecer. Sin embargo, si tenemos suficiente energía, ésta puede convertirse temporalmente en partículas durante su breve tiempo de vida.

De aquí es de donde surgen los mesones que mantienen unido el núcleo de un átomo. Surgen de la nada, como lo que se llama fluctuaciones de los campos cuánticos en el vacío. Cuando esto sucede, las reglas cuánticas dicen que cada partícula producida debe estar acompañada por su correspondiente antipartícula. En vez de quarks aislados que surgen del vacío, lo que tenemos son parejas de quark y antiquark. En otras palabras, mesones. Sin embargo, estos sólo pueden vivir durante un tiempo muy breve, el tiempo que permite la incertidumbre cuántica, hasta devolver su energía prestada al vacío y desaparecer de nuevo. Viven justo lo suficiente para ser objeto de un intercambio con el nucleón vecino, manteniendo así el núcleo unido (ésta es la razón por la que la analogía con el balón de fisioterapia puede ser útil). Sin embargo, la gama de valores de la fuerza nuclear resultante es tan limitada que nada que esté fuera del núcleo, ni siquiera el núcleo del átomo más próximo, puede llegar en ningún momento a sentir dicha fuerza.

Pero ¿cuál es la razón por la que los quarks sólo aparecen en tríos o parejas? Pues porque, a su vez, se mantienen juntos por el intercambio de otro tipo de partícula del campo cuántico. Cuando los físicos tuvieron que poner un nombre a estas partículas, se permitieron hacer una pequeña broma: las partículas se llaman gluones, ya que lo que hacen es pegar (glue) los quarks uno a otro. Los gluones son los cuantos que están asociados con una nueva clase de campo y actúan casi del mismo modo que todas las demás fuerzas que ya hemos mencionado. Sin embargo, existe una diferencia crucial. La fuerza de adherencia entre dos quarks se hace tanto más fuerte cuanto más alejados están uno de otro.

En vez de considerar el intercambio de gluones entre dos quarks como un chorro de partículas que los golpean para mantenerlos apartados, pensemos que son un trozo

de un elástico muy poderoso que une a los dos quarks. Cuando éstos se encuentran suficientemente cerca, el elástico se queda flojo y blando, con lo que los quarks empiezan a moverse como a empujones. Pero cuando los quarks intentan separarse (aunque sea sólo a la distancia que hay desde un extremo al otro en un protón), el «elástico» se estira y vuelve a ponerlos juntos. Cuanto más se aparten, más se estira el elástico y, por consiguiente, los empuja con más fuerza cuando tira de ellos para juntarlos.

El único modo en que un quark puede ser despachado de un nucleón o de un mesón es que haya algo que lo golpee tan fuerte —con tanta energía— que «rompa el elástico». En efecto, esto puede suceder cuando un electrón que se mueve casi a la velocidad de la luz choca contra un protón en un experimento como los que hemos descrito anteriormente. Sin embargo, esto sólo puede suceder si la energía aportada por el impacto del electrón es lo suficientemente grande como para convertirse en la masa de dos nuevos quarks (concretamente, una pareja de quark y antiquark). Cuando el «elástico» se rompe, cada extremo de dicho elástico roto se ata a uno de los dos nuevos quarks. Uno de los extremos se agarra al protón, restaurando así su estado normal. El otro extremo ata el nuevo antiquark al quark que se escapa; en vez de un único quark que se escapa del lugar del impacto, lo que vemos es un mesón. Nunca, en ninguna clase de circunstancias, vemos un quark solitario.

Por lo tanto, la fuerza de adherencia de los gluones es la fuerza real que mantiene juntos los componentes del núcleo, incluyendo a las parejas de quark y antiquark que forman mesones. La poderosa fuerza nuclear que actúa entre los nucleones es en realidad un vestigio relativamente débil de la fuerza de adherencia de los gluones, y ambas fuerzas no están consideradas como dos fuerzas de la naturaleza genuinamente distintas. Junto con el electromagnetismo y la gravedad, esto significa que hasta ahora hemos mencionado tres tipos de fuerzas genuinamente diferentes que actúan en el mundo de las partículas. Existe aún otra más que debemos mencionar: la llamada fuerza débil (porque es más débil que la fuerza nuclear). Ésta es la que menos se parece a lo que en la realidad cotidiana significa el término «fuerza». Reconociendo este hecho, los físicos se refieren a menudo a las cuatro fuerzas de la naturaleza llamándolas las cuatro «interacciones».

La función de la interacción débil no es tanto mantener las cosas unidas, como separarlas. En particular, ejerce interacciones con los neutrones, convirtiéndolos en protones. Esto recibe el nombre de desintegración beta y en este proceso el neutrón emite un electrón y una partícula llamada neutrino (en sentido estricto, se trata de un antineutrino). Más exactamente, el neutrón interacciona con el campo que está asociado a la interacción débil y emite uno de los cuantos de dicho campo. Estos cuantos se llaman bosones vectoriales intermedios y poseen masa, por lo que su alcance (el alcance de la interacción débil) está limitado por las reglas de la

incertidumbre cuántica. Posteriormente, el bosón se convierte en un electrón y un antineutrino. (El campo débil puede interactuar de otras maneras con otras partículas, pero no es necesario que hablemos aquí de esto.)

Lo más bonito de todo esto es que los físicos disponen de una simetría fundamental muy clara en el mundo de las partículas. Existen cuatro tipos de fuerzas fundamentales y también cuatro tipos de partículas fundamentales (los quarks superiores e inferiores, el electrón y el neutrino). Esto es todo lo que se necesita para explicar cualquier cosa que se pueda ver en el universo. Desafortunadamente, por razones que nadie ha podido entender, parece funcionar cierto tipo de duplicación (incluso triplicación) en las leyes de la física. En experimentos en los que se realizan choques de alta energía entre haces de partículas, los físicos han encontrado otras dos familias de partículas, con masas que son cada vez mayores. No son en ningún caso partículas que, del mismo modo en que los quarks se encuentran dentro de los nucleones, estuvieran «dentro» de las partículas a las que se hizo colisionar en estos experimentos. En vez de esto, son partículas que se han creado a partir de energía pura (como las parejas de quark y antiquark en el lugar dónde se rompe el elástico con el que comparábamos a los gluones). En casos extremos, estas partículas incluyen ciertas variedades que pueden no haber existido de forma natural en el universo desde el Big Bang. Hay una partícula que es la contrapartida del electrón, pero más pesada, y está asociada con su propio tipo de neutrino. Asimismo, también hay contrapartidas más pesadas para el quark superior y el quark inferior. Y luego, por si fuera poco, existe otro conjunto de todo esto, pero con más peso.

Ninguna de estas versiones más pesadas de las partículas básicas desempeña papel alguno en el modo en que el universo funciona actualmente (aunque pueden haber sido importantes en el Big Bang; véase el capítulo 12). Son todas ellas inestables y se desintegran progresivamente, convirtiéndose al final, después de liberar energía, en sus contrapartidas más ligeras. A los físicos que se dedican a la física de partículas les suministran muchos temas de investigación interesantes, así como un ámbito donde desarrollar muchas teorías sobre el comportamiento de estas partículas. Sin embargo, en este libro nos vamos a concentrar en hablar de cómo es el universo en el que vivimos, y no de las posibilidades más exóticas en cuanto a cómo podría haber sido, por lo tanto no diremos nada más, por ahora, sobre esas versiones más pesadas de las partículas fundamentales. En cambio, antes de salir de las profundidades de la física intranuclear para ir al mundo de las moléculas, quisiéramos hacer inventario de esas cuatro fuerzas fundamentales de la naturaleza y de cómo los físicos están intentando desarrollar una imagen global —una teoría unificada— para explicar el comportamiento de dichas fuerzas (o campos) y del conjunto básico de cuatro partículas, considerando todo ello en bloque.

La gravedad es la más débil de las cuatro fuerzas, pero es la más obvia en la vida

cotidiana y fue la primera que se estudió científicamente, por el modo en que se presenta y por su larguísimo alcance. En cualquier bloque de materia se suma la gravedad de todas y cada una de las partículas que haya en dicho bloque. No existe nada parecido a su neutralización, tal como sucede con las cargas eléctricas positivas y negativas que se neutralizan entre sí en un átomo. El alcance de la gravedad, como el alcance del electromagnetismo, es, en principio, infinito, porque los cuantos de su campo (gravitones) tienen masa cero, como los fotones. Sin embargo, esto no significa que la fuerza de un bloque de materia tenga el mismo valor en todos los lugares. En efecto, la fuerza gravitatoria de un objeto va disminuyendo en función de uno dividido entre el cuadrado de la distancia al objeto (la famosa «ley de los cuadrados inversos»); por lo tanto si nos situamos al doble de la distancia, la fuerza será sólo un cuarto de su valor; si estamos tres veces más lejos, la fuerza es un noveno de su valor, y así sucesivamente.

El electromagnetismo es mucho más fuerte que la gravedad, pero esto no es obvio en el entorno cotidiano, debido a que la electricidad y el magnetismo se presentan en dos variedades —carga positiva y negativa, polos norte y sur que se neutralizan entre sí. Sin embargo, al igual que la gravedad, tanto la electricidad como el magnetismo obedecen a leyes de cuadrados inversos, cuando no están completamente neutralizados. Como muestran las ecuaciones de Maxwell, son diferentes aspectos de una única fuerza, y las dos son transportadas, según el modelo del campo cuántico, por fotones que no tienen masa.

La mejor manera de entender las fuerzas relativas de la gravedad y el electromagnetismo consiste en pensar en lo que sucede cuando una manzana cae de un árbol. La fuerza que intenta mantener el átomo en la rama donde crece es la fuerza electromagnética, que actúa entre unos pocos átomos del tallo que sujeta la manzana a la rama. La fuerza que tira de la manzana hacia abajo es la fuerza de la gravedad, ejercida por todos los átomos del planeta Tierra tirando juntos y a la vez. La Tierra tiene una masa de poco menos de 6×10^{24} kilogramos. La gravedad de toda esta concentración de masa, actuando toda ella a la vez, es capaz de romper la unión electromagnética existente entre unos pocos átomos del tallo de una manzana.

Los valores relativos que alcanzan las cuatro fuerzas de la naturaleza se pueden expresar en función del valor de la fuerza nuclear. Si el valor de la fuerza nuclear se sitúa en uno, en números redondos el valor de la fuerza electromagnética sería alrededor de 10^{-2} (1 por 100 del valor de la fuerza nuclear), el valor de la fuerza débil sería 10^{-13} (una cienmilmillonésima del 1 por 100 del valor de la fuerza nuclear) y el valor de la gravedad sería solamente 10^{-38} . La fuerza nuclear es cien billones de billones de billones de veces más fuerte que la gravedad. Sin embargo hemos de recordar que la fuerza débil y la fuerza nuclear tienen un alcance muy limitado y actúan a una escala mucho menor que el diámetro de un átomo. El universo sería muy

diferente si tuvieran el mismo alcance que la gravedad y el electromagnetismo.

No obstante, esto es exactamente lo que, en opinión de los físicos, sucedió en el Big Bang, cuando el universo era aún muy joven. En el capítulo 12 veremos más acerca de esto. Por ahora, diremos que enlaza con la idea de los físicos que están intentando encontrar un modo de combinar todas las fuerzas en un solo paquete matemático: una teoría de campos unificada.

En cierto sentido, están ya a medio camino para conseguirlo. El electromagnetismo y la interacción débil ya se han combinado en un solo paquete que se llama la teoría electrodébil. Se desarrolló en la década de 1960 y ha sido el modelo para casi todas las iniciativas subsiguientes encaminadas a desarrollar una teoría de campos unificada.

La base de la teoría electrodébil es la similitud de las funciones desempeñadas por los cuantos del campo en el electromagnetismo (los fotones) y en la interacción débil (los bosones vectoriales intermedios). El electromagnetismo es una interacción que resulta especialmente fácil de describir mediante las matemáticas, porque sólo tiene una clase de cuanto de campo, el fotón, y no tiene ni carga ni masa. En la interacción débil los bosones vectoriales intermedios aparecen en tres variedades. Una de ellas tiene carga positiva, otra carga negativa y la tercera posee carga nula. Todas las variedades tienen masa. Sin embargo, en todo lo demás son como los fotones. Si no tuvieran los impedimentos que origina el límite de su alcance impuesto (a causa de su masa) por la incertidumbre, se comportarían exactamente igual que los fotones. No obstante, si el universo estuviera suficientemente caliente, los bosones vectoriales intermedios no tendrían que depender de la incertidumbre cuántica para existir. La energía de fondo del universo sería suficiente para fabricar bosones vectoriales intermedios a partir de la radiación caliente que llenaría el universo, siendo dichos bosones capaces de existir eternamente y tener un alcance infinito, exactamente igual que los fotones.

La teoría electrodébil describe cómo habrían interactuado las partículas en estas condiciones, estando en pie de igualdad los fotones y los bosones vectoriales intermedios. Y lo que es crucial, también describe cómo se habrían separado las diferentes fuerzas cuando el universo se enfrió y se expandió a partir del Big Bang, y predice que esta separación sería responsable de las diferencias entre las dos fuerzas que vemos actualmente, contando con que los bosones vectoriales intermedios poseen una cierta masa.

Los bosones vectoriales intermedios se fabricaron a partir de energía pura en experimentos con colisiones de partículas en el CERN, cerca de Ginebra, a principios de los ochenta, y se descubrió que tenían exactamente las masas que había predicho la teoría electrodébil.

El paso siguiente será hallar un tipo de teoría similar que englobe la fuerza

nuclear y la interacción débil. Esto significa poner a los gluones en un plano de igualdad con los fotones y los bosones vectoriales intermedios. En principio, tendría que funcionar el mismo planteamiento. Sin embargo, debido a que los gluones tienen todavía más masa que los bosones vectoriales intermedios, esto sólo funcionaría a temperaturas todavía más elevadas, que corresponderían a épocas anteriores aún más remotas durante el Big Bang.

La manera de comprobar estas teorías es, por supuesto, hacer que las partículas choquen entre sí en los aceleradores y, al hacerlo, recrear las condiciones en las que las fuerzas fundamentales están en un plano de igualdad. Tales experimentos son, como ya hemos mencionado, suficientemente potentes para comprobar la teoría electrodébil con bastante amplitud y asegurarse de que la teoría describe con precisión el modo en que se comportan las partículas reales. No se puede construir aceleradores lo suficientemente potentes como para probar las predicciones de las teorías equivalentes, que incluyen también la fuerza nuclear (estas teorías se conocen habitualmente como Grandes Teorías Unificadas, a pesar de que no incluyen la gravedad). No es demasiado difícil encontrar teorías de este tipo, pero una teoría que no esté comprobada mediante experimentos no tiene ningún valor práctico: «Si no concuerda con el experimento, es falsa». Ello resulta especialmente importante en este caso, porque hay que comprobar también más aspectos complicados que en la teoría electrodébil, ya que deben tenerse en cuenta ocho clases diferentes de gluones.

La variedad de las partículas implicadas permite a los físicos construir distintos tipos de Grandes Teorías Unificadas, de tal modo que cada una realiza diferentes predicciones sobre la manera en que debe comportarse la materia con energías muy altas. Sin embargo, con el fin de comprobar estas teorías y averiguar cuál es correcta (si alguna de ellas lo es), necesitaríamos un acelerador que fuera un billón (10^{12}) de veces más potente que los aceleradores que se utilizaron en los años ochenta para comprobar la teoría electrodébil. Incluso un acelerador tan grande como el sistema solar sería aún demasiado pequeño para hacer este trabajo. Como veremos en el capítulo 12, ésta es la razón por la cual los físicos que trabajan en física de partículas están centrando buena parte de su atención en la cosmología, utilizando sus teorías para predecir lo que sucedió en el Big Bang y comparando después sus predicciones con el modo en que aparece actualmente el universo.

Pero hay que tener en cuenta que tampoco ahora hemos incluido la gravedad en el paquete. El Santo Grial de la física teórica es una especie de Gran Superteoría Unificada que incluiría también la gravedad. Desgraciadamente, aunque los gravitones tienen masa cero, esto no resulta fácil; hay que ir a energías aún más altas para poder hacer que la gravedad sea tan intensa como las otras fuerzas. Debido a que la gravedad es tan débil, el escalón de energía que hay que subir para realizar aquí un trabajo con el mismo planteamiento es tan enorme que resulta difícil prever cómo se

podría hacer, por lo que un avance de este tipo ha de venir en una dirección completamente diferente, esto es, una teoría que reemplace la idea habitual de partículas, sustituyéndolas por puntos en los que se sitúan los cuantos del campo, como pequeños bucles de «cuerda».

Estas cuerdas forman bucles que son mucho menores que el tamaño de una partícula semejante a un protón, pero la cuestión clave es que sí tienen un tamaño definido; no son puntos matemáticos. En la teoría estándar de campos de cuantos se considera que las entidades auténticamente fundamentales, como son los quarks y los electrones (que, al menos hoy por hoy, siguen pareciendo fundamentales), presentan un tamaño nulo. El tamaño típico de un bucle de cuerda sería muy, muy pequeño, hasta tal punto que existe en la parte inferior de la escala en la que la incertidumbre cuántica es importante, lo cual significa que tienen diámetros de aproximadamente 10^{-33} cm. Un bucle de cuerda es mucho menor que un átomo, en la misma proporción en que un átomo es menor que el sistema solar.

Aunque no hay manera de comprobar directamente las predicciones de esta teoría de las cuerdas —al no poder concebirse experimentos que verifiquen estructuras a una escala tan pequeña—, sin embargo, hay dos razones por las que muchos físicos piensan actualmente que las cuerdas representan la verdad última y son los bloques básicos en la estructura de la materia.

La primera razón es que se libran de los molestos infinitos que plagan cualquier teoría basada en las ideas antiguas, según las cuales las partículas se consideran como puntos fundamentales. Esto significaría que las partículas tienen volumen cero y, en consecuencia, en algún momento dentro de la serie de cálculos, inevitablemente habría que dividir por cero. Por ejemplo, la fuerza eléctrica cumple una ley de cuadrados inversos. Su valor a cierta distancia de la fuente es proporcional a uno partido por el cuadrado de la distancia a dicha fuente. De este modo, a medida que nos acercamos a la fuente de un campo eléctrico, la fuerza aumenta al disminuir la distancia. Pero si la fuente tiene dimensión nula, la distancia puede disminuir recorriendo todos los valores hasta el cero. Si la fuente —por ejemplo, un electrón— es de tamaño nulo, la propia fuente siente una fuerza que es proporcional a uno dividido entre cero (aún peor: ¡uno dividido entre cero al cuadrado!). En otras palabras, infinito.

Existen modos de evitar esto, utilizando un truco llamado renormalización. En efecto, significa dividir un infinito entre otro para obtener un número finito. En primer lugar, imaginemos que infinito dividido por infinito es igual a 1, al igual que $2/2$ es 1, o $51234/51234$ es 1; pero el infinito es un material engañoso y se puede obtener cualquier respuesta que se desee a partir de este tipo de división. Por ejemplo, pensemos en un número que represente la suma total de los números naturales ($1 + 2 + 3 + \dots$). Obviamente es infinito. Ahora pongamos en esta serie el doble de cada

número y sumémoslos de nuevo. Está claro que la respuesta es, una vez más, infinito. Pero ¿cómo será este infinito comparado con el primer infinito? Podríamos pensar que, siendo cada número de la serie del segundo infinito el doble que su correspondiente en la primera serie infinita, el infinito que resulta de la suma de todos ellos tendría que ser el doble de grande que el primer infinito. Pero, pensemos otra vez. La segunda suma sólo contiene los números pares ($2 + 4 + 6 + \dots$). No contiene absolutamente ninguno de los números impares, por lo que sólo puede ser la mitad de grande que el primer infinito. Si dividimos este segundo infinito entre el primero, el resultado es $0,5$, no 2 (y desde luego no es 1).

Utilizando este tipo de trucos en la teoría estándar, se puede aplicar la renormalización para obtener resultados finitos que coincidan con las mediciones experimentales de la intensidad de una fuerza eléctrica. Sin embargo, se trata de un procedimiento que las circunstancias imponen a los físicos, y muchos físicos lo odian. Pero el problema de tener que manejar infinitos de este modo desaparece en la teoría de las cuerdas, porque aquí ya no hay que manejar puntos matemáticos que tienen tamaño nulo. No existe ninguna de esas incómodas fuentes puntuales (conocidas también como singularidades), por lo que no es necesaria la renormalización.

La segunda cosa importante en relación con la teoría de las cuerdas es que esta teoría predice la gravedad. Como ya hemos mencionado, en el antiguo planteamiento de la Gran Superteoría Unificada parece muy difícil encontrar un modo de incluir la gravedad, porque ésta es mucho más débil que las otras fuerzas. Cuando los científicos empezaron a jugar con la teoría de las cuerdas e intentaron utilizarla para conseguir una descripción de las fuerzas y partículas conocidas, supusieron que la gravedad resultaría también imposible de tratar en la nueva teoría y, al principio, ni siquiera intentaron tenerla en cuenta (esto sucedió en los años setenta).

Sin embargo, la forma en que funciona la teoría de cuerdas consiste en no tener una clase diferente de bucle de cuerda para cada tipo de partícula fundamental o de cuanto de campo, sino considerar a todas las partículas fundamentales y a los cuantos de campo como distintas formas de vibración del mismo tipo de cuerda, casi del mismo modo en que una única cuerda de violín puede tocar diferentes notas. A mediados de los años setenta, cuando unos cuantos físicos matemáticos estaban barajando esta idea como una concepción teórica abstracta, hallaron modos de describir de esta manera todas las entidades fundamentales conocidas (los quarks, los fotones, etc.). Un fotón correspondería a una vibración del bucle de cuerda equivalente a una nota de violín; un electrón correspondería a una vibración distinta (equivalente a una nota diferente en la cuerda del violín) y así sucesivamente. Pero también hallaron que existía otro modo según el cual los bucles de la cuerda podían vibrar y que no se correspondía con ninguna de las partículas y campos que

intentaban describir.

Al principio intentaron hallar un modo de librarse de esa forma de vibración no deseada. Luego se dieron cuenta de que correspondía exactamente al gravitón, es decir, al cuanto del campo gravitatorio. Así que resultó que la teoría de cuerdas, se quiera o no, incluye automáticamente a la gravedad.

Esto no significa que los físicos ya hayan conseguido su Santo Grial, su teoría total. Las complejidades matemáticas inherentes a la teoría de cuerdas hacen que sea difícil desarrollar completamente las ideas, y sigue existiendo la frustración de no poder comprobar las ideas mediante experimentos. Comparado con el modo en que se habían hecho anteriormente los progresos en la física de partículas, el descubrimiento de la teoría de cuerdas se ha realizado dando marcha atrás. Antes, desde el trabajo pionero de Rutherford, se habían hecho demostraciones sobre la estructura del átomo y luego se había demostrado la existencia de partículas más fundamentales, para desarrollar posteriormente teorías que explicaran los resultados de los experimentos. Sin embargo, esta vez la teoría ha llegado por el camino de las matemáticas puras, sin ningún vestigio de experimento con el que apoyar dicha teoría. Nunca ha existido un experimento en el que se hayan creado partículas que chocaran entre sí y rebotaran apartándose unas de otras de una manera que sugiriera automáticamente que lo que generaba los rebotes eran bucles de cuerdas vibrantes. La teoría de cuerdas se ha calificado excepcionalmente como un ejemplo de la física del siglo XXI que ha caído en el regazo de los físicos del siglo XX. No obstante, está ampliamente considerada como una alternativa viable al antiguo planteamiento de la teoría de campos que utilizaba la idea de partículas puntuales, y suprime definitivamente (o más bien nunca sufre) las dificultades que parecían hacer que una teoría completa de los campos cuánticos aplicada a la gravedad fuera un sueño imposible.

Independientemente de cómo se desarrollen las cosas durante los próximos cien años, hay algo que es cierto: es lo más lejos que podemos llegar, escribiendo a finales del siglo XX, en cuanto a comprobación de la estructura más interna de la materia se refiere. También es verdad que ninguna de estas teorías, por muy fascinante que sea, tiene ningún soporte directo en la naturaleza del mundo en que vivimos. Con el fin de comprender el mundo cotidiano que nos rodea, sólo necesitamos comprender el comportamiento de los átomos y sus componentes, considerando a los protones y los neutrones, así como a los electrones, como partículas fundamentales. Además sólo tenemos que preocuparnos de dos de las cuatro fuerzas fundamentales: la gravedad y el electromagnetismo.

Con el objeto de continuar nuestra descripción de la manera científica de comprender el mundo en el que vivimos, es hora de dirigir nuestra atención hacia afuera, comenzando de nuevo por los átomos, pero observando ahora cómo se unen para formar entidades cada vez más grandes. Trabajando en esta dirección, el objetivo

último de la ciencia es explicar la existencia del universo en sí mismo y cómo llegó a ser tal como es ahora. En algún lugar hacia la mitad de la escala de tamaños que va desde el mundo de las partículas fundamentales hasta el universo en su globalidad, encontraremos una explicación de nuestra propia existencia.

Cuando estábamos realizando comprobaciones en el interior del átomo, contemplábamos subunidades cada vez más sencillas del mundo físico a medida que nos centrábamos en escalas cada vez menores; sin embargo, nuestro primer paso saliendo de la escala del átomo para llegar al universo de grandes dimensiones tiene un efecto contrario, ya que observamos cómo el comportamiento de cosas tan sencillas como los átomos los capacita para unirse con el fin de constituir cosas tan complejas como son los seres humanos. El primer paso consiste en mirar de qué modo los átomos se adhieren unos a otros para formar moléculas. Nos vamos a la química.

Química

Toda la química se puede explicar mediante la física de la década de 1930; además basta con un modelo muy sencillo para dar una idea de por qué los átomos se unen, del modo en que lo hacen, para formar moléculas. Ni siquiera necesitamos preocuparnos por los neutrinos, aún menos por la fuerza nuclear o la fuerza débil. Desde luego, no tenemos que preocuparnos por la gravedad, que es, con mucho, demasiado débil para ser importante en las reacciones químicas. Tampoco debemos preocuparnos, al menos al principio, por la dualidad onda-partícula. Apenas hemos de tener en cuenta el hecho de que el núcleo de un átomo está constituido por dos tipos diferentes de partículas, el protón y el neutrón. Todo lo que necesitamos saber para comenzar a explicar la química es que, desde luego, los átomos contienen en su estructura electrones, que son partículas con carga negativa y están distribuidos de acuerdo con las reglas de la física cuántica a cierta distancia del núcleo central, cargado positivamente, y sometidos a la influencia de fuerzas electromagnéticas. Las bases de la química son increíblemente sencillas, pero la complejidad de las moléculas que se pueden construir de acuerdo con las reglas básicas sencillas es asombrosa.

La primera explicación que se logró para la tabla periódica de elementos se remonta en realidad a 1922, cuando Niels Bohr utilizó una mezcla de las viejas teorías de la física clásica (incluida la idea de electrón definido como una pequeña partícula) y las nuevas teorías de la física cuántica para desarrollar lo que es en esencia el modelo del átomo tal como se explica actualmente en las escuelas y que ya hemos mencionado en el capítulo 2. Esto sucedía diez años antes del descubrimiento del neutrón, hecho que muestra lo poco que se necesita saber sobre el interior del núcleo para entender la química.

El número atómico del núcleo (que, como sabemos en la actualidad, es igual al número de protones que hay en dicho núcleo) determina cuántos electrones hay en las capas más externas del átomo. Tiene que haber un electrón por cada protón, con el fin de que las cargas eléctricas se equilibren y el resultado sea un átomo eléctricamente neutro. Sea cual sea el número atómico, el primer electrón del átomo se encuentra en el nivel de energía más bajo, que se corresponde con el nivel de energía ocupado por el electrón solitario del átomo de hidrógeno, el átomo más sencillo que existe. El siguiente electrón se sitúa en el mismo nivel de energía, pero tiene un espín opuesto al del primer electrón. En este nivel sólo hay dos lugares disponibles. Sin embargo, hay más espacio para los electrones en el nivel de energía siguiente, en el sentido de alejamiento del núcleo. Así, los ocho electrones siguientes, correspondientes a los ocho números atómicos siguientes (es decir, a los ocho elementos siguientes de la tabla periódica) pueden encajarse todos ellos a más o menos la misma «distancia» del

núcleo. Después el patrón se repite (con algunas sutilezas que tienen interés principalmente para los especialistas).

Cada estrato de electrones situado alrededor del núcleo se llama «capa». Se trata de un nombre algo desafortunado, ya que implica (erróneamente) que las capas más externas no «ven» el núcleo, sino sólo la capa que tienen debajo; sin embargo, esta terminología es un remanente de principios de los años veinte y constituye una carga que hemos heredado. Una de las sutilezas, que comentaremos sin entrar en detalles, es que, debido al modo en que la energía de todo el átomo se altera a medida que se llenan sucesivas capas, se llega a un punto en que es posible introducir algunos electrones suplementarios en una de las capas internas. El resultado es toda una serie de elementos (las tierras raras) que presentan propiedades químicas muy similares, porque tienen las capas más externas idénticas, pero diferentes pesos y números atómicos. En estos elementos, cuando el número de protones del núcleo se va incrementado de uno en uno, los electrones suplementarios no forman nuevas capas en el exterior del átomo, sino que están escondidos en las capas internas.

La explicación de Bohr sobre la tabla periódica y sobre los detalles de espectroscopia que ya discutimos en el capítulo 2, incluía el hecho de que cada capa se llenaba en sucesión a medida que íbamos hacia átomos con números atómicos cada vez mayores (pero tuvo en cuenta debidamente los sutiles efectos que producen las tierras raras). Su idea clave, que hizo que la química fuera de una vez por todas una rama de la física, fue la constatación de que, por lo que respecta a sus interacciones con el mundo exterior (es decir, con otros átomos), prácticamente la única cosa que importa en un átomo es el número de electrones que hay en la capa ocupada más externa. Lo que sucede en zonas más profundas del átomo sólo tiene una importancia secundaria.

Ésta es la razón por la cual el hidrógeno, que únicamente posee un electrón, presenta propiedades químicas similares a las del litio, que tiene tres electrones, dos de los cuales están en la capa más interna y el tercero se sitúa en otra más alejada del núcleo. Es el motivo por el cual el flúor, que tiene dos electrones en la capa más interna y siete en la siguiente, muestra propiedades químicas similares a las del cloro, que tiene dos capas llenas (una que contiene dos electrones y otra que contiene ocho) y una tercera, la capa ocupada más externa, que contiene siete electrones más. Y así sucesivamente. Por supuesto, la existencia de las capas internas no se puede ignorar del todo, y tampoco se puede despreciar la masa global del átomo, ya que son factores que hacen que el cloro no sea exactamente igual que el flúor, ni el hidrógeno exactamente lo mismo que el litio. Sin embargo, en cada caso las similitudes son mucho más importantes que las diferencias.

Bohr no tenía idea de por qué una capa que contiene ocho electrones tendría que estar llena (o «cerrada», según su terminología), lo que no tiene importancia para

entender la química. Lo que importa es que, por la razón que sea, los átomos parecen tener capas cerradas. Posteriormente, aplicando la mecánica cuántica, fue posible explicar por qué esto tenía que ser así: una capa de electrones llena corresponde a un estado de baja energía para el átomo, y los estados de baja energía siempre son deseables. Sin embargo, no es necesario que nos preocupemos por esto, al igual que Bohr no tenía necesidad de preocuparse por ello en 1922. La idea de Bohr explicaba inmediatamente por qué a ciertos átomos les gustaba unirse para formar moléculas y (lo que es igualmente significativo) por qué otros átomos eran reacios a actuar de esa manera.

El mejor modo de explicar esto es poner un ejemplo. El hidrógeno sólo tiene un electrón en su única capa ocupada, pero le gustaría tener allí dos electrones, para cerrar la capa. El carbono tiene un total de seis electrones, dos de ellos en la capa más interna (que está, por consiguiente, cerrada) y cuatro en la segunda capa. Le gustaría tener ocho electrones en la capa externa ocupada. Una forma de que pueda cumplir su deseo consiste en unirse con cuatro átomos de hidrógeno para formar una molécula de metano (CH_4).

Imaginemos a cada átomo de hidrógeno (o más bien a cada núcleo de hidrógeno) asiéndose mediante la fuerza electromagnética a uno de los cuatro electrones externos del átomo de carbono, mientras que el átomo de carbono (el núcleo) se agarra a cada uno de los cuatro electrones de los cuatro átomos de hidrógeno. Entre los dos núcleos se comparten pares de electrones, de manera que cada átomo de hidrógeno tiene una participación en dos electrones, mientras que el átomo de carbono participa en ocho electrones (más sus dos electrones internos). Desde luego, éste resulta ser un estado más estable (energía más baja) que disponer de cuatro átomos de hidrógeno sueltos y un átomo de carbono también suelto.

El mismo razonamiento explica por qué sustancias como el helio (que tiene su capa más externa llena con dos electrones) y el neón (que tiene una capa exterior llena con ocho electrones, así como una capa interior con dos electrones y también llena) son reacios a cualquier interacción química. No tienen necesidad de ello, es decir, han alcanzado ya el equivalente químico del nirvana.

La clase de enlace químico que se describe mediante esta forma de compartir electrones entre átomos se llama enlace covalente. Un par de electrones compartidos constituye un enlace covalente. Sin embargo, como Bohr observó, existe otro modo de alcanzar el nirvana químico. Pensemos en un átomo que tiene un solo electrón en su capa ocupada más externa; por ejemplo, el sodio, que tiene dos capas llenas (una con dos y otra con ocho electrones) más un electrón solitario en el exterior de dichas capas. La manera más fácil de conseguir una capa exterior cerrada sería librarse de ese electrón más externo, para lo que bastaría con que éste tuviera algún lugar adonde ir.

Pensemos ahora en un átomo que tenga un solo hueco en su capa ocupada más externa; como por ejemplo el cloro, que tiene dos capas llenas (una con dos y otra con ocho electrones) más una tercera capa que contiene siete electrones. El modo más sencillo de formar una capa cerrada consistiría en agarrarse a un electrón suplementario, si hubiera uno disponible. Cuando el sodio y el cloro se unen, cada átomo de sodio puede dar un electrón a un átomo de cloro. Ambos alcanzan así el nirvana químico. Sin embargo, cada electrón que se intercambia lleva consigo una carga eléctrica igual a una unidad negativa, por lo que cada átomo de sodio se queda con una carga eléctrica residual +1, y cada átomo de cloro adquiere una carga eléctrica residual cuyo valor es -1. Dado que las cargas eléctricas opuestas se atraen entre sí, los átomos cargados (conocidos habitualmente como iones) se adhieren eléctricamente para formar cristales de sal común, es decir, cloruro de sodio. Esto se llama un enlace iónico.

Estamos viendo ejemplos claros y directos, pero en muchos casos el enlace químico entre los átomos de una molécula conlleva aspectos tanto del enlace covalente como del enlace iónico. Sin embargo, no es necesario que nos preocupemos por los detalles. Lo importante aquí es que todas las reacciones químicas se pueden explicar de este modo, por el hecho de compartir o intercambiar (o ambas cosas a la vez) electrones entre los átomos, en un intento de alcanzar el deseable estado de tener llenas las capas ocupadas más externas.

Como todos los mejores modelos científicos, el de Bohr formulaba una predicción que se podía comprobar. Esta predicción se comparó con los resultados de varios experimentos y el modelo pasó la prueba con un éxito rotundo.

Incluso en 1922, existían todavía algunos huecos en la tabla periódica de elementos, que correspondían a elementos aún no descubiertos cuyos números atómicos eran 43, 61, 72, 75, 85 y 87. El modelo de Bohr vaticinaba con todo detalle las propiedades de los elementos necesarios para llenar los huecos de la tabla, del mismo modo que Mendeleiev había formulado predicciones similares durante el siglo anterior. Sin embargo, resultaba crucial el hecho de que el modelo de Bohr, para el elemento número 72 (el hafnio) auguraba propiedades diferentes de las que predecían otros modelos rivales. Menos de un año después de haberse hecho esta predicción, el hafnio fue descubierto y bautizado, y además resultó que presentaba justo las propiedades que Bohr había predicho (en el caso de los otros elementos que faltaban, también resultó que tenían las propiedades vaticinadas anteriormente, aunque en estos casos las diferencias entre las predicciones de Bohr y las de otros modelos no eran tan claras).

Todo esto se realizó sin utilizar la idea de ondas de electrones, que no se había descubierto aún por aquel entonces. En una molécula como la de hidrógeno (H_2) se consideraba que cada átomo de hidrógeno contribuía al par con un electrón, estando

el par de electrones, más o menos, entre los dos núcleos de hidrógeno, de tal forma que cada núcleo podía disponer de los dos electrones. Sin embargo, la idea de ondas electrónicas aporta una manera de considerar las moléculas que es, en cierto modo, más fácil de entender, una vez que se haya aceptado que cualquier electrón puede realmente dispersarse por un volumen de espacio equivalente al tamaño del átomo o de la molécula, y que no está limitado a un solo punto en la órbita descrita alrededor del núcleo. El volumen que ocupa un electrón en un átomo o una molécula se denomina orbital y en el caso del átomo de hidrógeno el orbital es esencialmente una capa esférica que rodea al núcleo. Cada orbital puede estar ocupado por un máximo de dos electrones, de manera que el espín de cualquiera de ellos sea opuesto al espín del otro.

Actualmente tenemos una visión distinta de la molécula de hidrógeno, que no es la que ofrecía el modelo de Bohr de 1922. En vez de los dos electrones puntuales situados en el espacio que hay entre los dos núcleos, lo que podemos imaginar es que cada electrón rodea ambos núcleos.

Imaginemos que los dos núcleos están situados a una cierta distancia uno del otro, puesto que se mantienen alejados por la mutua repulsión de sus cargas positivas. El orbital que ocupan ahora los dos electrones se puede describir como si tuviera la forma de una especie de reloj de arena (con un cuello muy ancho) rodeando ambos núcleos, con un núcleo en cada mitad del reloj. Cada núcleo está completamente rodeado por el orbital y siente la presencia de los dos electrones en el orbital correspondiente, cerrando la capa más interna. Es como si la molécula de hidrógeno fuera un único átomo que contuviese dos núcleos. La razón por la que se forman las moléculas de hidrógeno es que los dos electrones de su configuración tienen una cantidad de energía menor que la que tendrían si estuvieran en dos átomos de hidrógeno separados. La química trata precisamente de esto: de minimizar la energía de los electrones.[5] Por supuesto, la energía ha de ir a algún lugar, y cuando dos átomos de hidrógeno chocan y se unen, con un reordenamiento de los electrones en el modo que se ha dicho, se libera algo de energía en forma de ondas electromagnéticas (o, si se prefiere, como fotones). Una parte de esta energía se convierte en energía cinética para las moléculas (la energía de su movimiento), haciendo que se muevan más rápido (lo que significa que incrementan su temperatura).

El hecho de que exista un estado de energía inferior no implica que siempre vaya a ser ocupado. Si estoy en la cima de una montaña, admirando el panorama, me encuentro en un estado de energía superior al que ocuparía si estuviera en el valle al pie de la montaña, porque estoy más alejado del centro de la Tierra (más lejos del centro de su campo gravitatorio). Sin embargo, siempre que tenga cuidado de no caerme, puedo permanecer en el estado de alta energía todo el tiempo que desee. En el caso de los átomos de hidrógeno que se combinan para formar moléculas, es muy

fácil acceder al estado de menor energía (basta con «rodar monte abajo»), ya que prácticamente cualquier choque entre dos átomos de hidrógeno a temperatura ambiente puede reordenar los electrones para formar moléculas de hidrógeno.

De un modo similar, los átomos de oxígeno se combinan muy fácilmente entre sí para formar moléculas diatómicas, O_2 . En este caso existe una sutil diferencia. La capa ocupada más externa de un átomo de oxígeno contiene seis electrones, por lo que se necesitan dos más para llenar dicha capa. Por consiguiente, aplicando el esquema de Bohr, en cada molécula de oxígeno hay dos pares de electrones (en total cuatro electrones, dos de cada átomo) que están compartidos entre los dos átomos, formando lo que se conoce como enlace doble (en un enlace simple interviene un electrón de cada átomo). Incluso en el modelo ondulatorio cuántico sucede que las capas interiores, que están completas, desempeñan una función poco importante a la hora de formarse los enlaces, y en la teoría ondulatoria se puede considerar que cada molécula de oxígeno está constituida por dos núcleos centrales (en este caso se trata de núcleos envueltos en una capa interior que está llena con solamente dos electrones), encontrándose todo ello rodeado por una capa de doce electrones, según aquella configuración en forma de reloj de arena con un cuello muy ancho.

Si comenzáramos por escarbar en una mezcla de átomos de hidrógeno y átomos de oxígeno, y quisiéramos unir moléculas a mano (¡utilizando unas pinzas muy finas!) para dar el nivel de energía más baja a los electrones en cuestión, obtendríamos moléculas de agua, H_2O . Recuérdese que cada átomo de oxígeno necesita dos electrones suplementarios para completar su capa más externa, mientras que cada átomo de hidrógeno necesita sólo uno. Por lo tanto, si un átomo de oxígeno forma enlaces con dos átomos de hidrógeno independientes, los electrones de los tres átomos consiguen un estado de energía más baja, que resulta ser aún más baja que en las moléculas de oxígeno o en las moléculas de hidrógeno. Según la teoría ondulatoria, los dos núcleos de hidrógeno junto con el núcleo de oxígeno se sitúan formando una V, de modo que el oxígeno está en el vértice de la V y hay un total de ocho electrones (seis de la capa más externa del oxígeno, más el único electrón de cada uno de los dos átomos de hidrógeno) con los que se completa una capa que rodea al conjunto de los tres componentes centrales.

Sin embargo, tanto las moléculas de hidrógeno como las de oxígeno son relativamente estables, por lo que a temperatura ambiente rebotarían las unas contra las otras, sin combinarse para formar moléculas de agua. Una manera de imaginarse gráficamente por qué esto tendría que ser así, aunque la configuración del agua tenga menos energía, es pensar que las distintas moléculas ocupan unos pequeños huecos en la ladera de una colina. El hueco que ocupa una molécula de hidrógeno y el que ocupa una molécula de oxígeno están ambos situados a mayor altura en la ladera de la colina que el hueco ocupado por una molécula de agua. Si de un golpe lanzáramos

fuera de sus huecos a la molécula de hidrógeno y a la molécula de oxígeno, éstas rodarían colina abajo hasta meterse en el hueco del agua. Pero necesitan ser lanzadas inicialmente mediante un golpe para salir de los huecos en los que anidan, es decir, las moléculas de hidrógeno y de oxígeno han de chocar entre sí lo suficientemente rápido como para que sus estructuras se rompan y puedan ser reorganizadas posteriormente para formar moléculas de agua.

Esto sucede —espectacularmente— cuando una mezcla de hidrógeno y oxígeno a temperatura ambiente entra en ignición por efecto de una chispa o una llama. El calor de la chispa o de la llama hace que unas cuantas moléculas se muevan con la rapidez necesaria para que la reorganización se produzca. Cuando esto tiene lugar, se libera más energía, la cual calienta las moléculas próximas y las impulsa a reaccionar del mismo modo. Así, una ráfaga de actividad química se extiende por la mezcla de oxígeno e hidrógeno como una explosión, y al final conseguimos obtener unas cuantas gotitas de agua.

Debido a que cada molécula de hidrógeno tiene dos átomos (H_2) y cada molécula de oxígeno también tiene dos átomos (O_2), mientras que cada molécula de agua posee un átomo de oxígeno y dos átomos de hidrógeno (H_2O), la explosión es más efectiva si comenzamos con una cantidad de hidrógeno que sea el doble que la de oxígeno, de tal forma que dos moléculas de hidrógeno se puedan combinar con cada molécula de oxígeno para producir dos moléculas de agua ($2H_2O$), sin que sobre nada. En este caso la explosión es realmente espectacular (por favor, ¡no intente hacerlo en su casa!).

Uno de los mayores triunfos de la física cuántica fue desarrollar una explicación completa del enlace químico hablando de las ondas electrónicas, trabajo en el que participó en gran medida Linus Pauling hacia finales de los años veinte y principios de los treinta. La gran aportación del planteamiento químico realizado en el marco de la mecánica cuántica es la posibilidad de calcular (al menos para las moléculas más sencillas), mediante las leyes básicas de la física, el cambio energético que se produce cuando las moléculas se reorganizan del modo que acabamos de explicar. Estos cálculos predicen los cambios energéticos que se miden en la realidad cuando tienen lugar las correspondientes reacciones químicas.

Erwin Schrödinger publicó en 1926 su descripción ondulatoria del electrón y, un año después, dos físicos alemanes, Walter Heitler y Fritz London, habían utilizado su ecuación de ondas para calcular el cambio energético global que se produce en uno de los procesos descritos anteriormente, a saber, cuando dos átomos de hidrógeno se unen para formar una molécula de hidrógeno. Sus resultados se aproximaban mucho a las cifras obtenidas experimentalmente. Otros cálculos posteriores, en los que ya se incluían los perfeccionamientos llevados a cabo por Pauling, dieron como resultado una concordancia aún más estrecha entre las predicciones teóricas y los datos

experimentales. Era el año 1927 y ello supuso un avance decisivo. Con anterioridad a esta fecha, todo lo que los químicos podían hacer era afirmar que, por razones desconocidas, los electrones tenían tendencia a emparejarse entre sí dentro de los átomos y que los átomos a su vez tendían a llenar sus capas de electrones. Pero en 1927 las ecuaciones les decían además por qué esto tenía que ser así. La razón era que estas configuraciones tienen menos energía, de un modo que se puede calcular y cuantificar, y además coincide con los resultados experimentales.

En principio, se puede aplicar el mismo tipo de cálculo a cualquier clase de moléculas, aunque en la práctica llega a ser muy difícil efectuarlo con precisión en los casos de moléculas complejas, por lo que los químicos tienen que recurrir a distintas técnicas de aproximación, de las que no es preciso hablar aquí. Lo que importa realmente es que los principios en que se basa el comportamiento de las moléculas se entienden muy bien. En vez de entrar en detalles relativos a distintas reacciones químicas, que serían simplemente variaciones sobre los temas que ya he comentado, prefiero avanzar en dirección a lo más interesante que ofrece la química: la parte de esta ciencia que se refiere a los seres vivos, incluidos nosotros mismos. En primer lugar, para establecer un marco, necesitamos saber un poco más sobre otras formas de atracción entre átomos y moléculas, concretamente sobre tipos de enlaces más débiles.

En algunos compuestos existen grupos completos de átomos que funcionan juntos, comportándose igual que un único átomo de un compuesto sencillo, como pueda ser el cloruro sódico (sal común, NaCl). Un ejemplo típico es el carbonato cálcico, la creta común (CaCO₃). El grupo CO₃ actúa en bloque como un átomo que lleva dos unidades suplementarias de carga negativa (dos electrones suplementarios) y se denomina ion de carbonato. El átomo de calcio desempeña en la molécula de carbonato cálcico el mismo papel que el sodio en la sal común (NaCl), salvo por el hecho de que lleva dos unidades de carga positiva, no una (de lo que se podría deducir, correctamente, que el átomo de calcio tiene dos electrones en su capa ocupada más externa y le encantaría deshacerse de ellos si pudiera encontrar un receptor dispuesto a aceptarlos).

El átomo de calcio ha entregado dos de sus electrones al ion de carbonato, que es un átomo de carbono unido mediante enlaces covalentes con tres átomos de oxígeno y que posee además dos electrones suplementarios suministrados por el átomo de calcio. Los cálculos cuánticos, en cuyos detalles no vamos a entrar aquí, confirman que se trata de una configuración con relativamente baja energía. De hecho, es tan estable que resulta hasta cierto punto fácil liberar el ion de carbonato separándolo de su compañero, de tal modo que pueda intervenir en reacciones químicas como una unidad independiente. Otro tipo de ion que se comporta de una manera similar es el amonio (NH₄), en el que un único átomo de nitrógeno está unido mediante enlaces

covalentes a cuatro átomos de hidrógeno, pero ha perdido uno de sus electrones, quedándose con una carga global de una unidad positiva. Una vez más, la razón por la que esta sustancia ha de ser estable, a pesar de su exceso de carga, es que tiene una configuración en la que el nivel de energía es particularmente bajo.

La forma más débil de atracción entre átomos y moléculas es la que hace que las «pequeñas partículas» de Feynman sean pegajosas, «atrayéndose unas a otras cuando están separadas por una distancia muy corta». Se trata de fuerzas originadas por la imperfección en la cobertura que la nube de electrones aporta al núcleo de un átomo (o a los núcleos que haya en una molécula). Reciben el nombre de fuerzas de Van der Waals, en honor al físico holandés Johannes van der Waals, que las investigó a finales del siglo XIX. Estas fuerzas se originan porque, en cualquier átomo o molécula, la nube de electrones con su carga negativa se extiende de forma dispersa alrededor del núcleo, que tiene carga positiva, con lo que cualquier nube de electrones situada en el vecindario (rodeando otra molécula u otro átomo) puede «ver» la carga positiva y ser atraída por ella hasta cierto punto, aunque los dos átomos o las dos moléculas se repelan mutuamente al llegar a la proximidad necesaria para que sus dos nubes de electrones interaccionen entre sí.

No tendremos mucho más que decir sobre las fuerzas de Van der Waals como tales; sin embargo, existe otro tipo de enlace muy importante que aún no hemos mencionado. Tiene una especial relevancia en la historia de las moléculas de los seres vivos y se puede considerar como una especie de efecto de Van der Waals extraordinariamente fuerte, aunque es más débil que los tipos habituales de enlaces covalentes o iónicos. Se denomina enlace de hidrógeno, porque sólo funciona en compuestos que contengan este elemento.

El hidrógeno representa un caso único, ya que es el único átomo participante en reacciones químicas que no tiene al menos una capa de electrones llena bajo la capa que interviene en dichas reacciones. Solamente posee una capa que está ocupada, la más interna. El helio, por supuesto, tampoco tiene más que una capa ocupada. Sin embargo, en el caso del helio esta capa está llena, por lo que éste no reacciona con ningún otro elemento, es decir, no se combina con otros átomos para formar moléculas. Se siente feliz tal como es. Incluso el litio, que es entre todos los demás átomos el más parecido al hidrógeno, aunque tiene un electrón solitario en su capa ocupada más externa, posee también dos electrones en su capa interna que, hasta cierto punto, protegen al núcleo frente al mundo exterior. En cambio, sin su único electrón, el núcleo de hidrógeno estaría desnudo. No tendría nada con que proteger su pudor, ni siquiera parcialmente; sería incapaz de disimular la fuerza de su carga positiva.

Examinar el ejemplo del agua es la mejor manera de entender el poder del enlace de hidrógeno y el papel que este enlace desempeña en la química. El agua, que como

es sabido resulta esencial para la vida (el nombre que recibe una zona desprovista de agua, desierto, es sinónimo de región desprovista de vida), presenta algunas propiedades muy peculiares, todas ellas gracias al enlace de hidrógeno. El factor más importante a la hora de determinar si una sustancia es un sólido, un líquido o un gas, cuando dicha sustancia se encuentra a temperatura ambiente, es habitualmente su peso —su peso molecular— que se calcula sumando los pesos atómicos de todos sus componentes. Cuanto más pesada sea una molécula, más energía necesita (más elevada ha de ser la temperatura) con el fin de poder desplazarse por todas partes con la libertad necesaria para ser un líquido o (aún con más libertad) un gas.

Toda molécula de agua está formada por dos átomos de hidrógeno (cada uno de ellos con peso atómico igual a uno) y un átomo de oxígeno (peso atómico igual a 16), y posee un peso total (en la escala en que el carbono-12 tiene un peso atómico igual a 12 unidades) de 18 unidades, o también 18 daltons, como se denominan las unidades en honor al pionero de la química John Dalton. Sin embargo, a temperatura ambiente, el agua es un líquido, a pesar de que un gran número de compuestos de pesos moleculares mucho mayores son gases cuando se encuentran en las mismas condiciones; por ejemplo, el dióxido de carbono, que tiene un peso molecular de 44 daltons, o el dióxido de nitrógeno, con un peso de 46 daltons, e incluso el propio oxígeno molecular, cuyo peso es de 32 daltons. Si se extrajera un átomo de oxígeno de la molécula de oxígeno y se sustituyera por dos átomos de hidrógeno, el peso de dicha molécula se reduciría en cerca de un 50 por 100, y además, paradójicamente, esto haría que, al mismo tiempo, las moléculas de agua resultantes se adhirieran entre sí para formar un líquido, en vez de volar libremente como las de un gas.

El modo en que el enlace de hidrógeno ocasiona todo esto se puede entender hablando de la geometría de una molécula de agua y mediante la descripción ondulatoria cuántica de los electrones. La forma de V de la molécula de agua presenta un ángulo realmente muy amplio, con una separación de $104'5^\circ$ (mayor que un ángulo recto) entre las posiciones de los dos núcleos de hidrógeno de la molécula. En la representación realizada mediante las ondas de los electrones, estos dos núcleos (simples protones) se agarran a la parte más interna del único átomo de oxígeno (un núcleo encajado en una capa interior cerrada formada por dos electrones), con un total de ocho electrones más formando una capa grumosa que rodea los tres componentes internos, como una ancha esfera con dos bultos en su superficie. Sin embargo, la parte central del oxígeno tiene un núcleo que contiene ocho protones, que son los que ejercen, con mucho, la mayor influencia sobre los electrones. La atracción electromagnética ejercida por cada uno de los núcleos solitarios de hidrógeno es débil en comparación con la que ejerce el núcleo del oxígeno, por lo que el resultado es que la nube exterior de electrones se concentra en el extremo de la molécula donde se halla el átomo de oxígeno. Esto no sería tan importante si los

núcleos de hidrógeno —los protones— tuvieran algunos electrones internos con los que revestirse, pero no es así. Si los tuvieran, no serían núcleos de hidrógeno.

Para otra molécula de agua, la actuación —y el efecto— de una molécula de agua cercana depende de cuál sea el extremo que apunta hacia ella. Si el extremo donde se encuentra el oxígeno está a la vista, lo que la otra molécula «ve» (o siente) es una nube de electrones que tiene una carga global negativa. Si el que apunta es el extremo correspondiente a los átomos de hidrógeno, lo que se ve, a través de un fino velo de electrones, es un par de simples protones, una carga positiva neta.[6] Así pues, cuando las moléculas del agua chocan entre sí, muestran una tendencia natural a unirse, ligándose el extremo de oxígeno de una molécula a uno de los átomos de hidrógeno de la molécula vecina, con una fuerza intermedia entre la del enlace covalente normal y las fuerzas de Van der Waals habituales. Debido a que el ángulo formado por los dos átomos de hidrógeno en la molécula de agua es tan grande, hay espacio suficiente para que cada uno de ellos se agarre de esta forma a un extremo de oxígeno de otra molécula de agua, sin que estas dos moléculas de agua se interpongan la una en el espacio de la otra.

En el agua en estado líquido este efecto es transitorio, ya que los enlaces están constantemente intentando formarse y romperse debido al movimiento de las moléculas; pero es suficiente para hacer que las moléculas tengan tendencia a adherirse, que pasen una al lado de la otra, rozándose, en parte como si resbalaran, pero también basta para hacerlas incapaces (al menos a temperatura ambiente y en todo el recorrido de temperaturas hasta los 100°C) de liberarse por completo y separarse volando independientes en forma de gas. Sin embargo, cuando el agua se congela convirtiéndose en hielo, los enlaces de hidrógeno ocasionan otro importante efecto. Cuando los movimientos moleculares se van haciendo más lentos, hasta ser capaces de formar una estructura cristalina, vibrando cada molécula suavemente en el lugar que le corresponde dentro de la retícula cristalina, es el enlace de hidrógeno el que determina de qué manera toma forma la retícula. El ángulo determinado entre los dos átomos de hidrógeno de una molécula de agua es justamente el adecuado para que las moléculas se sitúen en una formación muy abierta, en la cual cada átomo de oxígeno no sólo se encuentra unido a sus dos compañeros de hidrógeno habituales constituyendo una molécula, sino que también está unido mediante enlaces de hidrógeno a otras dos moléculas de agua; además cada átomo de hidrógeno está vinculado de esta manera a otra molécula de oxígeno, así como a sus compañeros moleculares habituales.

La estructura que resulta es similar a la estructura cristalina del diamante, aunque no es tan fuerte. Se trata de una estructura muy abierta, con mucho espacio entre los átomos, y el diseño habitual de la disposición en retícula es responsable, por ejemplo, de la hermosa y regular geometría que presenta el diseño de un copo de nieve. Sin

embargo, la estructura es tan abierta que cualquier cantidad de agua congelada (hielo) ocupa en realidad un volumen ligeramente más amplio que la misma cantidad de agua líquida. Por lo tanto, el hielo es menos denso que el agua y flota en ella.

Éste es un fenómeno cotidiano tan corriente que lo damos por supuesto. Pero detengámonos y pensemos en ello durante un momento. Sería realmente muy extraño que un bloque de hierro que se dejara caer en una cuba llena de hierro fundido fuera a flotar, en vez de hundirse hasta el fondo de la cuba. La mayoría de los sólidos son más densos que sus líquidos correspondientes, porque en los sólidos el movimiento térmico de los átomos y de las moléculas se hace más lento, y tanto los unos como las otras pueden apretarse más, aproximándose mutuamente. Sin embargo, en el caso del agua es esta pérdida de velocidad de los movimientos moleculares lo que permite que se formen de manera adecuada los relativamente delicados enlaces de hidrógeno, que son los que mantienen a las moléculas en sus lugares dentro de una redícula muy abierta. Si el hielo tuviera con el agua la misma relación que mantienen las formas sólidas de otros compuestos con sus formas líquidas correspondientes, los cubos de hielo de las bebidas se hundirían hasta el fondo, y el hielo que se forma en invierno en una charca o un lago estaría en el fondo y no formando una corteza en la superficie del agua. El océano Ártico no estaría cubierto de hielo, lo cual tendría efectos importantes en el clima, como ya veremos en el capítulo 8. Todo esto se debe al enlace de hidrógeno, que se puede explicar muy satisfactoriamente mediante la física cuántica, llegando hasta la fuerza del enlace y el espaciamiento de los ángulos en la redícula de un cristal del copo de nieve.

El hidrógeno constituye un caso único, porque es el único átomo que posee un solo electrón, y es el único átomo que puede actuar como el socio de carga positiva cuando los enlaces se forman de esta manera. Sin embargo, hay también otros átomos, como el oxígeno, que pueden actuar como socios de carga negativa en un enlace de este tipo. El más importante, junto con el oxígeno, es el nitrógeno. El átomo o el núcleo de hidrógeno puede, en efecto, formar un puente entre dos átomos que se encuentran en diferentes moléculas, entre dos átomos de oxígeno o entre dos átomos de nitrógeno (o, asimismo, entre un átomo de oxígeno de una molécula y un átomo de nitrógeno de otra molécula). De este modo, puede mantener unidas dos moléculas que en otro caso estarían separadas, aunque no las une con tanta fuerza como lo haría un enlace covalente normal. Sin esta posibilidad, como explicaré en el próximo capítulo, nosotros no estaríamos aquí. Aparte de la importancia del enlace de hidrógeno en cuanto a determinar las propiedades del agua, de la que depende nuestra vida, este enlace también determina la estructura de la molécula básica de la vida, el ADN. Pero ¿qué es lo que hace que las moléculas de los seres vivos, incluida la del ADN, sean diferentes de las moléculas no vivas, tales como el agua y el dióxido de carbono?

Se trata de la división más básica que se realiza dentro de la química y que ya era conocida por los que practicaban la alquimia, aunque siguió siendo un misterio hasta bien entrado el siglo XIX. Algunas sustancias —incluso compuestos tales como la sal común o el agua— se pueden calentar y después enfriar sin que varíen esencialmente. [7] La sal se puede calentar hasta ponerla incandescente, pero sigue siendo sal; el agua se puede evaporar convirtiéndose en un gas, pero se vuelve a condensar y es de nuevo agua; y así en otros muchos casos. Sin embargo, hay otro tipo de sustancias o compuestos que no siguen siendo lo mismo cuando se calientan. Por ejemplo, el azúcar o la madera. Si calentamos el azúcar, aunque sea suavemente, se quema y no deja de estar quemado cuando se enfría otra vez. En cuanto a un trozo de madera, todos sabemos lo que le pasa cuando lo calentamos.

La distinción entre las dos clases de sustancias se hizo formalmente en 1807 y su autor fue el químico sueco Jóns Berzelius, uno de los primeros en apoyar la versión de John Dalton de la teoría atómica. Berzelius constató que todas las sustancias del primer grupo estaban asociadas a sistemas no vivientes, mientras que el segundo grupo de sustancias se derivaba directa o indirectamente de sistemas vivientes. Por ello denominó al segundo grupo sustancias «orgánicas» y al primer grupo sustancias «inorgánicas». Los compuestos orgánicos que se definían de este modo eran mucho más complicados que los compuestos inorgánicos y tenían muchos más átomos en cada molécula; por todo ello, al principio se pensó que solamente podían ser producidos por organismos vivos gracias a la acción de alguna misteriosa fuerza vital. Hubo que esperar hasta 1828 para que el químico alemán Friedrich Wöhler descubriese, accidentalmente, que podía fabricar urea (uno de los compuestos constituyentes de la orina) calentando una sustancia relativamente simple, el cianato de amonio, que por aquel entonces se consideraba inorgánico.

En la segunda mitad del siglo XIX, poco a poco fue quedando claro que son aplicables exactamente las mismas reglas básicas tanto a los procesos químicos inorgánicos como a los orgánicos, y que las diferencias entre las dos categorías se deben totalmente a la complejidad de la mayoría de las moléculas orgánicas. También quedó claro que las moléculas orgánicas tienen algo más en común: todas ellas contienen carbono. Se cambió la definición de molécula orgánica diciendo que lo es toda molécula que contiene carbono, mientras que toda molécula que no contiene carbono se llama inorgánica. En sentido estricto, según esta definición, incluso una molécula tan sencilla como la del dióxido de carbono (CO_2) es orgánica, aunque algunos químicos podrían considerar simbólicamente que este compuesto es inorgánico.

Lo importante en todo este asunto es que la distinción entre compuestos orgánicos e inorgánicos, y el modo en que se denominan, indica que hay algo especial relacionado con el carbono como elemento. Los átomos de carbono son realmente la

base de las moléculas de la vida, pero entendiendo esto como un resultado de sus propiedades físicas, que actualmente se conocen bien, y no porque estos átomos sean la morada de una fuerza vital sobrenatural. La característica principal del átomo de carbono que hace que la química orgánica sea tan complicada es el hecho de que tiene cuatro electrones (la mitad de los que habría en una capa llena) en su capa ocupada más externa (y además una única capa interior completa con dos electrones). Esto significa que puede formar simultáneamente cuatro enlaces químicos, el máximo posible. Si tuviera menos electrones exteriores, habría menos electrones disponibles para formar enlaces; si tuviera más, la capa exterior se encontraría tan cerca de estar llena que lo importante sería el número de «huecos» que quedarán en ella para alojar electrones de enlace procedentes de otros átomos. Los cuatro electrones exteriores otorgan al carbono una capacidad de enlace máxima, incluida la capacidad de formar enlaces con otros átomos de carbono.

Avanzando aún más en la tabla periódica se encuentran otros átomos, principalmente el silicio, que también tienen exactamente cuatro electrones exteriores. Sin embargo, la diferencia está en que, entre la capa más externa y el núcleo central poseen más capas que ya están llenas, por lo que la influencia del núcleo sobre los electrones exteriores está debilitada y los enlaces que forman no son tan fuertes. El carbono es el átomo más pequeño que tiene exactamente cuatro electrones en su capa más externa, de manera que puede formar al mismo tiempo cuatro enlaces fuertes. Ésta es la llave de acceso a la vida.

Antes de que sigamos adelante para pasar a examinar las moléculas de la vida en sí mismas, vale la pena mencionar otra característica más del modo en que se forman los enlaces. Esta característica se pone de manifiesto en la manera como el carbono se combina con otros elementos. Tiene que ver con el hecho de que los átomos y las moléculas son objetos tridimensionales. Una vez más, se trata de un fenómeno que sólo se puede entender a la luz de la física cuántica. Además constituye un paso importante hacia la comprensión del funcionamiento de la vida y nos trae el mensaje de que nuestra existencia depende estrechamente del comportamiento del mundo cuántico.

Según la física cuántica básica, los orbitales que ocupan los cuatro electrones de la capa exterior de un átomo de carbono no tienen por qué ser todos iguales. Hay un orbital esférico (exactamente igual que el único orbital interior que forma la primera capa) y otros tres orbitales, cada uno de los cuales tiene la forma de un reloj de arena ancho (con un lóbulo a cada lado del núcleo central) y deben estar dispuestos formando entre sí ángulos rectos. Al haber sólo cuatro electrones, cada uno de ellos ocupa un orbital, aunque en cada orbital hay sitio para un segundo electrón que tenga el espín opuesto al del primero. Ésta es la razón por la cual el carbono puede formar cuatro enlaces. Según esto, cuando el carbono constituye enlaces tendría que haber

tres conexiones dispuestas en direcciones que forman entre sí ángulos rectos, y otra que no tuviera ninguna preferencia en absoluto por lo que respecta a la dirección. Sin embargo, se han realizado estudios sobre temas tales como la estructura de los cristales y en dichos estudios se demuestra que no sucede así.

No voy a entrar en detalles, pero resulta bastante lógico que la forma y la complejidad de un cristal (como la forma del copo de nieve) reflejen la forma de las moléculas que lo constituyen. Este tipo de estudios muestra que cuando el carbono constituye cuatro enlaces con otros átomos (en el caso más sencillo, con cuatro átomos de hidrógeno, para dar una molécula de metano, CH_4), los enlaces son idénticos entre sí y están dispuestos de manera simétrica apuntando hacia los vértices de un tetraedro imaginario en el que el átomo de carbono se sitúa en el centro.

La explicación se basa por completo en una comprensión más sofisticada del comportamiento cuántico del electrón y fue desarrollada por Linus Pauling en un trabajo fundamental sobre la naturaleza del enlace químico publicado en 1931. El orbital esférico correspondiente se llama orbital s y los tres orbitales perpendiculares reciben el nombre de orbitales p. Según sugirió Pauling, lo que sucede es que, en vez de existir dos tipos distintos de orbitales que operan el uno al lado del otro, la diferenciación es confusa y el orbital único s se combina con los tres orbitales p para producir un conjunto de cuatro estados idénticos e híbridos, conocidos como orbitales sp^3 . Esto refleja el tipo de dualidad cuántica que considera al electrón como una combinación de onda y partícula. Los orbitales no son s o p, sino una combinación de ambos en la proporción 1:3.

Por supuesto, Pauling hizo algo más que una mera especulación sobre estas ideas. Realizó todos los cálculos, utilizando las reglas cuánticas, y predijo la fuerza de los enlaces resultantes. Ésta coincide con la fuerza medida experimentalmente. Además, si lo que se desea es una explicación física de la razón por la cual se ha de preferir esta disposición, hay que decir que se debe a que la estructura mantiene las cuatro nubes de electrones en la parte exterior del átomo de carbono, con una separación mutua máxima, la cual corresponde al estado de energía más bajo que es posible para dichas nubes.

Esta hibridación de orbitales sólo puede darse en el caso de aquellos que tienen todos la misma energía. Se trata de una compartición equitativa de electrones entre los estados energéticos equivalentes que pueden existir y es una característica fundamental de la química orgánica, tal como explicó Pauling.[8] Este proceso se da incluso en el ion amonio, que ya hemos mencionado con anterioridad. El nitrógeno es el elemento siguiente al carbono en la tabla periódica, y tiene cinco electrones exteriores. Por lo tanto, «tendría que» formar tres enlaces, aceptando electrones de otros átomos, que es lo que hace cuando se produce, por ejemplo, el amoniaco (NH_3). Sin embargo, si el átomo de nitrógeno perdiera uno de sus electrones, sería como un

átomo de carbono con una unidad de carga positiva añadida en su núcleo. En ese caso sería capaz de formar una estructura tetraédrica de cuatro enlaces, que es justamente lo que hace en el ion amonio (NH_4). Ésta es la razón por la cual el ion amonio actúa como una sola unidad en los compuestos en que está unido a otro átomo mediante enlace iónico.

La capacidad de hibridación está relacionada con la idea de resonancia, según la cual la estructura de una molécula se puede considerar como una oscilación entre dos o más estados diferentes, de tal manera que se produce una especie de promedio de dichos estados. Esto sólo es posible si los distintos estados tienen esencialmente la misma energía y son versiones diferentes del estado energético más bajo que exista.

El ejemplo más sencillo es el de la molécula de hidrógeno. Dijimos anteriormente que el enlace entre dos átomos de hidrógeno en una molécula (H_2) es un enlace covalente; la imagen que se dio fue la de un par de electrones que eran compartidos por dos núcleos para formar un solo enlace. Después ofrecimos la descripción de una imagen alternativa, en la que los dos núcleos estaban rodeados por una sola nube electrónica. En 1928, Pauling ofreció una tercera explicación (un tercer modelo), en la que el enlace es iónico, pero incluye resonancia. Si el primer átomo de hidrógeno cede su electrón enteramente al segundo, se quedará con una unidad de carga positiva, mientras que el segundo ha ganado una unidad de carga negativa, por lo que los dos iones se adherirán el uno al otro (H^+H^-). Sin embargo, puede suceder exactamente lo mismo desempeñando estos átomos los papeles inversos (con las cargas inversas), es decir, si el segundo átomo de hidrógeno cede su electrón al primer átomo de hidrógeno (H^-H^+). Pauling demostró que el enlace se puede describir como un rápido cambio entre estas dos posibilidades, es decir, como una resonancia.

Daremos sólo un par de ejemplos más sobre cómo actúa la resonancia, antes de aplicar esta teoría, y otras, para describir cómo funciona la vida. Recordemos el ion carbonato (CO_3). Este ion posee una carga global de menos dos unidades, ya que ha ganado dos electrones de cualquier átomo con el que se haya unido mediante enlace iónico. Así, teniendo en cuenta sólo las capas más externas, diremos que tiene cuatro electrones del átomo de carbono, seis de cada uno de los átomos de oxígeno y dos electrones suplementarios: un total de veinticuatro electrones para compartir entre cuatro átomos del modo más eficiente en cuanto al nivel de energía. ¿Cómo estarán situados estos electrones?

Recordemos que los estados químicos preferidos —los de baja energía— son aquellos en que cada átomo tiene una capa exterior que ya está llena con ocho electrones (o, al menos, la ilusión de una capa exterior llena). Uno de los modos en que podría conseguirse esto consistiría en que cada uno de los dos átomos de oxígeno se pegara a uno de los dos electrones suplementarios, cediéndoles siete electrones, y

que luego cada uno formara un único enlace covalente con el átomo de carbono, produciendo cada uno de ellos una compartición de ocho electrones en total. Esto dejaría al átomo de carbono con dos electrones no emparejados mediante los cuales formaría un doble enlace con el átomo de oxígeno que quedaba, de manera que ambos átomos se hicieran la ilusión de que tenían una capa exterior completamente llena con ocho electrones. No es nada extraño que el ion carbonato sea una unidad estable.

Pero, si esto fuera el final de la historia, el ion carbonato no sería simétrico. Habría una carga negativa global en un extremo, y uno de sus enlaces con el oxígeno (el doble enlace, que es, desde luego, más fuerte que un enlace sencillo) sería más corto que los otros dos. Sin embargo, hay tres modos de conseguir el mismo tipo de distribución de carga, según cuál de las tres moléculas de oxígeno sea la que forma el doble enlace y no tenga exceso de carga negativa. Debido a que los tres modos posibles tienen exactamente la misma energía, se produce la resonancia y se puede considerar que el ion carbonato cambia rápidamente de una opción a otra, recorriendo las tres variaciones sobre el tema. El efecto global debe ser que las dos unidades de carga negativa se distribuyen equitativamente por todo el ion, y la longitud y la fuerza de cada enlace entre los átomos de oxígeno y el átomo de carbono debe ser la misma, equivalente a uno y un tercio de los enlaces individuales normales, estando los átomos de oxígeno espaciados uniformemente en torno al átomo de carbono. Todo esto se puede calcular con precisión utilizando la versión de Pauling de la química cuántica y los resultados de los cálculos se corresponden muy de cerca con los resultados de las mediciones experimentales efectuadas mediante la espectroscopia y otras técnicas. La resonancia constituye realmente un buen modelo. Por otra parte, en toda la química orgánica no hay nada que sea más importante que la determinación de la naturaleza que tiene la estructura conocida como anillo de benceno.

Hemos dicho anteriormente que los átomos de carbono pueden formar enlaces con otros átomos de carbono. De esta manera se puede generar en ocasiones una larga cadena de átomos de carbono, que «se dan la mano» unos a otros, pero teniendo cada uno de ellos dos enlaces libres, que salen de la cadena para unirse con otros átomos o moléculas. Sin embargo, debido al ángulo natural que aparece entre los enlaces de carbono, también es posible que estos átomos de carbono formen un anillo de seis átomos, el anillo de benceno. Su nombre procede del de un compuesto llamado benceno (descubierto por Michael Faraday en 1825), que tiene sus moléculas compuestas por seis átomos de carbono y seis átomos de hidrógeno. Los químicos sabían esto mucho antes de que supieran algo sobre los enlaces químicos, sencillamente por las mediciones de la cantidad de carbono y de hidrógeno que se combinan para dar una cierta cantidad de benceno. A primera vista, esta mezcla de átomos de carbono e hidrógeno a partes iguales en una sola molécula resulta muy

extraña. Cada átomo de carbono tiene capacidad para unirse con otros cuatro átomos (tanto como esto ya estaba claro en la década de 1820), por lo que, incluso si seis de ellos están ligados en una cadena, aún habrá «sitio» para que se unan catorce átomos de hidrógeno, como ya sucede, en realidad, en un compuesto llamado hexano. ¿Dónde se ha ido entonces la capacidad de formar enlaces suplementarios?

La respuesta le llegó al químico alemán Friedrich Kekule en una especie de visión, lo que él llamó «un sueño que le hizo despertar» y que tuvo en Londres en 1865 mientras viajaba en un ómnibus tirado por caballos. Al soñar despierto vio cadenas de átomos de carbono que bailaban por todas partes y, de repente, una de las cadenas dio la vuelta haciendo un bucle y se agarró al otro extremo formando un círculo. Esta visión condujo a Kekule a la idea de que los átomos de carbono de un anillo de benceno están dispuestos en círculo (o más bien, haciendo un hexágono). Cada átomo de carbono utiliza tres de sus enlaces para mantener el círculo cerrado, formando un enlace doble en un lado y un enlace simple en el otro. Esto deja a cada uno de los seis átomos de carbono con sólo un enlace libre, que sale del círculo y sirve para agarrarse a un átomo de hidrógeno.

Fue un modo brillante de ver las cosas, pero todavía quedaba un misterio difícil de aclarar, hasta que la física cuántica llegó al rescate. La estructura propuesta por Kekule para el anillo de benceno tenía enlaces dobles y simples alternando alrededor de los lados del hexágono. Pero los enlaces dobles han de ser más cortos que los enlaces simples. Además, es relativamente fácil romper una de las conexiones de un enlace doble, liberando un enlace que puede tomar parte en otras interacciones. Sin embargo, el anillo de benceno no muestra ningún rasgo de asimetría y ninguno de los enlaces es más propenso que otro a ser liberado de esta manera.

Desde luego, la explicación de esto es la resonancia. No ha de sorprender a estas alturas saber que los seis enlaces existentes entre los átomos de carbono en un anillo de benceno tienen la misma fuerza, concretamente una vez y media la fuerza de un enlace simple ordinario. En realidad la estructura del anillo de benceno es una resonancia híbrida entre dos posibilidades, una con un número par de enlaces dobles (comenzando a partir de cualquier lugar elegido arbitrariamente en el anillo) y un número impar de enlaces simples, mientras que la otra posibilidad tiene los enlaces dobles y simples a la inversa. El resultado es una estructura muy estable que constituye la base de un número extraordinariamente grande de moléculas más complicadas, que se obtienen reemplazando los átomos de hidrógeno de uno o más de los vértices del hexágono, en la parte exterior del anillo, mediante grupos más complicados de átomos (como, por ejemplo, el grupo metilo, CH_3).

Los anillos también pueden unirse unos a otros, ligándose un hexágono con otro hexágono a lo largo de uno de los lados, como baldosas hexagonales que estén colocadas lado con lado, pero esto nos lleva realmente más allá de los dominios de la

química elemental, introduciéndonos ya en el ámbito de la vida.

Biomoléculas

La capacidad de los átomos de carbono para formar anillos y largas cadenas (a veces, largas cadenas que llevan anillos incorporados a su estructura) es una característica fundamental de las moléculas en las que se basa la vida en la Tierra. Estas cadenas pueden ser realmente muy largas, pero bastará un par de ejemplos sencillos para dar una visión del tipo de química de que se trata en estos casos.

En una cadena de carbono, debido al ángulo que forman entre sí los enlaces químicos, resulta que los átomos de carbono que constituyen la columna vertebral de la cadena describen una especie de zigzag y hay otros átomos (o grupos de átomos) que sobresalen en los vértices de dicho zigzag, donde están unidos a los átomos de carbono. Un compuesto bastante sencillo de este tipo es el que tiene seis átomos de carbono en la cadena, cada uno de ellos unido a dos átomos de hidrógeno que sobresalen de la columna vertebral. En cada extremo de la molécula el último átomo de carbono está unido a lo que se llama un grupo «amina», formado por un átomo de nitrógeno y dos átomos de hidrógeno (NH_2). Este compuesto se denomina hexametilén-diamina («di» que quiere decir «dos», amina por el NH_2 y «hexametilén» porque contiene seis átomos de carbono con sus dos átomos de hidrógeno cada uno).

Una molécula ligeramente más corta y con el mismo tipo de estructura es el ácido adípico, que tiene sólo cuatro átomos de carbono en la cadena y un grupo COOH unido a cada extremo. Si una molécula de ácido adípico se encuentra con una molécula de hexametilén-diamina, es fácil (porque conduce a un estado de menos energía) que el ácido adípico libere un grupo OH de un extremo, mientras que el grupo NH_2 de un extremo de la hexametilén-diamina libera un átomo de hidrógeno, H . Los átomos liberados se unen formando una molécula de agua (H_2O), y los dos enlaces que quedan libres se unen a través del hueco que han dejado los átomos liberados, de tal modo que el átomo de oxígeno de un extremo de la cadena se une con el átomo de carbono del extremo de la otra cadena. Esto da como resultado una nueva cadena, de once átomos de longitud, con una subcadena de cuatro átomos de carbono y otra subcadena de seis átomos de carbono unida con un átomo de oxígeno a través del hueco.

Si se mezclan muchas moléculas de ácido adípico y muchas de hexametilén-diamina, el proceso se repetirá, añadiéndose a cada extremo de la cadena nuevos fragmentos de la sustancia correspondiente, para formar un cordón muy largo, en el que esta columna vertebral básica de once átomos y los átomos que van unidos a ella se repiten una y otra vez. El resultado, que puede contener miles de estas unidades básicas, constituye un ejemplo de una clase de moléculas de cadena larga llamadas polímeros. Este polímero en concreto es corriente y práctico a la vez; se conoce

habitualmente como nailon. También es especialmente sencillo, ya que la mayoría de los átomos que sobresalen en la columna vertebral son de hidrógeno, aunque éste se puede reemplazar mediante estructuras más complejas tales como el grupo amina, anillos de benceno u otras cadenas.

Los propios anillos de benceno son unas estructuras bastante planas, pero presentan una ondulación en el anillo producida por el ángulo existente entre los enlaces. Por otra parte, debido al ángulo exacto de los enlaces químicos del carbono, los átomos o grupos de átomos que están ligados por todo el anillo —cosa que puede ocurrir cuando algunos de los enlaces dobles de la estructura del benceno están rotos y libres para interactuar con otros átomos— se encuentran por encima y por debajo del plano del anillo, configurando una estructura en tres capas que recuerda vagamente a un bocadillo. También existen variaciones sobre el tema en las que uno de los átomos de carbono del anillo ha sido reemplazado por un átomo de otro tipo. En algunas de las variaciones más sencillas de este tipo, la mayoría de los átomos de carbono del anillo están cada uno de ellos ligados por un lado a un átomo de hidrógeno y por el otro a un grupo llamado hidróxilo (OH), con una alternancia de los lados en que se sitúa cada una de estas ligaduras a lo largo de todo el anillo. H y OH unidos forman agua, por supuesto, por lo que esta clase de compuestos se conocen comúnmente como carbohidratos o hidratos de carbono, lo cual significa «carbono con agua».

Los carbohidratos más sencillos se denominan azúcares. El más sencillo de todos ellos, la glucosa, es un anillo como el que acabamos de describir, con cinco átomos de carbono y uno de oxígeno. Cuatro de los átomos de carbono están ligados lateralmente a O y OH. El quinto tiene a un lado un átomo de hidrógeno y al otro lado sobresale un grupo más complejo (CH_2OH).

Es muy fácil para estos anillos unirse entre sí. Si dos grupos OH, cada uno de un anillo distinto, se unen, uno de ellos será liberado y se unirá al hidrógeno del otro grupo OH para dar H_2O . Esto deja al átomo de oxígeno de la segunda molécula de glucosa con un enlace libre que le servirá para acoplarse al lugar de donde ha sido desplazado el grupo OH de la primera molécula de glucosa. Dos anillos del tipo de la glucosa unidos de esta manera mediante un átomo de oxígeno forman una molécula de otro azúcar, la maltosa.

También es posible hacer una molécula con forma de anillo de cinco lados, justo como la estructura de la glucosa pero quitándole uno de los átomos de carbono y los grupos laterales que éste lleva acoplados. Esta molécula se llamaría ribosa. Otra molécula con forma de anillo de cinco lados, idéntica a la ribosa salvo que uno de sus grupos OH ha perdido el átomo de oxígeno, quedando un simple átomo de hidrógeno, se conoce como desoxirribosa, lo cual significa «ribosa de la que se ha eliminado oxígeno». Es la unidad básica que da su nombre a aquella que entre todas las

biomoléculas es la más importante por lo que respecta a los seres humanos: el ácido desoxirribonucleico, o ADN.[9]

Sin embargo, por lo que se refiere al peso (y dejando a un lado el agua, que constituye más de las tres cuartas partes del peso de los seres vivos, por ejemplo, nosotros mismos), las moléculas más importantes de nuestro cuerpo son las proteínas. Muchas de las proteínas son moléculas muy grandes y complejas; pero, como todas las biomoléculas complejas, están formadas por unidades más sencillas y subunidades, del mismo modo que el nailon y la maltosa están formados por componentes más sencillos. La complejidad de la estructura de biomoléculas tales como las proteínas, y sus facultades en cuanto a mantener la vida, se han estructurado y refinado durante vastos períodos de tiempo geológico (miles de millones de años) por efecto de la evolución. La selección natural ha adaptado estas moléculas para la realización de tareas específicas, basadas en el modo en que forman enlaces químicos. Dentro de poco veremos más sobre la evolución. Aquí queremos concentrarnos en lo que son estas estructuras y en cómo llevan a cabo sus tareas.

Además del carbono, que es el átomo más importante de las biomoléculas, éstas están provistas de gran cantidad de nitrógeno, otro átomo que puede formar varias conexiones interesantes (habitualmente tres a la vez) con otros átomos. En las rocas de la corteza terrestre, los elementos más frecuentes son el oxígeno (47 por 100 del peso), la sílice (28 por 100) y el aluminio (8 por 100). En nuestro cuerpo la mitad del peso seco es carbono, el 25 por 100 es oxígeno y algo menos del 10 por 100 es nitrógeno. Las proteínas contienen una proporción de nitrógeno aún mayor que la que contiene la totalidad del cuerpo (alrededor del 16 por 100 en peso).

La razón de esto es que todas las proteínas están formadas por unas unidades más simples conocidas como aminoácidos, y todos los aminoácidos contienen nitrógeno. Los aminoácidos tienen todos ellos la misma estructura básica, con cuatro grupos de átomos acoplados con los cuatro enlaces de un átomo de carbono. Uno es el grupo amina (NH_2), que da su nombre a los aminoácidos. Otro es el grupo del ácido carboxílico (COOH), que hace que las moléculas sean ácidas. Un tercero es siempre un átomo de hidrógeno aislado. El cuarto enlace del carbono puede acoplarse a toda una variedad de distintos grupos químicos, dando a los diferentes aminoácidos sus características específicas.

Los químicos sabían que las proteínas estaban formadas por aminoácidos mucho antes de que conocieran en profundidad la estructura detallada de las moléculas complejas, porque, cuando las proteínas se hierven en una disolución fuerte de un simple ácido o una simple base[10] (inorgánicos), se rompen los enlaces químicos que unen los aminoácidos entre sí en una cadena, y lo que queda es una sopa de aminoácidos que se puede investigar por medios químicos convencionales (aunque a veces tediosos).

Los aminoácidos comunes se han denominado según las sustancias en las que se descubrieron inicialmente, o se les han dado nombres que reflejan alguna propiedad obvia del propio aminoácido. El primer aminoácido que se identificó fue extraído del espárrago en 1806, por lo que se le dio el nombre de ácido aspártico. Otro aminoácido, extraído de la gelatina en la década de 1820, resultó ser de sabor dulce y se denominó glicina, de la palabra griega que significa «dulce» (cabe resaltar que éste es el aminoácido más sencillo, ya que tiene un solo átomo de hidrógeno unido al enlace de carbono «sobrante»).

En teoría podría existir una enorme variedad de aminoácidos y muchos de ellos en realidad los han fabricado los químicos. Sin embargo, por lo que concierne a los seres vivos, sólo veintitrés aminoácidos son importantes. Veinte de éstos se encuentran en todas las proteínas y otros dos están presentes en unas cuantas. El vigésimo tercero es de hecho una versión diferente de uno de los veinte primeros. En una molécula de cisteína el núcleo básico del aminoácido está acoplado mediante su enlace libre a otro átomo de carbono, que a su vez está acoplado a dos átomos más de hidrógeno y a un único átomo de azufre, el cual sobresale del extremo de la molécula y tiene además otro átomo de hidrógeno en el otro extremo. El átomo de hidrógeno que pende de una molécula de cisteína se puede combinar muy fácilmente con el átomo de hidrógeno que pende de otra molécula de cisteína, siendo posible así que escapen moléculas de hidrógeno (H_2) y los dos átomos de azufre quedan adheridos entre sí, formando lo que se conoce como un enlace disulfuro. El resultado es una molécula llamada cistina.

El mismo tipo de acoplamiento puede suceder entre otros aminoácidos, incluso entre aminoácidos de tipos diferentes. En el caso de la cisteína/cistina, los dos grupos que identifican al aminoácido se unen; pero en otros casos el grupo amina de un aminoácido puede acoplarse al OH del COOH de otro aminoácido, liberando agua y quedando el átomo de nitrógeno en esta ocasión para formar un puente (conocido como enlace peptídico) entre los dos aminoácidos residuales. A ambos lados de la molécula resultante se pueden acoplar de la misma manera nuevos enlaces, para formar una cadena conocida como polipéptido. Se trata de otra estructura en zigzag, cuya columna vertebral está formada por una pauta repetitiva de dos átomos de carbono, seguidos de un átomo de nitrógeno, dos átomos más de carbono, otro átomo más de nitrógeno, y así sucesivamente. Además, a lo largo de la cadena se puede acoplar una gran variedad de añadidos (incluso estructuras en forma de anillo), dependiendo de cuáles son los aminoácidos que se han unido para formar el polipéptido.

Una de las características diferenciadoras de este tipo de cadena es que el enlace peptídico, en el que un átomo de nitrógeno aparece ligado a un átomo de hidrógeno y a uno de carbono que siempre está acoplado a un átomo de oxígeno mediante un

enlace único (no importa mucho a qué se una el tercer enlace del nitrógeno), es una estructura rígida que se mantiene mediante la resonancia cuántica mecánica. La cadena entera puede girar en torno a los otros enlaces, pero el enlace peptídico no se puede retorcer. Como resultado, las cadenas polipépticas sólo pueden enrollarse de determinadas maneras [11] haciendo un ovillo (como ovillos de cuerda) para formar estructuras compactas. Aplicando los principios de la mecánica cuántica a estas estructuras, Linus Pauling consiguió averiguar cómo se enrollan las proteínas y pudo también analizar su estructura, abriendo así el camino para que otros investigadores estudiaran las biomoléculas por este procedimiento.

Una de las propiedades más importantes de una de estas cadenas se puede ver imaginándola estirada e ignorando el orden de la aburrida pauta repetitiva (CCNCCNCCN...) de los átomos a lo largo de la columna vertebral. La característica diferenciadora de una cadena polipéptica concreta (en particular, una proteína) se ve entonces como el orden de las distintas subunidades existentes a lo largo de la cadena, es decir, de los grupos acoplados a los laterales de la columna vertebral. Estos acoplamientos, conocidos a veces como radicales, son los fragmentos de los distintos aminoácidos que les dan su personalidad individual. A su vez, es el orden en que se encuentran estos radicales a lo largo de la cadena lo que les da a las proteínas su personalidad individual, y hace que estén enrolladas de maneras determinadas, permitiéndoles (o forzándolas a) tomar parte en reacciones químicas específicas, y además también impidiéndoles que interaccionen de otras maneras.

La variedad de las proteínas que se pueden generar a partir de unos veinte aminoácidos es enorme. Se puede hacer una comparación con el alfabeto de la lengua inglesa. En este alfabeto, con sólo veintiséis caracteres se puede construir un número enorme de palabras, incluidas todas las de este libro. Si cada aminoácido es equivalente a una letra de una palabra, el número de proteínas que se pueden hacer combinando los enlaces en la cadena mediante el alfabeto de los aminoácidos es mayor que el número de palabras que puedan aparecer en el mayor diccionario de la lengua inglesa, ya que las cadenas polipépticas pueden ser mucho más largas que una palabra inglesa normal (la hemoglobina, una molécula de proteína de tamaño medio, tiene un peso molecular igual a aproximadamente 67.000). Sin embargo, hasta donde sabemos, sólo una minoría de todas estas opciones se utiliza realmente en las proteínas que son tan importantes para la vida.

Existen dos líneas en la historia de la investigación de las proteínas. Una línea de ataque incluye la determinación de la estructura física de las moléculas (cómo se pliegan las cadenas polipépticas); la otra línea incluye la identificación de las subunidades de aminoácidos y su orden a lo largo de la cadena en cada proteína específica.

La primera persona que utilizó los rayos X para estudiar la estructura de los

cristales (al principio, estructuras muy sencillas, como los cristales de la sal común) fue Lawrence Bragg en 1912. Considerando los rayos X como ondas electromagnéticas, el modo en que estas ondas rebotan en los átomos de un cristal hace que las ondas interfieran entre sí, como las ondulaciones del agua en una charca, y la pauta de estas interferencias revela los detalles de la estructura del cristal. Bragg inventó la cristalografía mediante rayos X y recibió en 1915 el Premio Nobel de Física junto con su padre, William Bragg, por su trabajo conjunto en este campo. En la década de 1920 Lawrence Bragg desarrolló un conjunto de reglas para interpretar las pautas de rayos X producidas por cristales más complicados. Pero Linus Pauling, que trabajaba al otro lado del Atlántico, desarrolló las mismas reglas y fue el primero en publicarlas en 1929. Esto fue el comienzo de una larga y no siempre amistosa rivalidad entre el equipo de Bragg y el equipo de Pauling.

El paso siguiente consistió en aplicar la técnica de los rayos X para comprobar la estructura de las biomoléculas. Fue lo más natural comenzar por las proteínas, que, como ya hemos visto, son más comunes que cualesquiera otras biomoléculas. Las proteínas se presentan en dos variedades básicas: por una parte, estructuras largas y estrechas que mantienen el tipo de estructura alargada que se asocia generalmente con una cadena (el cabello es un buen ejemplo) y, por otra parte, estructuras globulares, en las que la cadena proteica básica está enroscada formando una bola.

Las primeras imágenes de la pauta de difracción por rayos X de una proteína fibrosa, una del grupo conocido como las queratinas, las obtuvo William Astbury (un antiguo discípulo de William Bragg) en la Universidad de Leeds a principios de la década de 1930. Las queratinas se encuentran en la lana, el cabello y en las uñas de nuestros dedos. Astbury descubrió que existe una pauta regular y repetitiva en la imágenes obtenidas mediante rayos X, lo cual significa que existe una estructura regular y repetitiva en la queratina; o, más bien, dos pautas distintas repetitivas, una correspondiente a fibras no estiradas (llamó a éstas alfa-queratina) y la otra a fibras estiradas (beta-queratina). Aunque la técnica no era aún lo suficientemente buena como para revelar la estructura exacta de las moléculas contenidas en la queratina, al menos servía para restringir las opciones posibles, descartando muchas posibilidades. Todo esto estimuló a varios científicos (especialmente los equipos liderados por Lawrence Bragg en Cambridge y por Linus Pauling en Caltech) para intentar hallar un modo de enrollar la cadena de una proteína con el fin de que encajara en las imágenes por rayos X.

Esto llevó mucho tiempo, en parte porque los investigadores tuvieron que retroceder hasta lo más básico, examinando la estructura de los enlaces existentes entre los distintos aminoácidos y averiguando cómo podían (o no podían) darle vueltas al asunto, y en parte a causa de la Segunda Guerra Mundial, durante la cual esta investigación prácticamente se detuvo. Después de la guerra, la técnica de los

rayos X se volvió mucho más precisa y esto permitió a los buscadores de modelos dar el empuje final necesario para determinar la estructura de este tipo especial de proteína. Fue Pauling el que ganó la carrera, hallando la estructura de la alfa-queratina y publicando una serie de trabajos en 1951 en los que explicaba cómo se unen las moléculas de proteína para producir estructuras aparentemente tan diferentes como cabello, plumas, músculos, seda y asta. La estructura que había descubierto este equipo fue denominada alfa-hélice, siendo una característica fundamental de dicha estructura básica el modo en que el enlace peptídico se mantiene rígido. Otra razón para la estabilidad de la alfa-hélice es que, en esta disposición especial de la cadena polipeptídica, el grupo NH de un enlace peptídico, gracias a su rigidez, está exactamente en el lugar adecuado para que el átomo de hidrógeno anide a lo largo del átomo de oxígeno en el enlace peptídico cuatro átomos de carbono más abajo en la cadena, de tal manera que puede formar un enlace de hidrógeno con el átomo de oxígeno. Cada uno de los enlaces peptídicos de la alfa-hélice está acoplado de esta manera con un vecino, y esto explica por qué tiene la molécula esa estructura repetitiva característica que muestra en sus personales imágenes de difracción de los rayos X.

Los diferentes tipos de queratina son el resultado de unas sutiles diferencias en la disposición de los aminoácidos específicos contenidos en las cadenas. Por ejemplo, en los tipos de queratina dura que forman sustancias tales como las uñas de nuestros dedos, hay muchos componentes de cisteína. Pero cuando dos moléculas de cisteína se tocan, como ya hemos visto, liberan hidrógeno y se unen mediante un puente de disulfuro, un auténtico enlace covalente. En las hileras de proteínas duras, las alfa-hélices se sitúan una al lado de la otra, fuertemente ligadas entre sí de esta manera, para formar resistentes capas de materia.

En el cabello los enlaces disulfuro operan de un modo ligeramente diferente, pues mantienen conjuntos de tres alfa-hélices entrelazadas y juntas, del mismo modo que unas tiras de cuerda se pueden retorcer juntas para hacer una cuerda más fuerte. Cuando al cabello se le aplica un tratamiento que rompe los puentes de disulfuros, éste se vuelve blando y se puede rizar fácilmente para darle un aspecto distinto. Después se puede tratar químicamente para restablecer los puentes de disulfuro y se quedará con su nuevo aspecto, de hecho, así es como los peluqueros realizan lo que se llama una «permanente».

En las beta-queratinas las cadenas polipeptídicas no forman hélices, sino que están en zigzag una a lo largo de la otra. En vez de tener enlaces de hidrógeno que se establecen dentro de una cadena, para mantener el aspecto helicoidal los enlaces se forman de un modo similar entre fibras proteicas vecinas, produciendo una estructura mucho más suave. Precisamente uno de los más conocidos ejemplos de esta estructura es famoso por su suavidad: la seda.

Siguiendo el éxito de Pauling con la alfa-hélice, la idea de buscar estructuras helicoidales en las moléculas biológicas disparó la imaginación de otros investigadores. La mayor recompensa fue conseguir determinar la estructura del ADN, del que se sabía por entonces (principios de la década de 1950) que era la molécula que llevaba la información genética de una generación a la siguiente. El ADN se encuentra en el núcleo de las células de los organismos vivos, por lo que se llama una nucleína. Es un compuesto ligeramente ácido, de ahí su nombre de ácido nucleico, y contiene desoxirribosa, por lo que se denomina ácido desoxirribonucleico, dando las siglas ADN.

La función principal del ADN en los procesos vitales comenzó a estudiarse con detalle a finales de los años veinte, realizándose en aquel tiempo investigaciones sobre el modo en que actúa la bacteria causante de la neumonía. Después de una enorme cantidad de duro y esmerado trabajo, en 1944 quedó claro ya que las diferencias entre los distintos tipos de bacterias de la neumonía eran producidas por desigualdades en el ADN de sus células. Dicho de otro modo, era el ADN lo que hacía diversas a las bacterias. Dado que las células de todos los seres vivos contienen ADN y que mucho tiempo antes se había constatado que aquello que hace a las especies distintas unas de otras está en sus células —en realidad está envuelto en los núcleos de las células de la mayoría de las especies, incluidos nosotros mismos— estaba claro que el ADN contenía el secreto de la propia vida. Pero ¿qué era exactamente el ADN? ¿Cómo se enrollan sus moléculas dentro del núcleo de las células? ¿Cómo hace el ADN para transmitir información de una generación a la siguiente?

El gran avance decisivo partió de dos investigadores del laboratorio de Bragg en Cambridge, Francis Crick y James Watson. Al final, el grupo de Cambridge había vencido al de Pauling, lo que constituyó una fuente de gran satisfacción en aquella época, al menos en Cambridge, aunque esto hoy en día parezca sólo una anécdota sin importancia dentro de la historia. Lo hicieron utilizando el mismo planteamiento que habían aplicado Bragg y Pauling para las proteínas: una combinación de fotografías por difracción de rayos X realizadas para hacerse una idea sobre el tipo de estructuras de que se trataba, y haciendo modelos para averiguar cómo podrían encajar juntos los distintos componentes de la molécula que estaban estudiando.

Hacia la década de 1930, algunos químicos que trabajaban en química orgánica habían conseguido averiguar cuáles eran las componentes del ADN, a pesar de que en aquella época todavía no eran plenamente conscientes del papel que desempeñaba este compuesto en los procesos biológicos; así, durante las tres primeras décadas del siglo xx, se aceptaba en general que dentro del núcleo el ADN actuaba como una especie de andamiaje para algunas moléculas proteicas que eran consideradas como las auténticas biomoléculas. Una molécula de ADN está formada exactamente por

tres componentes, pero hay muchos ejemplos de cada tipo de componente dentro de una sola molécula de ADN. El primer tipo de componente es el azúcar llamado ribosa desoxigenada o desoxirribosa, que otorga al ADN su nombre y consiste en una molécula de cinco lados que contiene cuatro átomos de carbono y un átomo de oxígeno, formando todos ellos un anillo. El segundo es un grupo de átomos conocido como grupo fosfato, que consta de un átomo de fósforo rodeado por cuatro átomos de oxígeno.[12] Y el tercero es un tipo de componente que se denomina base, aunque existen cuatro bases diferentes que están presentes en las moléculas de ADN (adenina, citosina, guanina y timina, a las que con frecuencia se alude sencillamente indicando sus letras iniciales).

Hacia la mitad de la década de 1930 estaba claro que el ADN se podía descomponer en unidades cada una de las cuales contenía una molécula de azúcar, un grupo fosfato y una base. Cada subunidad de este tipo se conoce con el nombre de nucleótido y parecía lógico suponer que estas subunidades estaban conectadas unas a otras para formar una cadena, de manera similar al modo en que están conectadas entre sí las subunidades de aminoácidos para formar una cadena en las moléculas proteicas. Estamos hablando de cadenas muy largas; sabemos actualmente que existen millones de átomos en una sola molécula de ADN (pero hay que tener presente que todos esos millones de átomos se presentan en exactamente cinco variedades: carbono, nitrógeno, oxígeno, hidrógeno y fósforo, que están dispuestos formando unas pautas realmente interesantes). Pero ¿cómo estaban dispuestos los nucleótidos para hacer una molécula de ADN?

Un grupo de bioquímicos que trabajaban en Cambridge dirigidos por Alexander Todd demostraron a finales de los cuarenta que los nucleótidos estaban en realidad conectados en una cadena de una manera específica. El esqueleto de la cadena está constituido por azúcar y grupos fosfato en orden alternante. Los azúcares y los grupos fosfato están conectados entre sí, de tal manera que un grupo base sobresale por el lateral de cada azúcar. Fue esta información fundamental la que utilizaron Crick y Watson, junto con imágenes obtenidas por difracción de rayos X, para descubrir la estructura global del ADN en 1953.

Las primeras fotografías del ADN obtenidas por difracción de rayos X las consiguió William Astbury en 1938. Sin embargo, se produjo una interrupción de los trabajos (de nuevo, causada en parte por la Segunda Guerra Mundial) hasta que, a principios de la década de 1950, un equipo del King's College de Londres retomó otra vez la investigación sobre la estructura del ADN siguiendo la misma línea. Las fotografías que fueron clave para aportar la información que Crick y Watson necesitaban las realizó una joven investigadora, Rosalind Franklin. Franklin habría resuelto por sí misma el problema de la estructura si Crick y Watson no le hubieran ganado por la mano[13] pero, debido a que murió joven (en 1958), no recibió su

parte del Premio Nobel que se otorgó a Crick y a Watson en 1962, y por consiguiente el papel que desempeñó Rosalind Franklin en toda esta historia no siempre se trata como merece.

Las fotografías realizadas mediante rayos X mostraban que la estructura de las moléculas del ADN tenían que ser helicoidales. La idea clave por la que el equipo de Cambridge ganó el Premio Nobel tenía dos componentes. En primer lugar, que la pauta que seguía la difracción significaba que la molécula era una hélice doble, con dos hebras retorcidas una en torno a la otra (una idea que también Franklin tuvo y mencionó en sus anotaciones). La segunda componente era que la disposición de las bases a lo largo de la columna vertebral de la molécula de ADN constituye para las dos hebras de la hélice un modo natural de mantenerse unidas, del mismo modo que la estructura de la alfa-hélice de una proteína se mantiene rígida gracias a los enlaces de hidrógeno.

Todo esto depende de la estructura detallada de las moléculas de las distintas bases. La timina y la citosina están hechas cada una de ellas de un anillo hexagonal en el que hay cuatro átomos de carbono y dos átomos de nitrógeno, junto con otras sustancias (hidrógeno, grupos metilo y otros) acopladas en torno a los lados del anillo. La adenina y la guanina están formadas cada una de ellas por un anillo hexagonal similar y este anillo se une a lo largo de uno de sus lados (como cuando dos baldosas se unen lado con lado) con un anillo pentagonal, que es como el primer anillo pero con un átomo menos de carbono. Por otra parte, hay otras sustancias, pero muy pocas, tales como el hidrógeno y los grupos amino, que se unen por los lados. En una aproximación, A y G tienen un diámetro que es el doble que el de C y T.

Si dos hebras distintas de ADN se colocan lado con lado, con las bases en el medio y alineadas de tal forma que allí donde en la primera hebra había una A, resulta que en la segunda siempre había una T, mientras que allí donde en la primera hebra había una C, resulta que en la segunda había una G. La cantidad de espacio ocupado por cada par de bases (AT y CG) sería el mismo, por lo que no ha de haber arrugas en la cadena.[14]

Si alineamos las moléculas de esta manera, sucederá una cosa que nos llamará la atención: la forma de las moléculas A y T será justamente la adecuada para que se formen entre ellas dos enlaces de hidrógeno. Además la forma de las moléculas C y G será la indicada para que se formen entre ellas tres enlaces de hidrógeno. Las dos hebras del ADN se mantendrán unidas mediante los enlaces de hidrógeno a lo largo de toda su extensión, encajando A con T al encontrarse una contra otra, y C con G de la misma manera, como una llave que encaja en la cerradura, o como juegos de clavijas eléctricas de dos púas y tres púas que encajan en sus enchufes respectivos. El puente TA es exactamente del mismo tamaño y de la misma forma que el puente CG. El esquema que tenemos ahora es bastante parecido al de los raíles paralelos de una

línea ferroviaria, unidos y separados por piezas cruzadas, como traviesas de la vía del ferrocarril. Aplicándole una torsión a esta estructura de dos hebras, obtenemos una hélice doble de ADN: la doble hélice.

Desde luego, esto no sucede por casualidad. La estructura del ADN no está formada por dos hebras aleatorias que se asocian al azar, sino que está construida a partir de unidades de nucleótidos, de tal forma que queda garantizado que A siempre está situada en posición opuesta a T y G siempre está en posición opuesta a C. El mejor modo de ver cómo sucede esto es observar el proceso en el que se copia el ADN cuando una célula se reproduce. No le supone un gran esfuerzo a la maquinaria química de la célula romper los enlaces de hidrógeno (que, como se recordará, son relativamente débiles) en un extremo de la hebra de ADN y comenzar a desenrollarla. Cuando se produce esto, cada extremo libre del ADN se adherirá de forma natural a los socios adecuados que encuentre en la sopa química de materia que rodea a la célula. Allí donde una hebra de ADN presenta una A se adherirá a un nucleótido T que se encuentre en los alrededores; donde presente una T, agarrará a una A que pase por allí. Donde se presente una C, se enganchará con una G (y viceversa). Cuando la doble hélice original se desenreda, cada una de las dos hebras forma una nueva hebra para que sea su pareja, y lo hace paso a paso a lo largo de la cadena, actuando como una plantilla.[15] Al llegar el momento en que se ha desenredado por completo, cada hebra ya ha finalizado la construcción de una nueva pareja, con lo que se consiguen dos moléculas de ADN idénticas donde sólo había una. Cuando la célula se divide, una copia de cada molécula de ADN se transmite a cada una de las dos hijas, con lo que la vida sigue adelante.

Según los trabajos de Crick y Watson, también era obvio de una manera inmediata que las secuencias de las bases a lo largo de una hebra de ADN (la pauta de las «letras» A, C, G y T) podía transmitir información, como las letras del alfabeto, o la secuencia de aminoácidos a lo largo de una proteína. A primera vista, un alfabeto con sólo cuatro letras podría parecer restringido. Pero tanto el código Morse como los ordenadores utilizan un alfabeto binario en el que existen únicamente dos letras (el punto y el guión en Morse; conectado y desconectado, es decir, «on» y «off», en el alfabeto binario de los ordenadores). Se puede expresar lo que se desee en un alfabeto binario (no digamos en un alfabeto de cuatro letras), con tal de que las palabras puedan ser lo suficientemente largas. Así, el ADN que se encuentra en el núcleo de una célula de nuestro cuerpo lleva una gran cantidad de información: una descripción completa de la construcción, el cuidado y el mantenimiento de un cuerpo humano.

El modo en que el ADN trabaja consiste en transmitir información a la maquinaria celular sobre cómo fabricar proteínas. Resulta que el código genético está escrito en realidad con palabras cuya longitud es sólo de tres letras (tres bases), porque todo lo que cada palabra tiene que hacer es especificar un aminoácido

concreto. Así, la secuencia de bases existente a lo largo de un breve segmento de una hebra de ADN podría escribir un mensaje como por ejemplo ACG TCG TCA GGC CCT. Este mensaje indica a la maquinaria celular que disponga cinco aminoácidos concretos en un cierto orden cuando construya una cadena proteica. Dado que un alfabeto de cuatro letras se puede utilizar para construir sesenta y cuatro palabras diferentes en las que aparecen tres letras, no hay ningún problema para codificar de esta manera veintitantos aminoácidos, ni tampoco para hacer unos pocos signos de puntuación especiales, como la palabra de tres letras que significa «detener» (como cuando se dice «detener la formación de la proteína ahora»). En efecto «detener» es un mensaje tan importante que existen tres modos de codificarlo: UAA, UAG y UGA (la U representa al uracil; explicaremos su presencia más adelante).

Desde luego es mucho más fácil decir «la célula fabrica cadenas polipépticas a partir de aminoácidos de acuerdo con las instrucciones codificadas en el ADN», que hacer realmente las proteínas y, además, aún no se ha llegado a entender perfectamente la totalidad del proceso. Sin embargo, sus pasos clave están bastante claros.

Lo que en realidad sucede cuando hay que fabricar una cierta proteína (y uno de los enigmas es precisamente cómo sabe la célula cuándo es necesario) es que la parte importante del ADN que está enrollado en el núcleo de la célula se despliega y se copia el mensaje correspondiente, utilizándolo como una plantilla y formando una cadena de nucleótidos a lo largo de la hebra desplegada para fabricar otra hebra del ácido nucleico. Esta nueva hebra no es en realidad de ADN, sino ARN, una molécula en forma de cadena larga, casi idéntica, pero en la cual la unidad de azúcar es la ribosa, no la desoxirribosa, y que posee otra base, el uracil, en vez de la timina. En el ARN sigue habiendo solamente cuatro bases, pero en todos aquellos lugares en que habría una T en el ADN, hay una U en el ARN (que es la razón por la cual la U aparece en el código genético; véase más arriba).

Después, la hebra de ARN fabricada de esta manera (llamada ARN mensajero) sale del núcleo de la célula pasando a la sopa química que está alrededor y que constituye el principal volumen de la célula, donde una estructura llamada ribo-soma se pone en marcha. El ribosoma se desplaza a lo largo de la hebra del ARN mensajero, leyéndolo igual que la cabeza magnética de un magnetofón lee la cinta que pasa por ella. Es el ribosoma el que interpreta cada palabra de tres letras del mensaje codificado en forma de aminoácido y reúne a los aminoácidos en el orden correcto para fabricar la proteína específica. Luego, el ARN mensajero se rompe de tal forma que sus componentes pueden reutilizarse en otra ocasión.

Mientras algunos biólogos moleculares abordaban la cuestión del código genético y averiguaban cómo funciona todo esto (sorprendentemente, el asunto duró hasta bien entrada la década de 1960), otros estaban descifrando las estructuras de otras

proteínas, incluidas las proteínas globulares que se hacen enrollando cadenas polipépticas. Las proteínas hacen casi todo lo que hay que hacer en nuestro cuerpo, incluso proporcionar la estructura del mismo. Por lo que respecta a los procesos biológicos, el papel más importante lo desempeñan unas proteínas llamadas enzimas, que son todas ellas proteínas globulares. Existen moléculas que incitan a otras moléculas a interactuar de distintas maneras. Utilizando la terminología de la química, se dice que actúan como catalizadores.

La importancia de las enzimas se puede explicar mediante un sencillo ejemplo. Imaginemos una molécula grande y aproximadamente esférica (la enzima) que tiene en su superficie dos oquedades de forma irregular. Una de las oquedades tiene justo la forma adecuada para contener otra biomolécula de tamaño menor (de la misma manera que una pieza de forma irregular que encaja en un rompecabezas); la otra oquedad tiene la forma adecuada para contener otra biomolécula diferente. Cuando las dos moléculas ocupan estos huecos, se encuentran alineadas una con otra de tal modo que se pueden formar enlaces químicos entre ellas. Así, las dos se emparejan y son enviadas al interior de la célula como una sola unidad, con el fin de que realicen la tarea bioquímica que tengan que hacer.

Es importante indicar que la enzima permanece invariable después de llevar a cabo todo esto y puede repetir la jugada una y otra vez. Las enzimas son como unos robots perseverantes que repiten la misma tarea química indefinidamente (me recuerdan siempre a las pequeñas escobas de Fantasía, que transportaban interminablemente sus pequeños cubos de agua). Algunas enzimas unen moléculas (incluso las cadenas polipépticas) y otras las descomponen, dando un corte con sus tijeras químicas en las uniones entre aminoácidos. Algunas transportan moléculas hasta donde éstas son necesarias o retiran productos de desecho y los llevan al vertedero correspondiente; también las hay que incluso transportan energía de un lugar a otro.

Las enzimas son importantes, pero son sólo un componente más del cuerpo y, como sucede con todo lo demás, su estructura está codificada mediante el alfabeto de cuatro letras del código genético, guardado bajo llave en el ADN que se encuentra en el interior de las células del cuerpo. Antes de que nos vayamos a un nivel de tamaño superior para ver cómo interactúan los cuerpos entre sí y con el medio ambiente en el que están inmersos, quisiéramos echar un breve vistazo a otro papel que desempeña el ADN: cómo se copia (de un modo ligeramente diferente a la división de una célula ordinaria en dos) y se transmite a la siguiente generación.

Hasta ahora nos hemos referido al núcleo de la célula y a su parte exterior, que es más amplia, sin explicar ningún detalle de la estructura de la célula. Dado que ya hemos utilizado el término «núcleo» de un modo similar para describir la estructura del átomo, la esencia de su significado es evidente aunque no hayamos especificado

ningún detalle. De hecho, Ernest Rutherford dio este nombre al núcleo del átomo precisamente porque este término ya se había utilizado en un contexto similar en la biología celular y deseaba hacerse eco de esta terminología. Sin embargo, con el fin de abordar el tema de la reproducción, deberíamos atar cabos explicando con un poco más de claridad cómo funciona una célula.

La célula es la unidad básica de la vida. Cada célula tiene en sí todos los atributos de la vida, incluida la reproducción; todos los órganos complejos de los seres vivos, sea cual sea su función, están compuestos por células. Un óvulo fertilizado de un animal, o la semilla de una planta, es una única célula que es capaz de dividirse y crecer para formar un organismo adulto. Esto incluye muchas tareas de división y multiplicación; en nuestro cuerpo existen alrededor de cien billones de células, aproximadamente un número de células mil veces mayor que el número de estrellas brillantes que hay en toda la Vía Láctea (véase el capítulo 10).

Toda célula está rodeada por una membrana, que la encierra y restringe el flujo de sustancias químicas hacia y desde el interior de la célula. Dentro de esta membrana hay una sustancia gelatinosa (el citosol), en la cual están localizadas y actúan muchos tipos diferentes de subunidades biológicas; por ejemplo, los cloroplastos (en las células vegetales verdes) que contienen el pigmento verde denominado clorofila y participan en la fotosíntesis. Además en el centro de la célula, envuelto en otra membrana, se encuentra el núcleo. Todo lo que está en el exterior del núcleo se llama citoplasma y es el lugar donde la célula realiza sus tareas, formando las biomoléculas a partir de compuestos sencillos, tales como el agua y el dióxido de carbono, siguiendo las instrucciones codificadas en el ADN. Pero, vamos a ignorar todo esto para concentrarnos en el propio ADN, que está almacenado dentro del núcleo en unas estructuras conocidas como cromosomas.[16]

Cuando los seres vivientes crecen, esto se debe a que sus células se dividen en dos y aumenta su número. Durante este proceso se copian todos los cromosomas del núcleo. Luego, la frontera entre el núcleo y el resto de la célula desaparece, desplazándose un conjunto de cromosomas hacia cada lado de la célula, donde se reúnen dentro de dos nuevas membranas nucleares. Finalmente, la célula se divide en dos células hijas, cada una de las cuales contiene un conjunto completo de cromosomas. Es imposible decir cuál de los dos conjuntos de cromosomas es el «original» (en efecto, debido al modo en que se realiza el proceso de copiar los cromosomas, no es posible establecer una diferencia entre los dos conjuntos). Donde antes había una célula, ahora hay dos, y ambas son igual de viejas (o de nuevas). Este proceso de división celular se conoce como mitosis.

Sin embargo, cada cromosoma lleva gran cantidad de ADN y es asombrosa la perfección con la que la célula lo desenrolla y lo copia durante la mitosis. En las especies cuya reproducción es sexual existen dos conjuntos de cromosomas, cada uno

de ellos heredado de un progenitor, con lo que se duplica el proceso de copia que hay que realizar. En los seres humanos, por ejemplo, existen exactamente veintitrés pares de cromosomas, y entre todos ellos llevan toda la información sobre cómo construir un cuerpo y cómo manejarlo. Esta información está codificada en unos segmentos de la cadena del ADN llamados genes (en el próximo capítulo veremos más sobre los genes y la evolución). El modo en que toda esta información codificada se almacena en los cromosomas es una obra maestra de embalaje.

Los cromosomas son una mezcla de ADN y proteína, pero, en contra de lo que se había pensado inicialmente, es la proteína la que aporta el andamiaje en el que se almacena el ADN. Estas proteínas especiales pertenecen a una familia llamada de los histones y se sabe que ocho moléculas de histones se unen formando una pequeña bola redonda. La doble hélice del ADN describe dos bucles en torno a esta bola y la sujeción para mantener la hélice en su sitio la realizan otros dos histones que están adheridos cada uno a un lado de la bola. A continuación, hay otra bola de histones que también tiene dos bucles de ADN envolviéndola, y así sucesivamente. Son como las cuentas de un collar y cada una de las bolas con los dos bucles de ADN envolviéndolas recibe el nombre de nucleosoma. Debido a que los cortos fragmentos de ADN que unen un nucleosoma con el siguiente son flexibles, la totalidad del collar se puede enrollar formando una estructura aún más compacta, casi del mismo modo en que un collar de cuentas se puede introducir enrollado en un espacio compacto. Además, también estos rollos se pueden enrollar a su vez para formar unas superbobinas.

Toda célula del cuerpo humano (salvo el óvulo y las células del esperma) contiene cuarenta y seis diminutos cilindros formados de este modo. Si se colocaran uno detrás de otro, los cuarenta y seis cromosomas cubrirían una distancia de tan sólo 0'2 milímetros. Sin embargo, si se desenrollara todo el ADN que almacenan y se colocara en fila, su longitud sería de más de 1'8 metros, es decir, más que la altura de la mayoría de las personas. El ADN se encuentra embalado con una longitud que sólo es aproximadamente una diezmilésima de la longitud que tendría estirado. No obstante, entre todo este ADN prietamente plegado, la maquinaria de la célula es capaz de encontrar el trozo de ADN que necesita cuando desea fabricar una proteína concreta, desenrollar el segmento del prieto ovillo del cromosoma que se precisa, copiar el mensaje del ADN en el ARN mensajero y volver a dejar todo limpiamente enrollado. En la mitosis se desenvuelve todo el lote, se copia y se vuelve a empaquetar en dos conjuntos de cromosomas en un intervalo de tiempo de unos pocos minutos. Sin embargo, cuando se dividen las células especializadas que participan en la reproducción sexual, éstas realizan una maniobra aún más impresionante conocida como re-combinación.

Cuando se fabrican células del esperma y óvulos, la producción se realiza

mediante un tipo diferente de división celular llamado meiosis. Describiremos específicamente lo que sucede en las células humanas, aunque el proceso es muy parecido en todas las especies que se reproducen sexualmente.[17] En vez de, sencillamente, duplicarse todos los cromosomas, en la meiosis los cuarenta y seis cromosomas primero se emparejan de tal forma que cada una de las veintitrés clases de cromosomas se sitúa al lado de su número opuesto (recuérdese que de cada progenitor se hereda uno de los conjuntos de veintitrés cromosomas).

Después de que los cromosomas se han duplicado, los trozos correspondientes de ADN se cortan de cada cromosoma de una pareja y se intercambian, formando nuevos cromosomas. Estos nuevos cromosomas son tales que cada uno de ellos contiene una mezcla de material genético (una mezcla de genes) de ambos progenitores de la persona en la que vive la célula. Este proceso se llama recombinación. La célula se divide entonces en dos células hijas, cada una de las cuales tiene un conjunto de cuarenta y seis cromosomas formando pares, pero entonces se inicia el segundo estadio de la división, sin que se realice ninguna copia del ADN, y en este segundo estadio se producen un total de cuatro células, tales que cada una contiene solamente un conjunto de veintitrés cromosomas, que son los «nuevos» cromosomas producidos mediante la recombinación.

En los machos tres de estas cuatro células suelen convertirse en esperma; en las hembras sólo una se desarrolla convirtiéndose en un óvulo y el resto se desecha. Lo importante de todo esto es que, tanto en el esperma como en el óvulo, existe solamente un conjunto de cromosomas y que cada uno de estos cromosomas contiene genes de cada uno de los progenitores de la persona en cuyo cuerpo se fabricó la célula. Cuando el esperma de un hombre y el óvulo de una mujer se unen para formar una nueva célula que tiene un conjunto completo de veintitrés pares de cromosomas, este óvulo fertilizado posee la capacidad de desarrollarse hasta formar un nuevo ser humano. En cada célula de este nuevo ser humano un conjunto de veintitrés cromosomas contiene una mezcla de información genética de los abuelos paternos y el otro conjunto de veintitrés cromosomas contiene una mezcla de información genética de los abuelos maternos. Pero en cada célula del cuerpo humano (excepto en las células sexuales) hay exactamente el mismo conjunto de cuarenta y seis cromosomas, portando cada cromosoma exactamente la misma información genética que su cromosoma equivalente de cualquier otra célula de dicho cuerpo.

En números redondos, existen 75.000 genes repartidos a lo largo de los veintitrés cromosomas humanos (unos contienen más genes y otros menos), lo que indica que la reproducción sexual y el proceso de recombinación aportan una enorme variedad de mezclas potenciales de genes. Ésta es la razón por la que no hay dos personas exactamente iguales (excepto en el caso de los gemelos idénticos, que resultan cuando un óvulo se divide completamente en dos después de la fertilización,

desarrollándose cada una de las células hijas para formar un nuevo ser humano). Sin embargo, dado que en toda célula hay dos conjuntos de cromosomas, existen dos versiones de cada gen, lo que añade un grado más de complejidad a esta historia.

Supongamos que existe un gen único que determina una característica física, por ejemplo, el color de los ojos. Esto es un poco simplista. En general, se supone que las características físicas (los detalles del fenotipo, según la jerga especializada) se producen por la interacción de varios, o muchos, genes (la totalidad del paquete de genes se llama genotipo); pero en nuestro ejemplo lo vamos a poner más sencillo. En el cromosoma correspondiente heredado de un progenitor, el gen que determina el color de los ojos puede «decirle» al cuerpo, por ejemplo, que tenga los ojos azules. Sin embargo, en el lugar equivalente del cromosoma equivalente heredado del otro progenitor, las instrucciones pueden ser que tenga ojos castaños. Estas versiones diferentes del mismo gen se llaman alelos. En este caso concreto, las personas que tienen el alelo de los ojos castaños en un juego de cromosomas tendrán ojos castaños. Una persona tendrá realmente los ojos azules sólo si los dos cromosomas llevan el alelo de los ojos azules.

Debido a que existen diferentes alelos prácticamente para cualquier gen (a menudo varios alelos distintos) y se barajan continuamente de una generación a otra en nuevos ordenamientos por efecto de la reproducción sexual y la recombinación, la naturaleza está experimentando continuamente con diferentes combinaciones de genes (diferentes genotipos), produciendo fenotipos ligeramente distintos. Al mismo tiempo, la naturaleza también está «inventando» nuevos genes, o, al menos, variaciones sobre viejos genes. Todo este proceso de copiar que se desarrolla cuando las células se dividen no es del todo perfecto, por lo que algunos fragmentos del mensaje que lleva el ADN sufren mutaciones. A veces, durante la meiosis, algunos fragmentos seccionados del ADN se pierden, o se recolocan en un lugar equivocado, o exactamente al revés. Esto tiene generalmente malas consecuencias. Es poco probable que una célula con el ADN dañado funcione de manera correcta y probablemente morirá mucho antes de que llegue a la fase de formar un nuevo tipo de cuerpo, es decir, un nuevo fenotipo. Pero, si el cambio es lo suficientemente pequeño, sólo tendrá un ligero efecto, para bien o para mal, y puede acabar quedando plasmado en el fenotipo de un nuevo individuo.

Un cambio así podría, por ejemplo, alterar la estructura de un tipo de proteína fabricada por el cuerpo, cambiando el orden de los aminoácidos a lo largo de su cadena polipéptica. Si esa proteína concreta fuera, por ejemplo, la hemoglobina, que transporta oxígeno en la sangre por todo el cuerpo, una ligera diferencia en la estructura de dicha proteína podría hacerla más eficiente, o menos eficiente, a la hora de realizar su función. Si se volviera más eficiente, el cuerpo que «poseyera» ese nuevo alelo sería más efectivo, pudiendo respirar con mayor facilidad y teniendo más

probabilidad de sobrevivir y dejar descendencia; y la mitad de esa descendencia llevaría el nuevo alelo en uno de sus cromosomas. Si la nueva versión de la proteína fuera menos efectiva haciendo su trabajo (en este caso, transportando oxígeno), el cuerpo probablemente se encontraría enfermo, siempre falto de respiración y con pocas posibilidades de sobrevivir y dejar mucha descendencia.

A nivel molecular, esto se llama evolución. Se trata de pequeños cambios en el mensaje del ADN (un cambio de un letra en el código del ADN podría ser suficiente para cambiar el aminoácido que va a un lugar determinado de una molécula de proteína) traducidos en fenotipos (cuerpos) que son ligeramente diferentes unos de otros y que se comparan entre sí utilizando un solo criterio: cuál de ellos deja más hijos (es decir, más copias de sus propios genes). La reproducción sexual sencillamente contribuye a favorecer este proceso barajando los genes para formar nuevos fenotipos basados en distintas combinaciones de los genes. Pero aquí no existe el problema del «huevo y la gallina». No hay duda de que primero aparecieron las moléculas y, cuando éstas evolucionaron, inventaron toda la parafernalia de cuerpos (incluidos los de los seres humanos) con el fin de ayudar en su propia reproducción. Los biólogos tienen un aforismo que dice: «Una gallina es la manera que tiene el huevo de hacer más huevos». Del mismo modo, un ser humano es sencillamente la manera que tienen los genes de hacer más copias de sí mismos.

Aunque estamos introduciendo el tema de la evolución haciendo lo contrario de lo que es habitual para contar la historia (vamos de abajo arriba, en vez de hacerlo de arriba abajo), esta manera es mejor para situar las cosas en su propia perspectiva, privando a todos los fenotipos (incluidos los seres humanos) de su papel estelar. Sin embargo, realmente no podemos ir más lejos con nuestro relato sobre las biomoléculas. Con el fin de comprender cómo funciona la evolución en el mundo en general, tenemos que dar otro paso hacia arriba, aumentando la escala, para ver cómo interaccionan los cuerpos con el entorno en que viven y entre ellos mismos.

La evolución

La evolución actúa a nivel de los genes. Obviamente tiene que haber existido una evolución en el mecanismo mediante el cual opera la célula, y los genes, por lo que sabemos de ellos, se han desarrollado como resultado de dicha evolución. Pero aquí no vamos a entrar en estos detalles. Si desea usted saber por qué una persona tiene los ojos azules y otra los tiene castaños, por qué los seres humanos son distintos de los otros monos, por qué todos los monos son diferentes de los lagartos o por qué los guisantes se dan en distintas variedades, lo que importa aquí son los genes. La evolución (por lo que respecta a los animales y las plantas) depende del modo en que los genes se copian y se transmiten a la generación siguiente (herencia biológica), y al hecho de que este proceso de copiar es (como ya hemos visto) casi perfecto, pero no del todo. Es el «casi» lo que garantiza que la descendencia se parezca a sus progenitores y que sean miembros de la misma especie; es el «no del todo» lo que permite que funcione la evolución, siendo a veces la causa de que surjan nuevas especies.

La primera persona que comprendió el modo en que actúa la herencia biológica fue un monje de Moravia, llamado Gregor Mendel. Mendel fue contemporáneo de Charles Darwin y llevó a cabo sus estudios sobre la herencia biológica (concretamente investigando el modo en que las características se transmiten de una generación a la siguiente en el caso de las plantas de guisante) al mismo tiempo que Darwin perfeccionaba su teoría de la evolución por selección natural. Desgraciadamente el naturalista inglés nunca supo de los trabajos de Mendel [18] y éste no se enteró de lo que había realizado aquél hasta una época tardía de su carrera, cuando ya había renunciado a la investigación científica para centrarse en las tareas administrativas que conllevaba su elección como abad del monasterio en 1868. Por este motivo, las dos grandes componentes de la manera moderna de entender la evolución no se aunaron hasta principios del siglo XX, cuando se redescubrió la obra de Mendel y cuando varios otros investigadores, de forma independiente, llegaron a las mismas conclusiones que él sobre la naturaleza de la herencia biológica. También se tuvo que esperar hasta la primera década del siglo XX para que se acuñara el término «gen», pero, de todos modos, lo utilizaremos para referirnos a la unidad básica de la herencia biológica, aunque estemos tratando de los trabajos realizados en el siglo XIX por Mendel y Darwin.

Es importante insistir en que Mendel no era un monje corriente. Procedía de una familia pobre, pero era extraordinariamente inteligente y estaba sediento de lograr hacer una carrera académica en el ámbito de la ciencia. La única forma en que podía conseguir algún tipo de educación superior era ordenarse sacerdote y formarse como profesor. Durante su proceso de formación como docente, estudió especialmente

física y aportó a su investigación sobre la herencia biológica un planteamiento realizado desde la mentalidad de físico, poniendo mucho cuidado en mantener separadas las líneas de reproducción y asumiendo con toda minuciosidad el modo correcto de interpretar sus resultados estadísticamente; algo que incluso los físicos estaban sólo empezando a abordar a mediados del siglo XIX y que era casi inaudito en biología en aquellos tiempos. Hay algo de irónico en todo esto, porque, hace unos pocos años, algunos modernos expertos en estadística, reexaminando los datos de Mendel, alegaron que sus resultados eran «demasiado buenos para ser ciertos» y que por fuerza tenía que haber amañado dichos datos. Al final resultó que estos expertos en estadística no entendían adecuadamente la biología y no habían considerado el hecho de que aproximadamente un guisante de cada diez que Mendel había plantado no habría germinado. Esta historia corrobora maravillosamente la afirmación de que Mendel sí comprendía, además de la biología, también la estadística.

Sin embargo, lo cierto es que, lamentablemente, siempre tenía que encontrar un hueco para la investigación científica en su tiempo libre, ya que su tarea principal era la que realizaba como profesor en Brünn (actualmente Brno); también necesitaba encontrar un hueco para el trabajo físico, porque tenía asignada una pequeña parcela de terreno en la huerta del monasterio. Además, cuando se convirtió en abad y podría, en principio, haberse permitido un espacio de tiempo mayor para dedicarse a sus investigaciones, estaba demasiado ocupado haciendo otras cosas y no podía investigar nada.

Los estudios sobre los guisantes no fueron el único aspecto del trabajo de investigación de Mendel, pero constituyeron con mucho la parte más importante de sus experimentos y son además aquellos en los que se basa el renombre de que goza Mendel actualmente. Trabajó con un total de aproximadamente 28.000 plantas, entre las cuales, según sus propias palabras, «examinó detalladamente» 12.835 ejemplares. Esta tarea se llevó a cabo principalmente en la segunda mitad de la década de 1850, cuando Mendel tenía algo más de treinta años. Anteriormente, otros investigadores habían cultivado un gran número de plantas y habían estudiado su descendencia; pero lo habían hecho más bien con escaso rigor, dejando que las plantas se criaran de forma natural e intentando luego interpretar la variedad de cruces que resultaban. Mendel trató cada planta como a un individuo, dándole un número propio en su cuaderno de anotaciones, y mantuvo las plantas separadas, realizando él mismo la polinización (extrayendo polen de una planta para llevarlo a las flores de otra), de tal modo que sabía en cada caso qué par de plantas eran las progenitoras de cada planta de la generación siguiente.

Entre toda la abundante y minuciosa información que Mendel reunió basándose en sus estudios, bastará un ejemplo clásico para demostrar de qué modo funciona la herencia biológica. La gran ventaja que ofrecen las plantas de guisante para este tipo

de estudios (como Mendel bien sabía antes de comenzar sus trabajos) es que poseen unas características claramente definidas cuya pista se puede seguir a lo largo de generaciones sucesivas. Por ejemplo, algunas semillas de guisante son verdes y otras son amarillas; algunas son lisas, mientras que otras son más rugosas, con superficies arrugadas. Entre sus muchos experimentos, en una sucesión de estudios Mendel tomó plantas de una variedad que tenía semillas lisas y las cruzó con plantas de otra variedad cuyas semillas eran rugosas. En las plantas hijas que se produjeron de esta manera todas las semillas eran lisas y la rugosidad parecía haber desaparecido en este cultivo. Pero, cuando Mendel utilizó plantas que procedían exclusivamente de esta generación hija como progenitoras en sus próximas series de cruzamientos, en la generación siguiente (las nietas de las plantas originales) el 75 por 100 de las semillas eran lisas y el 25 por 100 eran rugosas (las cifras exactas de Mendel fueron 5.474 semillas lisas y 1.850 semillas rugosas).

En términos genéticos, la explicación es sencilla (aunque, desde luego, fueron necesarios más experimentos para que esta sencillez se manifestara con claridad). En las plantas originales cuyas semillas eran lisas cada planta era portadora de dos copias (dicho con precisión, dos alelos) de un gen que determinaba el carácter liso de las semillas. En las plantas originales cuyas semillas eran rugosas cada planta era portadora de dos copias del alelo que determinaba el carácter rugoso de las semillas. En cada caso, no hay duda con respecto a la versión del gen que se va a expresar en el fenotipo.

Sin embargo, en la generación siguiente, cada planta hija hereda un alelo de esta característica procedente de cada progenitor. Cada hija es portadora de una copia del alelo que determina rugosidad y una copia del que determina el carácter de superficie lisa, y ambos alelos están uno junto al otro en cada célula. Lo que sucede es que en este caso el gen de la superficie lisa es dominante. Siempre que la planta lo lleve, este alelo domina sobre el alelo alternativo. Por lo tanto, en toda planta hija producida de este modo la superficie lisa es la que aparece expresada en el fenotipo (el alelo alternativo que está presente en el genoma, pero no se expresa en el fenotipo, salvo que se encuentre en ambas copias del gen, se llama recesivo).

No obstante, en la generación de las nietas los genes se mezclan de una manera bastante diferente. Cada nieta hereda un alelo de esta característica procedente de cada progenitora, y cada progenitora tiene dos alelos diferentes para transmitir. Hay una probabilidad del 50 por 100 de que una planta nieta en concreto herede ambos alelos de una progenitora y la misma probabilidad de que herede los dos alelos de la otra progenitora. Si llamamos R al alelo de la rugosidad y L al de la superficie lisa, las plantas hijas tienen cada una un genotipo (por lo que respecta a esta característica específica) que se puede escribir brevemente como RL. Cada una puede transmitir el R o el L; por consiguiente, la siguiente generación puede tener una de las cuatro

combinaciones de estos alelos: RR, RL, LR o LL (por supuesto, no existe una diferencia real entre RL y LR). Tres de estas cuatro combinaciones (el 75 por 100) incluyen al menos un alelo de superficie lisa, que consecuentemente se expresará en el fenotipo. Sólo una de las variaciones (que se presenta en el 25 por 100 de las plantas nietas) tiene dos alelos en cada célula que determinan ambos la rugosidad. Por lo tanto, sólo una cuarta parte de las plantas nietas tendrá semillas rugosas.

Esto es realmente todo lo que hay referente a la herencia biológica, salvo que, en individuos complejos, como nosotros mismos, miles de genes diferentes, muchos de los cuales se presentan en unas cuantas variedades distintas, se transmiten de hecho de una generación a la siguiente, y muchas veces es el efecto combinado de varios genes diferentes el que determina una característica del fenotipo, como sucede, por ejemplo, con la altura de un ser humano. La cuestión que resulta absolutamente crucial es que la reproducción sexual no implica que se realice una amalgama de las características de los progenitores, sino que la información genética se transmite en unidades discretas, casi similares a los cuantos. A veces existe una apariencia de amalgama (por ejemplo, cuando el hijo de un hombre alto y una mujer de corta estatura crece hasta una altura intermedia), pero esto es el resultado de la participación de muchos genes individuales en una misma tarea, algo así como el modo en que una pintura puntillista parece una capa lisa y continua a cierta distancia del lienzo, pero, cuando se mira de cerca, resulta que está formada por muchos puntos diminutos.

Sin embargo, alguna que otra vez puede suceder que un gen se copie de una manera imperfecta y, en consecuencia, se cree un nuevo alelo. Si esto conlleva que, como resultado, se produzca alguna ventaja en el fenotipo, este alelo se extenderá; si hace que el fenotipo sea menos eficiente, se extinguirá. [19] Así es como funciona la evolución por selección natural, tal como la descubrió Charles Darwin, aunque éste no sabía exactamente cómo se transmitía la información de una generación a la siguiente.

Como ocurre en el caso de Gregor Mendel, a Charles Darwin no siempre se le da todo el crédito que merece, a pesar de que, a diferencia de Mendel, su fama quedó sólidamente establecida estando él aún en vida. Algunas publicaciones populares que divulgan el trabajo realizado por Darwin sobre la teoría de evolución por selección natural insisten aún en retratarle como un joven despilfarrador, hijo de una familia acaudalada, que consiguió tanto por suerte como por méritos su famoso empleo a bordo del HMS Beagle para hacer un viaje alrededor del mundo. Pero esto queda muy lejos de la realidad. Es cierto que Darwin procedía de un ambiente acomodado y privilegiado. También es cierto que descuidó sus estudios en la universidad. Sin embargo, esos estudios fueron primero de medicina (por requerimiento de su padre, un médico que había tenido éxito en su carrera profesional) y tuvo que abandonarlos

debido a sus mareos (literalmente, por salir corriendo después de marearse la primera y única vez que comenzó a ver una operación quirúrgica en aquellos días, cuando aún no había anestésicos). Después inició los estudios de teología, el último recurso para un joven caballero de aquellos tiempos, que le preparaban para una vida tranquila como cura rural, pero que eran algo por lo que él no tenía ningún interés. No obstante, la manera de Darwin de descuidar sus estudios consistía en interesarse vivamente por temas que dentro de su carrera no tenía que estudiar, especialmente la geología y la botánica. Gracias a que sus tutores de Cambridge se dieron cuenta de que era excepcionalmente bueno en ambas materias (aunque no muy bueno como teólogo), fue recomendado al capitán Robert FitzRoy para que fuera su acompañante con tareas de naturalista en el Beagle.

El viaje duró desde el 27 de diciembre de 1831 hasta el 2 de octubre de 1836. Darwin tenía exactamente veintidós años cuando inició el viaje y veintisiete cuando volvió a Inglaterra. Este viaje alrededor del mundo le dio la oportunidad de observar fuerzas geológicas en acción y de ver cómo estas fuerzas habían moldeado la Tierra durante lo que tenía que haber sido un período de tiempo enormemente largo (mucho más largo de lo que la mayoría de la gente creía en la década de 1830, cuando aún se aceptaba ampliamente la idea de que la creación había tenido lugar en el año 4004 a.C., una fecha basada en cronologías bíblicas). El viaje también le mostró toda la profusa variedad de seres vivos que poblaban la Tierra en muchos hábitats diferentes. Muchos otros habían visto con anterioridad las mismas cosas que Darwin, pero fue su mente astuta la que consiguió encajar las piezas del rompecabezas y hallar una explicación sobre el modo en que había evolucionado aquella profusión de vida. La teoría de la selección natural requería precisamente el enorme período de tiempo que señalaba el registro geológico. La geología dio a Darwin como regalo el tiempo suficiente para que la evolución por selección natural realizara su tarea.

Hubo otra persona que tenía ojos para ver la evidencia que se mostraba ante ellos y una mente lo suficientemente sagaz como para unir las piezas del rompecabezas. Darwin tenía clara en su mente la mayor parte de su teoría, y la mayor parte estaba registrada ya en sus cuadernos de anotaciones antes de finalizar la década de 1830. Aunque fue dejando caer gradualmente unos pocos detalles en un estrecho círculo de amigos íntimos, se abstuvo de hacer pública su teoría, en gran medida debido a su preocupación por cómo afectaría todo ello a su esposa, Emma. Ésta era una cristiana con creencias tradicionales firmemente arraigadas, mientras que Charles era un ateo cada vez más convencido. Sin embargo, en la década de 1850, el naturalista Alfred Wallace, que estaba establecido en el Lejano Oriente, llegó a las mismas conclusiones que Darwin sobre la evolución, aunque veinte años más tarde, y desarrolló la teoría de la evolución natural casi en idénticos términos. Darwin era ya por aquel entonces un eminente naturalista con el que Wallace mantenía correspondencia; en este

contexto, Wallace envió a Darwin un esbozo de su teoría. Fue esta carta del joven naturalista lo que obligó al propio Darwin a salir de su reserva y a escribir su famoso libro *El origen de las especies*. Se publicó por primera vez en 1859 y desde entonces no ha dejado de reeditarse. Darwin tuvo la idea primero, como Wallace reconoció; pero Wallace, como Darwin reconoció a su vez, debería ser recordado siempre como el codescubridor, en igualdad de condiciones, de la idea de la evolución por selección natural.

Esta formulación es importante. En la década de 1850 ya se había llegado a tener amplias pruebas, a partir del registro fósil, de que las especies habían evolucionado durante el largo espacio de tiempo geológico. Aunque algunos se negaban a incluir a los seres humanos en la historia de la evolución, la idea de la evolución como tal no resultaba ya ni sorprendente ni rechazable. La importante contribución que hicieron Darwin y Wallace consistió en aportar un mecanismo para la evolución: la idea de la selección natural. La evolución es un hecho, como el hecho de que las manzanas caen de los árboles hacia el suelo. La teoría de la selección natural (ampliamente apoyada por una abundante cantidad de pruebas) es la explicación (o modelo) del hecho de la evolución, de la misma manera que la teoría de la gravedad (ampliamente apoyada por una abundante cantidad de pruebas) es la explicación (o modelo) del hecho de que las manzanas caen de los árboles hacia el suelo.

Tanto a Darwin como a Wallace se les ocurrió la idea en buena parte a través de la lectura del *Essay on the Principle of Population* («Ensayo sobre el principio de población, o principio demográfico») del reverendo Thomas Malthus, que se publicó por primera vez, de forma anónima, en 1798, y posteriormente en una versión ampliada que llevaba la firma de su autor. Malthus estaba impresionado por el modo en que las poblaciones, incluida la humana, tenían la capacidad de crecer en progresión geométrica, lo cual significa duplicar su número en cada espacio de tiempo de una cierta longitud. Si, por ejemplo, cada pareja de cada generación engendra cuatro hijos que viven hasta su madurez y tienen a su vez hijos, la población se duplicará en cada generación. En la época en que Malthus escribió este ensayo, esto le estaba sucediendo en la realidad a la población de América, donde los pioneros se estaban extendiendo poblando las «nuevas» tierras. La población humana de América del Norte se estaba duplicando una vez cada veinticinco años, principalmente como resultado de una reproducción rápida, no por la inmigración. Sin embargo, en el Viejo Mundo la población permanecía más o menos estacionaria, al menos en las comunidades rurales. ¿A qué se debía esto?

El mismo tipo de pregunta se puede plantear con respecto a cualquier otra especie, no sólo en relación con los seres humanos. Incluso los elefantes, que son los mamíferos terrestres que se reproducen más lentamente, podrían de este modo producir 19 millones de descendientes a partir de cada pareja inicial en 750 años. Sin

embargo, en la década de 1790 el mundo no estaba invadido por los elefantes. Más bien se podría decir que sólo había una pareja, aproximadamente, por cada pareja que hubiera existido en el año 1050. Sería válido un argumento similar para los robles o las ranas, para las rosas o los colibríes, en definitiva, para todas las especies vivientes de la Tierra. Malthus señaló que las poblaciones se mantienen dentro de unos límites debido a las enfermedades, a la acción de los depredadores y a la limitación de la cantidad de alimento disponible. Todas las especies, sean plantas o animales, produce una descendencia excesiva en relación con el espacio existente. Sin embargo, la gran mayoría de esos descendientes no sobreviven para alcanzar la madurez y reproducirse a su vez.

Malthus observó todo esto con pesimismo, argumentando que las desesperadas condiciones en las que vivía la gente pobre de las zonas industrializadas de Gran Bretaña eran naturales y que la hambruna y la enfermedad eran los mecanismos naturales, y por consiguiente «correctos», para mantener a la población dentro de unos límites. Si se mejoraran las condiciones de las personas pobres, sólo se reproducirían hasta que, de nuevo, la hambruna y la enfermedad mantuvieran a la población controlada. Por lo tanto, se afirmaba que dejar las cosas como estaban, para que la naturaleza siguiera su curso, significaría en realidad reducir la cantidad de personas que morían de hambre y sufrían enfermedades.[20]

Darwin y Wallace supieron ver más allá de este argumento superficial. Constataron que el exceso de miembros jóvenes de una especie, comparado con el número de los que realmente llegan a reproducirse, significa que existe una fuerte lucha por los recursos, una lucha por la supervivencia, que tiene lugar entre los miembros de la misma especie. Como Darwin escribió en uno de sus cuadernos de notas en el otoño de 1838, el día en que leyó por primera vez el Ensayo de Malthus:

Por término medio, un año con otro, todas las especies tienen el mismo número de individuos que mueren por la acción de los halcones, del frío, etc. Incluso la disminución en número que se produzca en una especie de halcones debe afectar al momento a todo el resto. El objetivo final de toda esta presión debe ser la selección de una estructura adecuada... existe una presión como de cien mil cuñas que intenta introducir a la fuerza cualquier tipo de estructura adaptada en los espacios definidos por la economía de la naturaleza, o más bien que intenta crear esos espacios expulsando a otras estructuras más débiles.

Las especies se adaptan a lo que Darwin llamó «la economía de la naturaleza» encajándose en lo que se conoce como nichos ecológicos. Una carpa se adapta bien en un nicho acuático y a un oso le va bien un nicho especial en tierra. Pero, incluso si

un oso captura a un miembro concreto de una familia de carpas y se lo come, el oso y la carpa no están compitiendo entre sí. Para cada uno de ellos el otro es sencillamente parte del entorno, como el tiempo atmosférico. Si una carpa, hipotéticamente, desarrollara un sentido que la hiciera alejarse de la orilla y sumergirse en aguas profundas cuando un oso se acercara, competiría con éxito con otra carpa que se quedara cerca de la orilla y fuera devorada. Un oso que tuviera una especial habilidad pescando carpas sería el mejor compitiendo con otros osos que fueran menos habilidosos y pasaran hambre. En cada caso, los genes que hacen que esa carpa, o ese oso, consigan tener éxito se propagarían, porque en cada caso estos individuos tendrían más probabilidades de vivir mucho tiempo y prosperar, dejando más descendencia tras ellos que otros miembros de su misma especie.

En esto consiste la adaptación, en el sentido darwiniano. No se trata de la adaptabilidad de un atleta, que incluye fuerza física y agilidad (aunque esto también puede desempeñar un papel), sino de la adaptabilidad de la llave al agujero de la cerradura, o de una pieza de rompecabezas en la imagen global. Las especies como tales, y los miembros individuales de dichas especies, se adaptan a sus nichos ecológicos.

A lo largo de la mayor parte de la historia de la vida en la Tierra, esto significa que la evolución funciona principalmente para mantener a las especies adaptadas, cada vez mejor, a sus nichos. La materia prima de la evolución proviene de la variabilidad dentro de la herencia biológica que ya hemos comentado, cuando los genes se copian (a veces de una manera imperfecta), se barajan y se transmiten de una generación a la siguiente. Debido a que la tarea de copiar el ADN no siempre se lleva a cabo a la perfección, todavía se produce la evolución en algunas especies, tales como las bacterias, que se reproducen asexualmente (aunque esta evolución se produce más lentamente en términos de generaciones, lo cual está bien, pues si no nosotros no estaríamos aquí). Pero la competición entre individuos garantiza que exista una selección natural entre toda la variedad que se reúne en cada generación, donde el que se adapta mejor al medio ambiente es el que mejor sobrevive y deja más descendencia. Si, por ejemplo, un largo pico le sirve de ayuda a un colibrí para conseguir más néctar y sobrevivir para reproducirse, entonces cada generación de pájaros que tengan los picos más largos, aunque sea sólo un poco más, se verá favorecida en comparación con las de pájaros que tengan sus picos más cortos. A lo largo de muchas generaciones, la longitud media de los picos crecerá en los miembros de esa especie.

Entonces ¿de dónde vienen las especies nuevas? Cuando se ve funcionar toda la potencia de la evolución por selección natural es en los momentos en que se producen cambios en el medio ambiente o cuando los individuos se trasladan a un medio ambiente distinto. El ejemplo clásico, que Darwin observó en las islas Galápagos y

que contribuyó a formar su idea sobre la evolución, es la variedad de diferentes tipos de pinzones encontrados en distintas islas.

Los pájaros son todos ellos reconocibles como pinzones, aunque se encuentren en islas diferentes, pero en cada isla la especie está excelentemente adaptada a la vida que lleva allí. Hay muchas islas y muchas variedades de pinzones, pero aquí vamos a ver sólo un par de ejemplos. El alimento existente difiere cuando pasamos de una isla a la siguiente. En una isla resulta práctico un pico largo para sondear y poder acceder al alimento disponible. En otra isla las semillas que allí existen se aprovechan mejor utilizando un pico ancho y duro para abrirlas cascándolas. En cada isla los pinzones tienen justo el tipo de pico que es necesario para aprovechar mejor las reservas de alimentos disponibles.

Para Darwin estaba claro (y después se ha probado mediante estudios moleculares del ADN) que los pinzones de las Galápagos eran todos ellos parientes cercanos, descendientes de unos pocos pájaros originales que habían llegado allí desde tierra firme. En el transcurso de muchas generaciones las presiones malthusianas de la lucha por la supervivencia habían producido diferentes especies a partir del grupo inicial. Darwin afirmó que, considerando un período de tiempo aún más largo, el mismo proceso podría explicar la evolución de la vida sobre la Tierra (incluidos los seres humanos) a partir de un antepasado común.

Todavía hay gente por ahí que piensa que la evolución «no pasa de ser una teoría», del mismo modo que al tío Arturo se le puede ocurrir una idea de chiflado sobre cómo mejorar el cultivo de sus rosas. «¿Dónde está la prueba?», preguntan. Existen pruebas en abundancia, pero han quedado sepultadas en revistas y libros técnicos, a los que no tienen fácil acceso estos incrédulos al estilo de santo Tomás. Sin embargo, hay en especial un libro, *The Beak of the Finch* («El pico del pinzón») de Jonathan Weiner, que describe el proceso de la evolución, exactamente del modo en que Darwin supuso que se producía, tal como se desarrolla ante los propios ojos de los biólogos. Resulta maravilloso el hecho de que son precisamente los pinzones de las Galápagos, a los que el propio Darwin hizo famosos, las especies que aparecen en este libro, y que en él se puede ver precisamente la evolución de estas aves a medida que se van adaptando a las condiciones cambiantes del medio ambiente.

La historia del programa de investigación, de veinte años de duración, que puso de manifiesto cómo funciona la evolución, es en sí tan espectacular que no es necesario darle ningún realce. Rosemary Grant, su esposa, Peter, y sus colegas han estado yendo una temporada tras otra a las islas Galápagos desde 1970 y verdaderamente conocen ya de vista a cada uno de los pinzones de al menos una de las islas. Durante todo este tiempo han estado realizando los árboles genealógicos de estos pájaros y saben cuáles han tenido éxito en la cría y cuáles no. Además, han atrapado a todos los pájaros, excepto a uno o dos, y los han medido y fotografiado,

dejándolos luego en libertad para que volvieran a su vida salvaje. Han observado cómo las poblaciones disminuían en períodos de sequía y cómo experimentaban un auge en tiempos de abundancia; también han visto cómo, para una especie concreta, un cambio de menos de un milímetro en la longitud del pico de un ave puede ser la causa de la diferencia entre crecer y dejar abundante descendencia, o morir antes de tener la posibilidad de reproducirse. Los pájaros de mayor tamaño, dotados de los picos más grandes y capaces de apañárselas con las semillas más duras fueron los únicos capaces de sobrevivir a las peores sequías.

Esta historia se ha actualizado de repente, en términos tecnológicos, gracias al trabajo de Peter Boag, que ha estudiado el ADN de muestras de sangre de los pinzones y ha llegado a estar en situación de poder ver realmente las diferencias del código genético que determinan los diferentes diseños del pico y hacen que un pinzón sea eficiente comiendo semillas, mientras que otro sea especialista en sorber el néctar de las flores de cactus.

Weiner también menciona la persistente resistencia a aceptar las teorías darwinianas y la medida en que, incluso entre científicos que no son biólogos, no se comprende a menudo el funcionamiento de la evolución. Los químicos que habían inventado nuevos pesticidas se quedaron atónitos cuando la población de insectos desarrolló una resistencia a dichas sustancias, sin ver que la clave estaba en la evolución, que hace que cualquier nuevo método para matar, a menos que extermine por completo la especie, hace surgir una población resistente al producto mortífero. Un resultado de la incomprensión es que hay granjeros cultivadores de algodón, precisamente en los estados de EE.UU. donde la resistencia a las ideas darwinianas es más fuerte, que han de luchar cada temporada con las consecuencias de la evolución, que está en plena actividad en sus propios campos.

Por la misma razón, en los hospitales, las bacterias que causan enfermedades son cada vez más resistentes a fármacos tales como la penicilina. Los fármacos matan todas las bacterias vulnerables, pero, por definición, los supervivientes a un ataque son los que no son vulnerables con respecto al fármaco. Cuantas más bacterias vulnerables se maten, más oportunidades se dan a las otras bacterias para que se propaguen, y hay que tener en cuenta que las bacterias se reproducen más rápidamente que los seres humanos, compensando la desventaja que supondría en otros casos el no tener una reproducción sexual. Para un biólogo convencido de la teoría de la evolución, la sorpresa no es que después de medio siglo de uso la penicilina esté perdiendo eficacia, sino que mantenga todavía algo de dicha eficacia.

Pero ¿por qué son algunas personas tan hostiles a la teoría de la evolución? Quizá no sea a la teoría a lo que son hostiles. En uno de los pasajes más expresivos de su libro, Weiner menciona cómo uno de los científicos que tomaban parte en la investigación relataba lo sucedido en un largo vuelo aéreo durante el cual estuvo

charlando con su vecino de asiento y le explicó con todo detalle sobre qué trataba su trabajo.

«Durante todo el tiempo que duró el vuelo, mi compañero de viaje se iba emocionando cada vez más. “¡Qué teoría más ingeniosa! ¡Qué ingenioso es todo esto!”, decía. Finalmente, cuando el avión estaba ya aterrizando, le dije que aquella teoría se llamaba teoría de la evolución. Se puso rojo.»

Si usted tiene una mentalidad abierta, *The Beak of the Finch* no es el mejor texto para aprender cosas sobre la evolución: los libros de Richard Dawkins serían más adecuados. Sin embargo, es el libro ideal para recomendárselo a cualquiera de esos incrédulos que preguntan: «¿Dónde está la prueba de la evolución?», y da una visión amena de uno de los más importantes trabajos de investigación de los últimos veinte años, siendo ésta la razón por la que he reflexionado aquí tan extensamente sobre los temas tratados por Weiner. Se puede decir que describe literalmente la evolución en marcha, tratándola exactamente como Darwin y Wallace sugirieron.

En las islas Galápagos las distintas especies han evolucionado porque sus antepasados se trasladaron a un medio ambiente nuevo. Pero, a veces, durante la larga historia geológica de la Tierra, ha sido el medio ambiente nuevo el que, en cierto modo, ha «llegado» al lugar en que estaban unas especies determinadas y las ha forzado a evolucionar o morir. En varias ocasiones durante la historia de la Tierra se han producido acontecimientos en los que muchas especies han sido completamente aniquiladas, en un proceso en que los supervivientes se dispersaban y evolucionaban para adaptarse a los nuevos nichos ecológicos disponibles. El más famoso de estos acontecimientos, y también el más importante para nuestra propia especie, se produjo hace sesenta y cinco millones de años, ocasionado casi con toda seguridad por el impacto de un cometa o un asteroide contra la Tierra. El resultado fue que los dinosaurios quedaron exterminados y dejaron el camino expedito para que los mamíferos se dispersaran ocupando nuevos nichos ecológicos y evolucionando hacia nuevas formas, entre las que apareció (finalmente) el *Homo sapiens*.

Los mamíferos habían estado ya en la Tierra durante más de cien millones de años antes de producirse la catástrofe que acabó con la era de los dinosaurios, al final del período cretácico, hace unos sesenta y cinco millones de años. Sin embargo, debido a que a los dinosaurios les iba tan bien, las oportunidades de los mamíferos estaban limitadas. Había dinosaurios equivalentes a grandes herbívoros, como son actualmente los elefantes y los ciervos, y había otros que se corresponderían con los modernos depredadores, como son los leones y los lobos. Los mamíferos estaban limitados al papel de pequeñas criaturas, parecidas a las musarañas, que correteaban en la maleza y se alimentaban fundamentalmente de insectos. Pero «el

acontecimiento que se produjo a finales del cretáceo» hizo desaparecer todas las grandes especies y dejó vacíos sus nichos ecológicos. En consecuencia, hubo espacio para que los mamíferos se propagaran, en el sentido ecológico del término, para asumir los roles que habían dejado vacantes los dinosaurios ausentes. Hace sesenta y cinco millones de años, fueron suficientes tres millones de años para que algunos de aquellos mamíferos parecidos a las musarañas evolucionaran hasta convertirse en criaturas del tamaño de los perros, a las que seguían, pisándoles los talones en su evolución, murciélagos, roedores y animales ungulados. Hace unos cincuenta millones de años ya habían aparecido los elefantes ancestrales, que tenían el tamaño de los cerdos actuales.

La velocidad de esta temprana difusión y adaptación de los mamíferos en aquella época tuvo su causa en gran medida en la variedad de oportunidades que les ofreció la desaparición de los dinosaurios. Los procesos que llevaron específicamente a una familia de mamíferos, los primates, a llegar a producirnos a nosotros, los seres humanos, se vieron también afectados por cambios en la geografía del planeta, cuando los continentes cambiaron su posición y se produjeron cambios climáticos, que en parte se originaron como resultado de esos cambios geográficos. En los dos capítulos siguientes veremos más sobre esos cambios que afectaron a todo el planeta.

Nuestro propio lugar en la evolución, y en particular nuestra relación con nuestros parientes más cercanos, los monos africanos, queda reflejada no sólo en el registro fósil, sino también en comparaciones directas entre el ADN de nuestros cuerpos y el de dichos monos. Esta comparación muestra que más del 98 por 100 del ADN de los seres humanos, los gorilas y los chimpancés es igual; así pues las diferencias que nos hacen ser humanos de una manera única sólo ascienden a un poco más del 1 por 100 de nuestro ADN. Se han utilizado estudios de moléculas del ADN de muchas especies para averiguar con qué rapidez se han producido los cambios en el ADN durante la época más reciente de la evolución, y esto muestra que la escisión en tres líneas que conduciría a los seres humanos, los chimpancés y los gorilas actuales tiene que haberse producido hace unos cinco millones de años, en una época en la que los bosques de África oriental que habitaron nuestros antepasados estaban secándose y se encontraban en período de retroceso, lo cual obligó a dichos antepasados a adoptar nuevos estilos de vida y a adaptarse a las condiciones cambiantes con el fin de sobrevivir.[21]

Los seres humanos aparecieron casi exactamente a medio camino en nuestra visión científica general del universo, lo cual no es una coincidencia. Es el resultado del modo que hemos elegido para observar las cosas a diferentes escalas de distancia, comenzando por lo pequeño y yendo hacia arriba. Un núcleo atómico tiene un radio de aproximadamente 10^{-15} m y un ser humano tiene una altura de alrededor de un metro. Por consiguiente, una persona es alrededor de 10^{15} veces mayor que un núcleo

atómico. Un año luz tiene una longitud de aproximadamente 10×10^{15} m. Por lo tanto, cuando consideramos distancias cuya proporción con el tamaño de una persona es como la relación que surge al comparar a una persona con un núcleo, ya hemos salido al universo a gran escala, al dominio de las estrellas. En este sentido, los seres humanos se encuentran aproximadamente a medio camino, en cuanto a tamaño, entre el mundo de los núcleos y las partículas y el mundo de las estrellas.

Los seres humanos también son más o menos todo lo grandes que pueden ser para estar presentes y seguir llevando un modo de vida activo sobre la superficie de la Tierra. Recuérdese que es necesario el tirón gravitatorio de todo el planeta para romper los enlaces eléctricos que mantienen una manzana unida a un árbol y hacerla caer al suelo. Pero la gravedad en la superficie de la Tierra es, desde luego, justo lo suficientemente fuerte como para romper enlaces eléctricos de este modo. De una manera similar, si nos caemos y nos rompemos una pierna, la rotura tiene como causa el hecho de que el tirón gravitatorio en la superficie terrestre, que nos da nuestro peso, es lo bastante fuerte como para interrumpir las fuerzas eléctricas que mantienen unidos los átomos y moléculas de nuestros huesos. Los niños se caen a menudo, pero no se hacen mucho daño, porque su altura queda más cerca del suelo y no tienen que caer de tan lejos. Sin embargo, cualquier cosa que tenga una altura de mucho más de dos metros se ve en serios problemas si se cae. Para un mamífero el único modo de hacerse mucho más grande es ser robusto y pesado (como un elefante) o flotar en el agua (como las ballenas).

No obstante, hay otro modo de considerar el tamaño de un ser humano en relación con el universo. Lo más importante de todo, por lo que respecta al tema de este libro, es que los seres humanos son los sistemas más complejos que podemos encontrar; de hecho son los sistemas más complejos de todo el universo conocido. Esto también se debe en gran medida al intercambio entre la gravedad y otras fuerzas de la naturaleza. En números redondos, como ya hemos mencionado con anterioridad, existen cien billones de células en nuestro cuerpo que están trabajando conjuntamente para formar un organismo vivo. El funcionamiento de estas células tiene lugar en su totalidad mediante la actividad electromagnética, que es la base de todas las reacciones químicas. El número de células que hay en el cuerpo permite la complejidad de sus funciones y posibilita la especialización, existiendo muchas células que trabajan conjuntamente para realizar tareas específicas, incluida la de formar un cerebro grande y complejo. Pero todas las interesantes actividades que realizamos son el resultado de procesos químicos dirigidos por fuerzas electromagnéticas. Como veremos pronto, cuando se observan objetos a una escala mayor, como los planetas y las estrellas, se constata que la gravedad aplasta y anula toda (o, al menos, en gran parte) la interesante estructura electromagnética. Un planeta contiene más átomos que un ser humano, pero no tiene una estructura tan compleja.

A escala de los núcleos y las partículas, las cosas son relativamente sencillas, porque en las interacciones sólo intervienen cada vez unas cuantas partículas. A escala del ser humano, las cosas resultan complicadas e interesantes, por el hecho de que es posible para una estructura realmente bastante delicada de cien billones de células funcionar conjuntamente e interaccionar como si fueran una unidad, y también hay moléculas tan complejas como el ADN que pueden llevar a cabo sus tareas. Sin embargo, otros objetos, como los planetas y las estrellas, son también relativamente sencillos, porque a escala planetaria y a otras escalas aún mayores la complejidad molecular va desapareciendo, ya que la gravedad la destruye y hace que volvamos a la sencillez de la química inorgánica, en el mejor de los casos. En el interior de una estrella, ni siquiera es posible esta pequeña complejidad, por lo que volvemos a la simplicidad de la física de partículas.

En su momento veremos más sobre todo esto. Primero deseo describir algunas de las interacciones en las que participan los sistemas más complejos del universo conocido, es decir, nosotros mismos. Los seres humanos somos demasiado complicados para que nos sean aplicables el tipo de reglas científicas que hacen tan predecible la órbita de la Luna alrededor de la Tierra. No es posible decir exactamente cómo responderá un ser humano ante cualquier influencia exterior cuando ésta afecta al comportamiento consciente, salvo que sea algo tan sencillo como caer bajo la influencia de la gravedad. Sin embargo, la evolución por selección natural es con mucho lo bastante poderosa como para explicar en general por qué nos comportamos de la forma en que lo hacemos, planteando preguntas del estilo de cómo y por qué eligen las personas a sus parejas para el acto sexual, o por qué el altruismo tiene que resultar positivo desde el punto de vista de la evolución y por qué se producen conflictos en muchas familias entre los padres y sus hijos adolescentes.

Charles Darwin fue la primera persona que utilizó argumentos evolutivos para intentar comprender el comportamiento humano y escribió (ya en 1839, aunque el ensayo no se publicó en aquella época):

Mirando al hombre de la misma forma que un naturalista miraría a cualquier otro mamífero, se puede llegar a la conclusión de que tiene instintos parentales, conyugales y sociales, y quizás otros más... La historia de cada raza humana muestra esto, si la juzgamos por sus hábitos, como a cualquier otro animal. Estos instintos consisten en un sentimiento de amor [y simpatía] o benevolencia con respecto al objeto en cuestión. Sin tener en cuenta su origen, vemos en otros animales que estos instintos consisten en una simpatía activa tal que el individuo se olvida de sí mismo y ayuda, defiende y actúa para los demás en detrimento de sí mismo.

Esta idea de «mirar al hombre, como un naturalista miraría a cualquier otro mamífero» es el meollo de la forma moderna de comprender la evolución. Los seres humanos son animales y han sido formados por el mismo tipo de fuerzas evolutivas que han formado a otros animales. El estudio, realizado de este modo, de todas las formas de comportamiento social de todos los animales, incluidos los seres humanos, se denomina a veces sociobiología, y uno de los triunfos de la teoría de la evolución es que la manera moderna de entender la genética y la herencia, combinada con la idea de selección natural, es capaz de explicar el origen del comportamiento altruista (en los seres humanos y en otras especies) que el propio Darwin esbozó en 1839.

Aquí no tenemos espacio más que para ofrecer un esbozo del modo en que funciona la sociobiología, pero al menos esto tendría que convencer al lector de que la idea tiene un contenido.[22] El llamado altruismo aflora de muchas maneras en la naturaleza, pero nos limitaremos a dar aquí dos ejemplos. En primer lugar, ¿por qué un pájaro que forma parte de una bandada, cuando ésta se encuentra buscando alimento en un campo da un grito de aviso si se acerca un depredador? Se podría pensar que al hacer esto llamaría la atención sobre sí mismo, lo cual haría más probable que lo mataran y lo devoraran, por lo que no llegaría a transmitir sus genes a la siguiente generación. En segundo lugar, ¿por qué los seres humanos se sienten a veces motivados a poner en riesgo sus propias vidas para ayudar a un extraño, como cuando alguien se zambulle en río para rescatar a un niño que se está ahogando? A primera vista, esta manera de actuar tampoco parece un buen modo de garantizar que la persona transmitirá copias de sus propios genes.

Pero hay que tener en cuenta que en una bandada de pájaros lo más probable es que todos ellos estén emparentados unos con otros y que muchos de los genes de un individuo estarán también presentes en muchos otros individuos de la bandada. Si uno de estos genes (o varios funcionando juntos) induce a un individuo a dar un grito para salvar a los demás, aunque este pájaro sea devorado, muchas copias de estos conjuntos de genes sobrevivirán en los cuerpos de los pájaros que huyen volando, dejando la búsqueda de alimento para otro día.

El caso del niño que se está ahogando y es rescatado por alguien que no es familiar suyo es un poco más complicado, pero muestra cómo nuestro comportamiento sigue estando configurado por nuestro pasado evolutivo, a pesar de que la sociedad ha cambiado radicalmente durante los últimos siglos.

Hasta tiempos bastante recientes, en términos evolutivos, la mayoría de la gente vivía en tribus o pueblos. Por esta razón, un niño al que viéramos en peligro tenía bastantes probabilidades de ser un pariente nuestro.[23] Un conjunto de genes que inducía la reacción instintiva de lanzarse al rescate se difundiría entre la población de la tribu o el pueblo, siempre y cuando los beneficios (entendidos como el aumento de la probabilidad de que el conjunto de genes se propagara si el niño sobrevivía)

compensaran los costes (entendidos como la disminución de la probabilidad de que el conjunto de genes se propagara si el salvador se ahogaba). Esto alcanza el máximo de sentido si el niño que se está ahogando y su salvador tienen un parentesco lo suficientemente cercano como para que haya, de hecho, muchas copias de los genes que ambos llevan, incluido el gen del «altruismo». Entonces ¿qué probabilidades existen?

Desde luego, hay una gran cantidad de genes que están presentes en un gran número de individuos (como el alelo de los ojos azules). Pero si queremos personalizarlo, dado que usted ha heredado la mitad de los genes de su padre y la otra mitad de su madre, hay una probabilidad del 50 por 100 de que cualquier gen de su madre esté también presente en usted, y lo mismo sucede con un gen de su padre. También hay una probabilidad del 50 por 100 de que usted comparta un gen con uno de sus hermanos y una probabilidad de un octavo de que comparta un gen determinado con un primo hermano (un hijo de la hermana de su madre), y así sucesivamente. Esto nos lleva a mencionar un comentario que se atribuye a J. B. S. Haldane, y que supuestamente éste realizó en una taberna en algún momento durante la década de 1950. Mientras discutía sobre el problema del altruismo tomando cerveza con unos amigos, le preguntaron si arriesgaría su vida por un hermano. Después de pensarlo durante un momento, Haldane respondió: «No por un hermano; pero sí que lo haría por dos hermanos u ocho primos».

La cuestión central es que, por término medio, cualquier comportamiento que garantice la supervivencia de dos de nuestros hermanos u ocho primos nuestros, garantiza la supervivencia de todos nuestros propios genes. En la vida real no siempre tenemos necesariamente que arriesgar nuestra vida para rescatar a un hermano o un primo, o un extraño, incluso en el caso de que nos zambullamos en un río para salvarlos de ahogarse. Puede existir un elemento de riesgo, pero también muchas probabilidades de que tanto el que se está ahogando como el socorrista salven la vida. En términos estadísticos, si hay una probabilidad mayor que el 50 por 100 de que al acudir al rescate sobrevivamos, vale la pena arriesgarse para salvar a un hermano en peligro.

Ahora bien, nadie está sugiriendo que la gente se quede parada en las orillas de los ríos sopesando cómo de próximo es su parentesco con el niño que se está ahogando y haciendo luego el cálculo de las probabilidades de que todo vaya bien antes de zambullirse para rescatarlo. Existe toda una gama de reacciones humanas ante la crisis, lo mismo que con respecto a la mayoría de las situaciones. Hay gente que se va de la orilla, hay quien se echa a temblar y otros se lanzan sin pensarlo dos veces. La cuestión es que todo este patrón de respuestas individuales ha evolucionado para tener en cuenta las peculiaridades genéticas que existían cuando vivíamos en comunidades más reducidas. Los genes que hacen a las personas demasiado

temerarias, desde luego, han desaparecido; sin embargo, hasta cierto punto, también han desaparecido los genes que hacían a la gente demasiado prudente.

El equilibrio fundamental que ha surgido es el que hace que actualmente en la mayoría de las intervenciones de rescate el socorrista haya de tener unas buenas probabilidades de sobrevivir, mientras que el que se está ahogando tiene muy poca probabilidad de sobrevivir sin ayuda. Así, el equilibrio incide a favor de la difusión del gen (o genes) que fomentan este tipo de altruismo. De hecho, el altruismo es un brillante ejemplo de lo que Richard Dawkins llamó «el gen egoísta» en funcionamiento. «Egoísmo» significa en este caso garantizar que las copias del gen sobrevivan, no importando en qué cuerpo lo hagan. Por este motivo, el egoísmo genético en realidad nos hace, en determinadas circunstancias, no ser egoístas a nivel de individuos. Sin este egoísmo genético, nos parecería totalmente antinatural que alguien arriesgara su propia vida para salvar a un niño que se está ahogando.

¿Qué hay en cuanto al sexo? ¿Por qué es la proporción entre machos y hembras en la especie humana y en otros mamíferos tan cercana al 1:1? Después de todo, un solo macho podría, en principio, fecundar a muchas hembras y producir mucha descendencia. La especie «no necesita» tantos machos como hembras para mantener el número de individuos de una generación a la siguiente. Pero mirémoslo desde el punto de vista de los genes. Imaginemos una especie en la que cada macho reproductor tiene un harén de diez hembras, y ninguno de los otros machos se reproduce (lo cual no es muy diferente de lo que sucede en realidad en especies como la del ciervo rojo). Se podría pensar que la proporción entre nacimientos de machos y hembras en una situación así tendría que ser 1:10, de tal manera que las madres no malgastaran sus esfuerzos criando machos que nunca se reproducirían. Pero ¿la madre obtendría más ventajas criando más hembras, que estarían seguras (siempre y cuando alcanzaran la madurez) de transmitir copias de sus propios genes a generaciones posteriores? Sin embargo, cada macho que procrea produce diez veces el número de descendientes que pueda producir una hembra. Por lo tanto, si alguna vez existió una población en la que cada madre producía diez hijas por cada hijo, una mutación que hiciera que una hembra produjese algunos machos más constituiría una ventaja, porque estos hijos procrearían muy abundantemente, más que cada una de las hijas.

La mutación se propagaría y la selección natural garantizaría la evolución de la pauta que observamos en la realidad en especies tales como el ciervo. El número de descendientes machos y hembras nacidos en cada generación es el mismo, porque, aunque los machos tienen (en nuestro hipotético ejemplo) sólo una posibilidad de cada diez de reproducirse, cada macho que se reproduce genera exactamente diez veces más descendencia (diez veces más copias de sus genes) que cada hembra. Por razones obvias, esto se llama una estrategia evolutivamente estable o EEE; cualquier

mutación que hace que las madres produzcan más hembras o más machos constituiría una desventaja, por lo que desaparecería. En la concepción siempre existe una probabilidad del 50 por 100 de que los descendientes sean machos y una probabilidad del 50 por 100 de que sean hembras. Esto no significa que cada madre no pueda tener una sucesión de hijos o una sucesión de hijas, del mismo modo que la probabilidad del 50 por 100 de obtener cara cuando se lanza una moneda no impide que resulte una sucesión ocasional de caras (o cruces) si se realiza el lanzamiento un número suficiente de veces.[24]

Mediante argumentos similares se explica la lucha generacional que tiene lugar en tantos hogares. Dado que cada progenitor transmite la mitad de sus genes a cada hijo, por término medio una pareja necesita producir dos hijos que sobrevivan hasta la madurez para asegurarse de que se transmiten todos sus genes. Obviamente, por lo que respecta a sus propios genes egoístas, sería beneficioso producir más descendencia, como una especie de precaución para el caso de que les suceda algo a los dos primeros. Sin embargo, esto sólo sería ventajoso si el esfuerzo de criar a los descendientes «suplementarios» no deja a los padres tan exhaustos que vayan a descuidar a su primogénito, con lo que éste podría morir en la infancia. No obstante, cada pareja sólo se «preocupa» realmente (en términos de evolución) por la supervivencia de sus propios genes. Cuando el primogénito se hace mayor, son necesarios menos esfuerzos para criarlo y hay menos riesgo de que muera, aunque una parte de los esfuerzos se dedique a criar más hijos. En cuanto los descendientes son capaces de mantenerse por sí mismos, lo mejor, en términos de éxito evolutivo de los progenitores, es librarse de dichos descendientes y criar otro hijo. Sin embargo, desde el punto de vista de los descendientes lo mejor es seguir dependiendo y consiguiendo ayuda de los progenitores. Esto sucede en los pájaros y sucede en los seres humanos, especialmente en sociedades más «primitivas». En muchas sociedades modernas persiste el conflicto instintivo, incluso aunque, gracias a nuestro moderno entorno cultural, los padres pueden no tener en absoluto la intención de producir realmente más descendencia.

El hecho de que unas pautas evolutivas sencillas se hayan complicado debido al desarrollo de la sociedad moderna es una de las razones por las que la aplicación de la sociobiología a los seres humanos dista mucho de ser fácil, y a veces resulta controvertida. Reflexionamos sobre nuestras acciones (al menos, algunas ocasiones) en vez de actuar siempre instintivamente; sin embargo, debido a que el pensamiento ha evolucionado, y es obviamente ventajoso en términos evolutivos, eso también es, en principio, un tema válido para la investigación sociobiológica. Los seres humanos son, como ya hemos dicho, unas criaturas muy complejas, los sistemas individuales más complejos que se conocen. Después de haber traído al lector desde la sencillez del mundo subatómico hasta este punto máximo de complejidad, es hora de seguir

adelante, yendo hacia escalas aún mayores, para observar nuestro planeta en conjunto y su posición en el cosmos. La buena noticia es que a medida que vayamos saliendo al universo en grandes dimensiones, las cuestiones científicas básicas, de ahora en adelante, se irán haciendo de nuevo más fáciles de entender.

Albert Einstein comentó una vez que «el eterno misterio del mundo es su comprensibilidad... el hecho de que sea comprensible es un milagro». Ahora podemos ver que la razón por la que el universo —que es lo que Einstein quiere decir cuando utiliza el término «mundo»— resulta comprensible es que es sencillo. El porqué de esta sencillez puede ser aún un misterio; sin embargo, porque es en sí mismo sencillo pueden comprenderlo unas criaturas tan sencillas como nosotros mismos (aunque seamos complicados en relación con los estándares del universo). Comenzaremos nuestro viaje al espacio examinando la parte sólida de la Tierra que tenemos bajo nuestros pies, aunque no es siempre tan sólida como podríamos pensar.

Nuestro planeta

un escenario en continua transformación

Para el sistema de referencia de los seres humanos, la Tierra resulta grande. Tiene un diámetro de cerca de 12.800 km y su contorno es una circunferencia de 40.000 km, en números redondos. Sin embargo, la mayoría de los hechos que nos afectan directamente acontecen en una capa muy delgada próxima a la superficie terrestre: la corteza de la parte sólida del planeta y la fina capa de atmósfera que la rodea. Para obtener una perspectiva de esta zona en la que surge la vida, imaginemos que se hace un corte a través de la parte sólida de la Tierra, del mismo modo que un cocinero corta una cebolla por la mitad, y que se observa la estructura estratificada que se encuentra bajo nuestros pies.

Sabemos algo sobre la estructura interna de la Tierra gracias a las vibraciones producidas por los terremotos, es decir, por las ondas sísmicas que se desplazan a través de las rocas y sufren reflexiones y refracciones al atravesar los límites entre las diferentes capas de roca, de forma análoga al modo en que la luz experimenta reflexiones y refracciones al atravesar la superficie de un bloque de vidrio. De la misma manera que un físico podría averiguar cuáles son algunas de las propiedades del vidrio que forma un prisma haciendo pasar la luz a través de él y estudiando de qué modo ésta se refracta, así también pueden los geofísicos descubrir algunas de las propiedades de la Tierra observando las ondas sísmicas. Sin embargo, este reconocimiento mediante «rayos X» resulta muy tosco, pues la producción de ondas sísmicas no admite control alguno por parte del ser humano, sino que depende de los terremotos que se produzcan de forma natural; asimismo, requiere un gran despliegue de observatorios sísmicos distribuidos por todo el globo terráqueo para realizar el seguimiento de las ondas sísmicas cuando éstas se producen. En muchos aspectos, sabemos más sobre las estrellas del cielo que sobre el centro de la Tierra, porque, al fin y al cabo, las estrellas podemos verlas.

La máxima profundidad a la que una persona se puede situar bajo la superficie terrestre es la del fondo de la mina más profunda, es decir, sólo 4 km; y la mayor profundidad a la que se ha podido llegar mediante la perforación para sondeos en el interior de la corteza terrestre no alcanza los 20 km bajo la superficie. Aunque las características del campo de gravedad y del campo magnético terrestre aportan algo de información suplementaria sobre lo que sucede en el interior de la Tierra, todavía dependemos de las ondas sísmicas para la mayor parte de la investigación que se ha

de realizar.

Lo que muestran las ondas sísmicas es una estructura en capas construida alrededor de un núcleo sólido interno que tiene un radio de aproximadamente 1.600 km (no se sorprenda si encuentra cifras ligeramente diferentes en distintos libros: los radios exactos que podemos citar son todos ellos un poco imprecisos, lo cual es un reflejo de las dificultades que surgen a la hora de efectuar observaciones sísmicas). Este núcleo interno está rodeado por un núcleo exterior líquido que tiene un espesor de más de 1.800 km. El núcleo en su totalidad es muy denso, probablemente sea rico en hierro y tiene una temperatura de casi 5.000° C. La circulación de este material conductor de la electricidad en el núcleo exterior líquido es claramente responsable de la generación del campo magnético terrestre, sin embargo, nadie ha podido nunca desarrollar un modelo que explique de una forma completamente satisfactoria cómo funciona este proceso.

La elevada temperatura del núcleo es en parte una consecuencia del modo en que se formó la Tierra, que surgió como una bola caliente de materia fundida constituida por muchos objetos menores que colisionaron y quedaron adheridos unos a otros cuando se formó el sistema solar (verán más de esto en el capítulo 9). Una vez que se generó una corteza fría alrededor de la bola de roca fundida, dicha corteza empezó a actuar como una capa aislante que retenía el calor interior, de tal forma que éste sólo podía escapar muy lentamente hacia el espacio. A pesar de esto, si no se produjera algún tipo de inyección continua de calor, el interior de la Tierra no podría estar tan caliente como se encuentra aún actualmente, más de cuatro mil millones de años después. El calor suplementario proviene de isótopos radiactivos (fabricados originariamente durante la agonía de algunas estrellas), que se desintegran convirtiéndose en elementos estables y emiten energía durante este proceso. También esta fuente de calor se agotará en un período de aproximadamente diez mil millones de años y la Tierra se irá enfriando gradualmente hasta quedarse helada y convertida en una masa sólida; pero, para entonces, habrá muerto también el propio Sol, con lo cual el enfriamiento del interior de la Tierra será lo que menos preocupe a los seres inteligentes que puedan ser testigos del acontecimiento.

El volumen de la Tierra está constituido en su mayor parte por la capa que cubre al núcleo. Esta capa se llama manto. La totalidad del manto tiene un espesor de menos de 3.000 km, pero generalmente se considera que está formado por dos componentes (que se distinguen por tener propiedades sísmicas ligeramente diferentes): el manto inferior (con un espesor de alrededor de 2.300 km) y el manto superior (de unos 630 km de espesor). Entre los dos suman el 82 por 100 del volumen de la Tierra y dos tercios de su masa. Por encima del manto, como una piel que envolviera la parte sólida de la Tierra, se encuentra la corteza. Ésta tiene un espesor de alrededor de 40 km bajo los continentes de nuestro planeta, pero de sólo 10 km

bajo los océanos. Si tomamos un valor medio de 20 km, en números redondos el espesor de la corteza representa menos de un tercio del 1 por 100 (0'33 por 100) de la distancia existente entre la superficie y el centro de la Tierra.

Los geofísicos saben mucho más sobre la estructura de la Tierra en las proximidades de la superficie, que sobre el núcleo (aunque «mucho más» tampoco es mucho). A una profundidad de entre 75 km y 250 km, aproximadamente, existe una zona del manto en la que las ondas sísmicas viajan ligeramente más despacio que en las zonas que se encuentran justo encima y debajo de dicha zona. Se trata de una zona de debilidad causada por una fusión parcial de las rocas en la zona de baja velocidad, que se llama astenosfera (la parte del manto situada bajo la astenosfera recibe el nombre de mesosfera). Es posible que las rocas se fusionen parcialmente en esta zona debido a una relación entre presión y temperatura (el mismo tipo de relación explica por qué el núcleo exterior está fundido mientras que el núcleo interior es sólido). Aunque las rocas están más calientes a más profundidad dentro del manto, la presión es también mayor y las rocas se encuentran en estado sólido; por otra parte, aunque la presión es menor encima de la astenosfera, también lo es la temperatura, por lo que resulta que aquí también son sólidas las rocas. En la astenosfera la combinación de la presión con una temperatura de unos 1.100° C, que es la temperatura de un alto horno, es justo la adecuada para hacer que las rocas estén semi-fundidas.

Todo la zona situada encima de la astenosfera se denomina litosfera y el hecho más significativo es que, por la debilidad de la astenosfera, algunos trozos de la litosfera pueden cambiar de lugar, en cierto modo, flotando sobre la astenosfera. Esta libertad de algunos trozos de la litosfera para desplazarse sobre la astenosfera es lo que hace que los continentes se desplacen (muy lentamente) sobre la superficie de la Tierra, produciendo así impresionantes cambios en la geografía de nuestro planeta en el transcurso del tiempo geológico. Sin embargo, el modo en que se produce este movimiento de los continentes no se había comprendido hasta hace relativamente poco; en realidad, es relativamente reciente el hecho de que la mayoría de los geólogos y geofísicos hayan aceptado plenamente la evidencia de la deriva continental, aunque hacía tiempo que unos pocos estaban convencidos de que los continentes se mueven sobre la superficie de nuestro planeta.

Desde los primeros tiempos de la cartografía fiable a escala global, han sido abundantes las especulaciones acerca de las causas de la disposición actual de los continentes. En 1620, poco más de un siglo después de que Cristóbal Colón hiciera su famoso viaje como descubridor, Francis Bacon llamó la atención sobre la similitud existente entre el perfil de la costa oriental de Sudamérica y la costa occidental de África, diciendo lo siguiente:

Cada región tiene istmos y cabos similares, lo cual no es una mera

casualidad. Asimismo, el Nuevo Mundo y el Viejo Mundo concuerdan en lo siguiente: ambos mundos son amplios y anchos si nos acercamos hacia el norte y, sin embargo, se van haciendo estrechos y puntiagudos hacia la parte sur.

No hay nada que demuestre que Bacon considerara realmente estas peculiaridades como la consecuencia de que los dos continentes hubieran sido en algún tiempo una única masa de tierra que se rompió, separándose y desplazándose por el planeta, pero si una plantilla con el perfil de Sudamérica se coloca sobre una esfera (o se simula este movimiento en un ordenador), el saliente que forma Brasil encaja claramente por debajo del saliente de África occidental. América del Norte, si se le aplica un ligero giro, encaja casi igual de bien en Europa, rellenando con Groenlandia el hueco que quedaría en el norte.

La primera versión que se publicó de un mapa que mostraba esta coincidencia fue obra de Antonio Snider, un americano que trabajaba en París en 1858. Su idea consistía en lo siguiente: cuando la Tierra se enfrió, los continentes habían constituido una sola masa de tierra situada a un lado del planeta y, debido a que esta posición era inestable, aquel único continente se quebró y sus fragmentos se separaron y fueron arrastrados a sus posiciones actuales durante una catástrofe que Snider relacionaba con la historia bíblica de Noé y el diluvio universal.

En las décadas siguientes, otros científicos[25] discutieron, de forma pacífica, algunas variaciones sobre el tema. En todo caso, el que logró un puesto merecido en la historia como «padre» de la idea de la deriva continental fue el alemán Alfred Wegener, quien, a diferencia de otros que se limitaron a especular con las distintas posibilidades, ideó un modelo detallado —que no era del todo correcto, pero eso no viene al caso, porque pocos modelos lo son— y, como hizo todo lo posible para convencer a otros geólogos, organizó un gran escándalo con sus teorías, ya que no se limitó a publicarlas calladamente y abandonarlas a la deriva para que se hundieran o salieran a flote por sí mismas. Wegener vivió desde el año 1880 hasta 1930 y publicó por primera vez sus teorías sobre la deriva continental en una monografía en alemán el año 1915. Esta publicación casi no se conoció fuera de Alemania, a causa de la Primera Guerra Mundial, siendo la tercera edición, que se publicó en 1922 y se tradujo al inglés en 1924, la que realmente dio inicio al debate moderno sobre este tema.

Es la geología la que nos ofrece lo que actualmente parece, en una mirada retrospectiva, la prueba clave de que los continentes que se encuentran ahora en las orillas opuestas de los océanos estuvieron en otro tiempo unidos. Por ejemplo, hay formaciones rocosas en el oeste de África que coinciden con formaciones rocosas de Brasil de un modo tan claro que es como si los perfiles de los dos continentes se

aproximaran hasta tocar el uno con el otro sobre una hoja de periódico y luego se rasgara el papel a lo largo de la juntura, separando los dos trozos: sería fácil reconstruir el encaje original sin más que colocar juntos los dos trozos de papel, de tal forma que las correspondientes líneas del texto del periódico se unieran. De un modo similar, los «papeles de periódico» geológicos aparecen de tal manera que se pueden leer las frases desde un lado hacia el otro de la juntura, cuando los continentes «vuelven a colocarse juntos» en una reconstrucción.

A pesar de esta prueba, la idea de la deriva continental siguió siendo un punto de vista minoritario durante unos cuarenta años, en parte debido a que nadie ofrecía una explicación convincente sobre cómo habían podido desplazarse los continentes para separarse. Como mínimo, parecía improbable que hubieran podido abrirse camino por los fondos oceánicos, como grandes transatlánticos surcando el mar; y aunque pudiéramos imaginarnos que la Tierra se partió y los continentes quedaron separados, si todo el planeta se estaba expandiendo, ¿qué es lo que pudo hacer que se expandiera de este modo —unos dos tercios— en cerca de 200 millones de años? Por lo tanto, aunque esta teoría fue objeto de amplios debates en los años veinte y treinta, hasta la década de los sesenta no se obtuvo una prueba irrefutable de que el océano Atlántico se ensancha cada vez más.

En la formulación moderna la deriva continental constituye una parte de un gran bagaje de teorías que se conocen en su conjunto como tectónica de placas. Todo esto surgió durante la década de los cincuenta como resultado de un estudio minucioso de la corteza terrestre que se encuentra debajo de los océanos, utilizando técnicas sismológicas. Para realizar estos estudios locales no es preciso esperar a que los terremotos hagan que la tierra tiemble, sino que basta con improvisar una serie de explosiones para comprobar la corteza mediante las ondas sonoras que generan. Cuando los geólogos fueron capaces de hacer esto a gran escala y saliendo al mar a gran distancia de la costa, fue cuando descubrieron lo delgada que es la corteza oceánica, que en algunos lugares no tiene más que entre cinco y siete kilómetros de espesor. También descubrieron lo accidentado que es el fondo marino, con sus montañas, con cañones submarinos y, lo más importante de todo, con grandes cordilleras oceánicas de miles de kilómetros de longitud y alturas que sobrepasan en varios kilómetros la altura media del fondo oceánico. Como arquetipo tenemos la dorsal Centroatlántica, que, como su nombre indica, se encuentra aproximadamente a medio camino entre Europa y América, extendiéndose de norte a sur por la zona intermedia del océano Atlántico. A lo largo de la parte central de esta cordillera existe un rift [26] activo salpicado en toda su extensión por muchos lugares en los que hay actividad volcánica submarina.

En 1960 Harry Hess, de la Universidad de Princeton, explicó esto y todas las demás características recientemente descubiertas en relación con los fondos marinos,

refiriéndose al primer modelo de lo que llegó a conocerse como expansión de los fondos marinos (hablando con mayor exactitud, expansión del fondo oceánico, pero la aliteración, sea-floor spreading, hace que lo anterior suene bien), resucitando la idea de deriva continental dentro de un nuevo contexto. Según este modelo, las dorsales o cordilleras oceánicas se originan por corrientes de convección ascendentes en la materia fluida del manto terrestre.[27] Estas lentas corrientes de convección transportan los materiales hasta la boca de la dorsal oceánica, donde se derraman hacia el exterior a cada lado de la misma, empujando y alejando a los continentes, y dando lugar a la formación de nuevas cuencas oceánicas jóvenes constituidas a partir de los materiales infrayacentes expulsados.

La convección, por supuesto, afecta a unos materiales que afloran en un lugar y a otros que deben hundirse en alguna otra parte, ya que, para que el tamaño de la Tierra haya seguido siendo aproximadamente el mismo durante la historia geológica reciente, se tiene que destruir parte de la corteza terrestre en algún lugar, más o menos en la misma proporción en que los materiales salen de las dorsales oceánicas. Hess llamó la atención sobre los sistemas de profundas fosas que marcan el borde de algunos océanos, especialmente al oeste del Pacífico. Sugirió que en estos lugares la delgada corteza oceánica está hundiéndose de nuevo bajo el borde del continente, para fundirse finalmente, integrándose en la astenosfera y cerrando así el ciclo de convección. Este modelo ofrecía también una explicación de toda la actividad geológica que se desarrolla a lo largo del Pacífico occidental, donde, como ya sabe cualquiera que haya visitado el Japón, los volcanes y los terremotos son fenómenos frecuentes. El argumento de esta historia es que el Atlántico se está ensanchando, a una velocidad de aproximadamente dos centímetros al año, mientras que el Pacífico se está encogiendo, lo cual tiene como consecuencia que América del Norte se desplaza lentamente hacia Asia.

Incluso al principio de la década de los sesenta, la teoría de la expansión del fondo marino no hizo que se descartara inmediatamente todo lo anterior; no hubo ninguno de esos momentos en que se grita ¡Eureka! y todo el mundo dice: «¡Claro! ¿Cómo no se me ocurrió?». La prueba convincente llegó unos pocos años más tarde, cuando los geólogos estudiaron el fondo marino utilizando una técnica nueva para demostrar las propiedades magnéticas de las rocas.

El magnetismo de las rocas es un instrumento clave para la geología continental. El campo magnético terrestre no es constante, sino que cambia a lo largo del tiempo geológico, debilitándose en ocasiones para hacerse luego más fuerte en el mismo sentido, o invirtiéndose a veces completamente, de tal modo que lo que ahora es el polo Norte magnético ha sido en algunos períodos del pasado el polo Sur magnético. Este proceso se comprende sólo de una manera precaria, si bien es cierto que tiene alguna relación con el modo en que los materiales conductores de la electricidad

situados en el núcleo líquido exterior se hunden profundamente en la Tierra formando remolinos. Por desgracia, la expresión «inversión de los polos» se entiende a menudo de forma equivocada, por lo que algunos alarmistas dicen a veces que la Tierra puede dar un vuelco en el espacio o que la corteza en su totalidad puede empezar a resbalar sobre el globo terráqueo (mientras que el magnetismo permanece invariable), de tal modo que Australia y Europa intercambiarían sus posiciones. Desde luego, no es esto lo que sucede. Lo que estamos diciendo es simplemente que se produce un cambio en la dinamo interna de la Tierra, de tal forma que el campo magnético desaparece progresivamente y vuelve a aparecer en sentido contrario, mientras que los continentes siguen estando donde están durante el breve lapso de tiempo en que esto sucede.

En tierra, cuando las rocas se depositan (por ejemplo, el magma que fluye hacia el exterior de un volcán y después solidifica), los estratos se apilan situándose unos sobre otros. Cuando una capa de rocas fundidas se detiene, adquiere el magnetismo del campo magnético terrestre vigente en ese momento, es decir, su magnetismo natural coincide con la dirección del campo magnético terrestre en el momento de su depósito. Cuando los niveles estratigráficos están bien conservados, los cambios producidos en el magnetismo terrestre se pueden apreciar examinando los estratos del más moderno al más antiguo (es decir, de arriba abajo) y observando las inversiones que presenta la orientación de los minerales magnéticos de las rocas.

A menudo, sin embargo, los estratos aparecen plegados y desordenados debido a procesos geológicos, como por ejemplo el levantamiento de una montaña. En estos casos, una vez que se han calibrado las inversiones magnéticas en los estratos bien conservados, se puede invertir la técnica. Hasta cierto punto, la ordenación y la datación correctas de los estratos desordenados se pueden reconstruir examinando sus propiedades magnéticas y comparándolas con las propiedades magnéticas de los estratos bien ordenados.

Cuando, por primera vez, mediante buques geofísicos, se remolcaron unos magnetómetros por el mar con el fin de medir el magnetismo de las rocas de los fondos oceánicos, estos aparatos revelaron, sin embargo, unas pautas muy diferentes y, desde luego, sorprendentes. Las rocas del fondo marino del Atlántico eran portadoras de un magnetismo fósil que se manifestaba en franjas de norte a sur, de tal forma que una franja tenía la misma orientación que el campo magnético terrestre actual, mientras que la siguiente presentaba la orientación

opuesta, y así sucesivamente. Y, lo que es más, la pauta de estas franjas a un lado de la dorsal Centroatlántica es como la imagen en un espejo de la pauta que siguen en el otro lado de dicha dorsal.

La conclusión es inmediata: las bandas magnéticas siguen la orientación que tenía el campo magnético terrestre durante el tiempo que tardaron las rocas en solidificarse

después de salir a presión, como la pasta de dientes cuando sale del tubo, repartiéndose por ambos lados de la dorsal. Durante millones de años las rocas estuvieron saliendo a presión y adoptando un tipo de magnetismo. Entonces, en un abrir y cerrar de ojos geológico (unos cuantos miles de años), el campo se invirtió y, durante el siguiente intervalo de millones de años, las rocas expulsadas adoptaron la orientación magnética opuesta. Esta pauta repetitiva ha dejado en las rocas del fondo marino una grabación del cambiante magnetismo terrestre, literalmente casi como una grabación en cinta magnetofónica, apareciendo las rocas más jóvenes junto al eje de la dorsal y las más antiguas a cada lado del océano. De hecho, el descubrimiento de esta pauta diversificadora del magnetismo del fondo marino fue lo primero que reveló el modo en que el campo magnético terrestre se había invertido repetidamente durante la historia reciente de nuestro planeta. Por otra parte, desde aquellos días pioneros de la década de los sesenta, las franjas magnéticas del fondo oceánico se han comparado con la pauta magnética que se observa en las series estratigráficas afloradas en el continente. Todo encaja a la perfección.

Al aparecer estas nuevas pruebas, llegaba la hora en que todo estaba a punto para aceptar la teoría de la deriva continental, en su nueva formulación en términos de tectónica de placas. Uno de los momentos clave, cuando ya un grupo mayoritario de expertos empezaba a moverse a favor de la idea, tuvo lugar en un congreso organizado por la Royal Society en 1964, en el cual Edward Bullard, de la Universidad de Cambridge, presentó una de las primeras reconstrucciones mediante ordenador electrónico del encaje de los continentes situados a ambos lados del Atlántico. En un sentido estricto, esta reconstrucción no era ni más (ni menos) persuasiva que lo que fue en 1858 la reconstrucción dibujada a mano por Snider, excepto por el hecho de que utilizaba el borde de la plataforma continental para definir el borde de cada masa terrestre, en vez de utilizar como guía la línea costera actual; pero se ve que ya existía una cierta mística de los ordenadores en 1964 y, además, ya se había preparado el terreno mediante todo el trabajo realizado acerca de la expansión del fondo marino. Bullard aprovechó la marea y cautivó la imaginación de sus colegas, por lo que el año 1964 se considera en general el año en que se le abrieron las puertas a la teoría según la cual los continentes se mueven. Las últimas dudas crónicas se disiparon en la década de los setenta, cuando los satélites que utilizaban telémetros de láser llegaron a ser lo suficientemente precisos como para medir el movimiento de los continentes y confirmaron que, entre otras cosas, el Atlántico Norte efectivamente se está ensanchando a una velocidad de un par de centímetros por año.

La expresión «tectónica de placas» se utilizó por primera vez tres años después de que Bullard presentara su mapa en el congreso de la Royal Society, en un trabajo publicado en la revista científica *Nature*, en la cual Dan McKenzie y R. L. Parker

reunieron en una coherente recopilación todas las nuevas teorías geofísicas. A finales de la década de los sesenta, ya estaba completo el modelo en lo esencial. Dicho modelo explica que el volumen de la corteza terrestre (oceánica y continental) está formado por unas pocas placas, observándose una actividad geológica relativamente pequeña en la parte central de las mismas. Las placas encajan entre sí como las piezas de un puzzle y cubren de esta manera toda la superficie de la Tierra; sin embargo, a diferencia de un puzzle, la actividad que tiene lugar en las zonas limítrofes entre unas placas y otras cambia su pauta a medida que transcurre el tiempo. Existen solamente seis placas principales y doce menores, cubriendo por entero todas juntas el globo terrestre.

El modo en que cambia la pauta de las placas depende de lo que suceda en sus zonas fronterizas, que se conocen como márgenes o bordes de placa. Hay tres tipos de márgenes. Los márgenes constructivos o divergentes son los lugares, situados en las dorsales oceánicas, en que se está creando corteza oceánica nueva y ésta se extiende por ambos lados, de tal forma que vemos dos placas que se están alejando la una de la otra. Por el contrario, los márgenes destructivos o convergentes se localizan en las profundas fosas en que la delgada corteza del fondo oceánico se desliza bajo el borde de una espesa plancha de corteza continental, metiéndose en el manto con un ángulo de unos 45° , con lo que vemos dos placas que se están moviendo la una hacia la otra, siendo una de ellas aniquilada. Finalmente, en los márgenes conservadores o transcurrentes, las placas ni se crean ni se destruyen, sino que se limitan a deslizarse, pasando una junto a la otra.

Existen placas que pueden estar formadas sólo por corteza continental, o sólo por corteza oceánica, o por una mezcla de ambas; pero tanto los márgenes constructivos como los destructivos sólo pueden existir asociados a la corteza oceánica. Actualmente los materiales continentales ni se crean ni se destruyen mediante procesos tectónicos, con la excepción (posiblemente) de procesos poco importantes de formación de montañas que se desarrollan a lo largo del lado continental de un margen destructivo, donde los volcanes están activos. Nadie puede decir con seguridad cómo surgió por primera vez la corteza continental, es decir, qué fue lo que creó los primeros pedazos de corteza continental que podían dar lugar de esta manera a la formación de montañas. Sin embargo, la explicación más verosímil es que la joven Tierra resultó marcada por los impactos de grandes asteroides y que los primeros minicontinentes fueron las costras que se formaron sobre estas cicatrices. Fuera lo que fuese aquello que los originó, si una parte de la corteza continental es transportada hasta un margen destructivo por una placa que está siendo tragada allí, esta corteza no puede ser engullida por la fosa y, además, el margen destructivo pronto deja de actuar.

La totalidad de la corteza oceánica se crea en las dorsales en expansión,

exactamente a la misma velocidad que es tragada en las fosas profundas. Si una fosa profunda resulta bloqueada por materiales continentales, las repercusiones se sienten por todo el cinturón tectónico de convección y, en compensación, debe aminorarse o detenerse el proceso de expansión en algún otro lugar. En su totalidad, la pauta de actividad tectónica es una característica variable de la Tierra.

El mejor modo de explicar todo esto consiste en poner ejemplos. El mar Rojo y el valle del Rift de África oriental son parte de un complejo sistema de fallas que afectan a la corteza terrestre, donde la actividad expansiva ha comenzado en tiempos relativamente recientes, observándose cómo brotan materiales calientes de debajo de la corteza y se expanden hacia los lados por convección, agrietando la corteza mientras lo hacen. El mar Rojo tiene el aspecto de un océano en miniatura e incluso presenta una «cresta» central formada por una dorsal que se expande. Aunque es demasiado pronto para saber exactamente cuál de las fallas de África oriental se desarrollará completamente en un futuro geológico, sí que parece probable que dentro de unas decenas de millones de años uno de estos rifts se combinará con el del mar Rojo para convertirse en un océano tan grande como el Atlántico, desgajando de Asia la masa de África.

Por contra, en la parte occidental de América del Norte, el continente, moviéndose hacia el oeste como resultado de la expansión del Atlántico Norte, ha sobrepasado una dorsal en expansión que existía anteriormente y que ahora está escasamente activa, aunque todavía se pueden sentir los vestigios de su antiguo poder. Esto sucedió porque la actividad expansiva del Atlántico Norte se desarrollaba más rápidamente que la expansión de la dorsal equivalente situada en lo que actualmente es el Pacífico Norte. El continente americano recibía un empuje hacia el oeste a una velocidad mayor que la velocidad a la que la dorsal Pacífica podía crear una nueva corteza, por lo que durante mucho tiempo la corteza del océano Pacífico ha estado desapareciendo, hundiéndose en una profunda fosa que se extiende a lo largo de la costa oeste de América del Norte, y el proceso en que era engullida se desarrollaba a mayor velocidad que la producción que tenía lugar en la dorsal en expansión. Es como si una escalera mecánica, cuyo suelo móvil surgiera con una velocidad constante en la parte de abajo, fuera «tragada» al doble de velocidad en la parte superior, de tal forma que la altura de la escalera disminuyera progresivamente en sentido descendente.

En el Pacífico la dorsal en expansión sigue estando activa más al sur, siendo como un cuchillo que se clava en California, donde ocasiona la compartimentación de la península de la Baja California y la actividad de la famosa falla de San Andrés. Sin embargo, en general, el límite entre la placa norteamericana (que incluye la mitad oeste del fondo oceánico del Atlántico Norte) y la placa del Pacífico (que consiste esencialmente en la totalidad del fondo oceánico del Pacífico) es un margen

transcurrente, en el que la placa del Pacífico gira ligeramente y roza hacia el norte al pasar junto a la placa continental. En un pasado geológico relativamente reciente, la actividad asociada con esta colisión entre placas levantó las Montañas Rocosas; actualmente es responsable de la actividad sísmica y volcánica por toda esta costa (y, precisamente, otras interacciones similares en las fronteras entre placas explican la actividad sísmica que se produce justo en las proximidades del borde del Pacífico, que en ocasiones ha recibido el apelativo de «anillo de fuego»).

Esta actividad que tiene lugar en California está asociada con un desplazamiento hacia el norte a una velocidad de aproximadamente seis centímetros al año de la corteza situada al oeste de la falla de San Andrés. Esta falla incluye un pequeño fragmento del continente que se desgajó del resto de América del Norte por efecto del último vestigio de actividad de la vieja cordillera en expansión. Lo que actualmente es San Francisco se situaba hace 30 millones de años a más de mil kilómetros hacia el sur, en la época en que los sistemas de dorsales y fosas empezaron a aniquilarse unos a otros (suponiendo y es bastante suponer que el proceso haya continuado desarrollándose de forma constante durante todo ese tiempo). Sin embargo, dicho desplazamiento hacia el norte no es constante; en algunos lugares, las dos placas se adhieren una a la otra al mismo tiempo y permanecen así durante décadas, a veces durante siglos, y posteriormente se mueven de un tirón hacia adelante, relajándose así la tensión acumulada. Éste es el motivo por el cual el sur de California es propenso a sufrir grandes terremotos; por ejemplo, si la región situada en torno a Los Ángeles se «queda pegada» durante cien años, cuando se mueve, necesita lanzarse seis metros para alcanzar la posición de las placas correspondiente en ese momento. Y, desde luego, un desplazamiento lateral de seis metros es suficiente para originar un terremoto muy importante.

También hay fallas inactivas en muchos lugares del mundo. Por ejemplo, en Escocia, el Gran Glen indica el lugar donde hubo un borde transcurrente entre placas, muy similar al que existe actualmente en California, pero, en este caso, se trataba de dos trozos de corteza continental que se deslizaban pasando el uno al lado del otro. La actividad existente a lo largo de este límite cesó hace mucho tiempo (aunque también Escocia sufre pequeños temblores de tierra de vez en cuando) y las dos placas se han soldado entre sí; sin embargo, el límite entre dos continentes que antiguamente fueron distintos permanece marcado por el rift del propio Glen.

También existen casos más complicados, como los procesos tectónicos que se desarrollan en juntas triples, en las que se encuentran tres placas, o donde una dorsal en expansión converge con una fosa destructiva formando ángulos rectos. Pero aquí no es necesario que entremos en tantos detalles. Cada placa puede adoptar distintas formas y puede estar delimitada por cualquier combinación de márgenes constructivos, destructivos o transcurrentes, siempre que el conjunto contribuya a

mantener constante la cantidad total de corteza. Dado que la placa del Pacífico carece de corteza continental y se está destruyendo en su margen occidental, es bastante posible (incluso probable) que un día se destruya por completo, cuando América del Norte y Asia colisionen entre sí para formar un nuevo supercontinente. Para entonces, es posible que el Atlántico se haya ensanchado hasta conseguir el tamaño que tiene el océano Pacífico actualmente, y que África pueda estar ya separada de Arabia por un océano cuyo tamaño sea el que tiene el Atlántico Norte hoy en día.

Tales colisiones y adaptaciones de los continentes, seguidas de una ruptura del supercontinente para quedar de nuevo en trozos más pequeños, mientras se forman nuevos lugares de expansión, parecen haber ocurrido varias veces a lo largo de la historia de la Tierra. Cuando los continentes chocan entre sí, se alzan cadenas montañosas. El Himalaya, que todavía está creciendo en altura actualmente, es el resultado directo del avance de la India hacia el norte y su choque con Eurasia; las montañas de Chipre se están levantando por efecto de la presión debida a que los continentes europeo y africano se mueven el uno hacia el otro aplastando el Mediterráneo. También se forman montañas, como hemos dicho anteriormente, a lo largo de los bordes de los márgenes destructivos. Así pues, allí donde haya cadenas montañosas que actualmente se encuentren alejadas de cualquier límite entre placas, podemos asegurar que dichas cadenas marcan los emplazamientos de antiguos límites entre placas desaparecidos hace tiempo. «Rasgando a lo largo de la línea de puntos» de una cadena montañosa y separando los continentes el uno del otro, los geofísicos han podido hacerse alguna idea sobre el aspecto que tenían los continentes hace mucho tiempo. Cuanto más antiguas sean las cadenas montañosas, más tiempo habrá transcurrido desde que el límite era geológicamente activo.

Existe también otra fuente de información relativa a las posiciones ocupadas por los continentes en el pasado. El registro fósil de los seres vivos revela si la roca en que se hallan estaba cerca de los trópicos o en climas más fríos. El descenso del hielo por los glaciares en las épocas de glaciación ha dejado cicatrices que pueden verse en continentes que actualmente se encuentran alejados entre sí, pero que han tenido que estar juntos formando una sola masa de tierra cuando los glaciares estaban sobre ellos. Los vestigios magnéticos también muestran cómo se han desplazado las rocas, porque el magnetismo siempre se establece actuando en dirección norte-sur en el momento en que las rocas se enfriaron, pero puede encontrarse ahora apuntando hacia cualquier dirección, lo cual muestra que el continente en el que se han hallado las rocas giró cuando se desplazó. De este modo, utilizando una gama de indicios geológicos, ha sido posible esbozar, al menos a grandes rasgos, cómo ha cambiado la geografía de nuestro planeta durante, más o menos, los últimos 600 millones de años, aunque el esbozo es, por supuesto, más impreciso cuanto más retrocedamos en el tiempo, por lo que poco se puede decir sobre la geografía en tiempos aún más

remotos.

Los indicios geológicos muestran que, hace más de 600 millones de años, la mayor parte de los territorios que actualmente constituyen América del Sur, África, la península de Indostán, la Antártida y Australia estaban agrupados en un único supercontinente, que se conoce como Gondwana. Para entender esto con una cierta perspectiva, pensemos que hace 600 millones de años es cien millones de años antes de la evolución de los primeros peces; en aquella época, el continente de Gondwana estaba situado aproximadamente cruzando el ecuador. Se desplazaba lentamente hacia el sur como un bloque (gracias a la actividad de una o varias dorsales en expansión desaparecidas hace tiempo) y pasó por el polo Sur hace unos 450 millones de años, ocasionando la formación de glaciares que han dejado sus cicatrices a través de todos los fragmentos actuales del continente de Gondwana.

Entre tanto, cerca del ecuador y en bajas latitudes del hemisferio norte, se estaban uniendo otras piezas del puzzle continental. Por aquella época, América del Norte y Groenlandia estuvieron unidas la una a la otra durante cientos de millones de años. Hace unos 400 millones de años, en la época en que las plantas terrestres estaban empezando a extenderse sobre los continentes, este trozo de materiales continentales colisionó con lo que ahora es parte de Europa (principalmente Escocia y Escandinavia) y las piezas se soldaron para constituir lo que ahora se conoce como el continente de la «Old Red Sandstone» (antiguas areniscas rojas), llamado así por las rocas que se formaron durante la colisión. Para entonces el continente de Gondwana ya había cruzado el polo Sur y se estaba desplazando hacia el norte, mientras los vertebrados se trasladaban a tierra firme y evolucionaban los primeros reptiles. Hace poco más de 250 millones de años, el continente de Gondwana y el continente de las antiguas areniscas rojas chocaron y quedaron soldados el uno al otro. Poco después, el último gran trozo de corteza continental que permanecía independiente, lo que hoy en día es Asia, colisionó con la parte norte de este supercontinente y quedó soldado a Europa, ocasionando el surgimiento de los montes Urales durante este proceso. Todos los modernos continentes se unieron constituyendo un único supercontinente, conocido actualmente como la Pangea, que se extendía desde el polo Sur, cruzando el ecuador, hasta altas latitudes al norte.

Sabemos mucho más sobre el modo en que la actividad tectónica ha cambiado la geografía del globo desde la época de la Pangea, que sobre épocas anteriores, porque se destruyeron muchos de los indicios de la actividad tectónica anterior, o al menos quedaron confusos, debido a las colisiones que dieron lugar a la formación de la Pangea. Hubo glaciares tanto en el extremo norte como en el extremo sur de esta masa terrestre, y la circulación de los océanos quedó muy restringida por el modo en que el continente se situaba, de norte a sur, impidiendo a las corrientes oceánicas su circulación por todo el planeta. En aquel tiempo, conocido como el período pérmico

tardío en el cómputo del tiempo geológico, desaparecieron de la Tierra muchas formas de vida, probablemente como resultado del frío y del modo en que bajó el nivel del mar cuando las aguas quedaron presas en los casquetes de hielo. Esta crisis fue tan dramática para la vida que los geólogos la han considerado literalmente como el fin de una era: el paleozoico.

Todos los intervalos de la escala geológica del tiempo se definen de esta manera, en términos de cambios en el registro fósil; así, las eras son largos intervalos cuyos límites están marcados por cambios muy dramáticos en las formas de vida de la Tierra, mientras que los períodos son intervalos más breves (subdivisiones de las eras) marcados por extinciones menos importantes de formas de vida, salvo cuando el final de una era y el final de un período coinciden. Hace aproximadamente 250 millones de años, en la era que siguió al paleozoico —el mesozoico— la vida comenzó a recuperarse cuando la Pangea inició su fragmentación y el clima del planeta se hizo más cálido. Los primeros mamíferos aparecieron realmente a comienzos del mesozoico, pero durante cientos de millones de años ocuparon un segundo plano, ya que fueron los dinosaurios, al menos por lo que se refiere a animales terrestres, los que llegaron a dominar esta era.

La Pangea, en realidad, había empezado a fragmentarse casi en el momento en que terminó de formarse. El primer rift se produjo a lo largo de una línea que corresponde al Mediterráneo y al Caribe actuales, desgajándose lo que más o menos podía ser el viejo continente de Gondwana, que de nuevo quedó separado de los continentes del norte, y rompiéndose luego en los fragmentos que conocemos hoy en día. En el norte, el supercontinente de Laurasia, que se individualizó al romperse la Pangea, empezó también a fragmentarse, separándose en trozos menores. Así pues, el mayor «rift» se situó entre América del Norte y Europa. Finalmente, Groenlandia se separó de América del Norte y se individualizó cuando el continente norteamericano se desplazó hacia el oeste. La apertura del Atlántico Norte fue uno de los últimos capítulos de este fraccionamiento, ya que la parte septentrional de América del Norte, Groenlandia y Europa seguían estando todavía unidas en el norte hace unos 65 millones de años, en la época de la extinción de los dinosaurios. Se dio la coincidencia de que esta última gran ruptura de la parte norte del supercontinente se produjo aproximadamente al mismo tiempo que la separación entre Australia y la Antártida, con el último rift importante en lo que quedaba del continente de Gondwana. Desde entonces, coincidiendo casi exactamente con la era de los mamíferos, la historia de la deriva continental se refiere a los continentes casi tal como los conocemos ahora, habiéndose movido éstos lentamente hacia las posiciones en que hoy en día se encuentran.

El ensanchamiento del Atlántico Norte hasta alcanzar su tamaño actual ha necesitado más o menos los últimos 65 millones de años, y este intervalo de tiempo,

desde la extinción de los dinosaurios, marca una nueva era (que aún no ha terminado) dentro del cómputo del tiempo geológico. Se trata del cenozoico, la era de los mamíferos. En comparación con algunas de las cifras que hemos estado barajando, 65 millones de años suena como un breve intervalo, y realmente lo es, dentro del contexto geológico. Pero, durante el cenozoico, la era de los mamíferos, la mitad de todos los fondos marinos de la Tierra ha sido tragada por las profundas fosas oceánicas en el cinturón de convección de los fondos marinos y se ha destruido, quedando reemplazada por la nueva corteza oceánica que se expande a partir de las dorsales oceánicas. Los océanos cubren dos tercios de la superficie de la Tierra; así pues, esto significa que, desde la época de los dinosaurios, se ha reciclado y renovado un tercio de la totalidad de la corteza que constituye la superficie de la Tierra.

Es la actual actividad tectónica de la Tierra la que hace que algunas regiones sean más propensas a los terremotos y a la actividad volcánica que otras zonas del planeta. La actividad sísmica se produce generalmente en los márgenes o bordes (de todo tipo) de las placas, siendo la actividad volcánica especialmente frecuente allí donde la corteza oceánica se está hundiendo en profundas fosas, fundiéndose, con emisión de masas de rocas calientes y fundidas que ascienden para salir violentamente a través de cadenas montañosas que se forman en el continente vecino, donde la corteza continental se está desplomando al alcanzar dicho continente una fosa.[28] Como muestra el destino que tuvieron la Pangea y el continente de Gondwana, a largo plazo ningún lugar de la Tierra está a salvo de la actividad sísmica. A muy largo plazo, pueden abrirse nuevos rifts en cualquier lugar, incluso en medio de un continente, dividiéndolo de la misma forma en que África está siendo dividida actualmente a lo largo del Valle del Gran Rift. Incluso lugares como las Highlands en Escocia, los Apalaches, o los montes Urales (que se encuentran hoy en día totalmente alejados de las fronteras entre placas) pueden sufrir aún el fragor de la actividad sísmica cuando las viejas juntas que había entre antiguos continentes se asienten un poco más firmemente en su lugar.

Toda esta actividad tiene una importancia inmediata a la hora de explicar la historia de la vida en la Tierra y también con respecto a nuestro tema principal: la relación entre la humanidad y el universo. La vida ha existido en la Tierra casi desde el momento en que el planeta se enfrió: se han encontrado rastros de bacterias, como restos fosilizados, en rocas que tenían más de tres mil millones de años. No está claro cómo surgió exactamente la vida en nuestro planeta, pero la actividad geológica puede haber desempeñado aquí también un importante papel. Según la vieja teoría, una mezcla de sustancias químicas que se elaboró en un mar poco profundo (o lo que Darwin denominó «una pequeña laguna caliente») y que recibió energía de la luz del Sol y de fuertes tormentas eléctricas pudo haber producido las primeras moléculas capaces de autorreproducirse. Un punto de vista minoritario (pero respetable) es que

los primeros organismos reproductores se formaron en algún otro lugar del universo, quizás en nubes de materia en el espacio, y fueron traídos a la Tierra por unos cometas. El mérito de esta teoría es que significa que transcurrieron miles de millones de años antes de que la Tierra se formara, durante los cuales podían haberse producido los necesarios procesos prebióticos y explica por qué la vida surgió en la Tierra con tanta rapidez después de formarse el planeta: no tuvo que empezar de la nada. Sin embargo, una de las últimas propuestas parte del descubrimiento de unas extrañas formas de vida que viven actualmente en zonas muy profundas de los océanos, y consiguen su energía, no de la luz del Sol, sino del calor que emerge de chimeneas volcánicas, grietas en los fondos marinos que producen masas de agua hirviendo y una química activa en la que interviene el sulfuro de hidrógeno. Existe la posibilidad de que una actividad de este tipo pueda haber producido las primeras moléculas autorreproductoras en nuestro planeta.

Sabemos muy poco sobre la historia de los primeros tiempos de la vida en la Tierra y esto se debe a dos razones: la primera es que las formas de vida tenían un cuerpo blando y carecían de esqueletos u otras partes duras que hubieran podido fosilizarse al morir dichas formas de vida. En segundo lugar, la vida existió al principio solamente en el mar, por lo que los restos fósiles decantados en el fondo marino debieron desaparecer, ya que desde entonces el fondo ha sido reciclado varias veces por la actividad tectónica, salvo cuando esta actividad ha levantado el suelo marino y lo ha convertido en parte de un continente, como está sucediendo hoy en día en el Mediterráneo.

Los límites en la escala cronoestratigráfica están marcados por los cambios en el registro fósil de la vida en la Tierra. El más relevante de estos límites es el que se sitúa entre el precámbrico, que representa toda la historia de la Tierra desde sus orígenes hasta hace unos 600 millones de años, y todo lo que es más reciente. Durante casi cuatro mil millones de años, la evolución tuvo lugar en los océanos, en forma de criaturas unicelulares de cuerpo blando que nos han dejado unos cuantos restos para proceder a su estudio. Sin embargo, hacia el final del precámbrico, la vida empezó a diversificarse en formas multicelulares, como medusas, gusanos, esponjas y otras similares que, aunque escasas, aparecen en el registro fósil. Luego llegó la explosión cámbrica —no tanto una explosión de vida, como una explosión en el registro fósil—, que comenzó hace unos 600 millones de años, cuando las criaturas vivientes desarrollaron características tales como conchas que fosilizaban sin dificultad y (asimismo importante) que son fácilmente reconocibles hoy en día. [29] Sin embargo, todo esto sucedió mucho antes de que la vida emergiera y fuera a tierra firme, donde los continentes no eran más que rocas estériles. Aunque la vida llegó finalmente a transformar la Tierra, las mismas rocas han desempeñado también un papel en la evolución de la vida, ya que su lenta deriva por la superficie del planeta

ha aportado cambios notables en el medio ambiente en que la vida se desarrolla, y esto ya sucedía incluso antes de que la vida se trasladara a tierra firme.

Hace aproximadamente 440 millones de años, se produjo una extinción masiva de la vida en la Tierra, en la que desaparecieron muchas especies. La catástrofe es tan evidente en el registro fósil, que se utiliza para marcar el final de un período geológico, el ordovícico. Parece que esta extinción ocupa el segundo lugar en importancia dentro de las extinciones de formas de vida que ha sufrido nuestro planeta. Sucedió justamente en la época en que el continente de Gondwana, que se había formado durante la última parte del ordovícico, se desplazó sobre el polo Sur y quedó en parte cubierto de hielo.

Cuando un período glacial se origina de esta forma, no son sólo las regiones polares las que resultan afectadas. El hielo únicamente puede crecer sobre los polos de nuestro planeta si se corta el aporte de las corrientes oceánicas que traen agua caliente de las latitudes tropicales. La manera más directa en que puede suceder algo así es tener una masa de tierra sentada a horcajadas sobre uno de los polos.[3 0] La tierra también proporciona una superficie sobre la cual puede depositarse la nieve. En vez de caer sobre el mar y fundirse, puede acumularse, formando capas de hielo y glaciares a medida que transcurren los siglos. Así pues, el propio hielo y la nieve pueden contribuir al enfriamiento, porque reflejan y rechazan el calor del Sol que, de otra forma, calentaría la superficie del planeta, ya sea esta superficie terrestre o marina. El resultado es que toda la Tierra, no sólo la zona polar que se ve afectada directamente, siente el frío, y las pautas de los vientos y del clima cambian en todo el globo.

Por lo que respecta a la vida terrestre, como veremos más adelante, un efecto secundario que a veces resulta de crucial importancia es que, debido a que todo el planeta se enfría, se produce menos evaporación en los océanos, por lo que los vientos son más secos y en el aire es menor la humedad que podría caer sobre la tierra en forma de lluvia. Esto no era demasiado importante para la vida al final del ordovícico, puesto que no había vida en tierra. Sin embargo, lo que sí vemos claramente en el registro fósil de los organismos marinos de la época es que especies que estaban adaptadas a condiciones de aguas frías se trasladaron hacia el ecuador. Ahora bien, las especies que estaban adaptadas a condiciones de aguas cálidas —la mayoría— no tenían hacia dónde ir. En consecuencia, desaparecieron.

Fue después de este desastre cuando la vida no sólo se recuperó, sino que se trasladó a tierra firme, y en este proceso también desempeñó un papel la deriva continental. Cuando el continente de Gondwana se desplazó abandonando el polo Sur, la carga de hielo que llevaba se fundió. Las aguas cálidas volvieron a las altas latitudes del sur y las glaciaciones terminaron. Al mismo tiempo —o, más bien, después de unas cuantas decenas de millones de años— el polo Norte quedó cubierto

por los océanos, y los fragmentos de tierra que actualmente constituyen América del Norte, Europa y Asia se dispersaron en torno al ecuador. Entre ellos había mares poco profundos que bañaban sus costas. El intervalo transcurrido entre hace unos 440 millones de años y hace, más o menos, 360 millones de años (que abarca dos períodos geológicos: el silúrico y el devónico) fue especialmente caliente, quizá porque la actividad volcánica asociada con estos movimientos continentales había hecho que la atmósfera fuera rica en dióxido de carbono, un gas que capta el calor del Sol por el llamado efecto invernadero, y esto favoreció el florecimiento de la vida sobre el planeta. Sin embargo, la característica más importante del medio ambiente en aquel momento (desde nuestra propia perspectiva) puede haber sido la poca profundidad de los mares que bordeaban entonces las masas de tierra. Esto significa la existencia de grandes superficies que estaban totalmente cubiertas por el agua durante la pleamar y quedaban descubiertas con la bajamar.

La cantidad especialmente grande de dióxido de carbono que calentaba la Tierra durante el silúrico también favoreció el crecimiento de las plantas, ya que éstas absorben dióxido de carbono durante la fotosíntesis y lo utilizan para fabricar sus tejidos. Carl Sagan lo expresó en una de sus más memorables frases: «Un árbol está en gran medida hecho de aire». Muchas plantas crecen con más fuerza cuando aumenta el aporte de dióxido de carbono. En el silúrico, las plantas que crecían en deltas fluviales y estuarios poco profundos sometidos a las mareas no estarían cubiertas de agua durante la bajamar, por lo que correrían el riesgo de secarse; pero así, también recibirían más directamente la luz solar, que es el otro ingrediente esencial para la fotosíntesis. En la lucha por la supervivencia, las plantas que pudieran tolerar condiciones de mayor sequedad, tendrían también una clara ventaja y se desplazarían a tierra firme, saliendo del biológicamente rico dominio mareal y obteniendo así un aporte ininterrumpido de luz solar.

La vida animal no tardó en seguir a las primeras plantas en su desplazamiento a tierra firme, por una razón que resulta obvia: los mares estaban demasiado poblados, mientras que la tierra, una vez que en ella hubo plantas que aportaran alimento, ofrecía algo así como una ruta alternativa. Vale la pena detenerse un momento para reflexionar sobre esto. Las formas de vida que tienen más éxito son aquellas que están excelentemente adaptadas a sus nichos ecológicos y nunca tienen que cambiar mucho. Muchas de las bacterias que viven actualmente en la Tierra son idénticas en lo esencial a aquellas predecesoras cuyas que ya estaban por aquí cuando el mundo era joven, hace tres mil millones de años. Son las grandes supervivientes, el auténtico éxito de la evolución.

Sin embargo, cuando a los sujetos se les hace duro competir para encontrar alimento y otros recursos, deben, en el sentido literal, buscar nuevos pastos. [31]

El pez que tuvo más éxito siguió siendo pez, y sus descendientes son hoy peces

todavía. Fueron aquellos peces a los que no les iba demasiado bien en los mares, los que se vieron impulsados hacia la frontera, las marismas sometidas a las mareas, donde evolucionaron para convertirse en anfibios, que se trasladaban tierra adentro para comerse los insectos que habían seguido a las plantas cuando éstas salieron de los mares poco profundos. Del mismo modo, fueron los anfibios que tuvieron menos éxito quienes debieron desplazarse de nuevo, adaptándose completamente a la tierra seca y convirtiéndose en reptiles, y así sucesivamente. Somos descendientes de una larga serie de criaturas que no desempeñaban muy bien sus papeles y tenían que adaptarse o morir, una larga serie de cuasi-fracasos, en términos de la evolución (no diremos fracasos totales, porque en este caso no habrían dejado descendiente alguno). Y lo que es aún peor, para ponernos en nuestro lugar, nuestro linaje no fue el primero que se trasladó a tierra firme. Fue a unos moradores del mar segmentados, con pesadas conchas, a los que les resultó más fácil salir del mar, una vez que fue posible disponer de alimentos fuera de este medio. Algunos antepasados de los miriápodos se encuentran entre los primeros animales que «conquistaron» la tierra firme; y hay que recordar que las cucarachas estaban ya bien instaladas en tierra hace 300 millones de años, mientras nuestros antepasados apenas habían dejado de chapotear en los bajíos.

De hecho, nuestros antepasados directos, los vertebrados, se trasladaron a tierra justo después de otra gran extinción, que marcó el final del período devónico, hace aproximadamente 360 millones de años. Una vez más, en gran medida se puede echar la culpa de las extinciones a la deriva del continente de Gondwana, que hizo que este continente girara y retrocediera sobre el polo Sur. En muchos aspectos, esta crisis que sufrió la vida en el devónico tardío fue una réplica de lo que había sucedido al final del ordovícico. Afectó fundamentalmente a la vida marina, en el sentido de que hubo formas de vida, adaptadas a temperaturas frías, que fueron llevadas hacia el ecuador, y otras formas de vida que estaban adaptadas a los trópicos fueron aniquiladas. También hay pruebas de que la Tierra podía haber chocado con un gran objeto procedente del espacio, un asteroide o un cometa, durante la época de las glaciaciones que estuvo asociada con la deriva del continente de Gondwana, al retroceder éste pasando sobre el polo Sur.

El efecto principal de un impacto como éste pudo ser el de extender un velo de polvo en la parte alta de la atmósfera (polvo fino, como polvos de talco), que reflejaría el calor procedente del Sol y produciría un enfriamiento aún mayor del planeta. Cualesquiera que fueran las causas exactas de este enfriamiento, no hay duda, sin embargo, de que la Tierra volvió a entrar en un fuerte período de enfriamiento al final del devónico, que dio como resultado la extinción de muchas formas de vida. En la recuperación posterior a esta gran extinción, al salir los supervivientes distribuyéndose por los nichos ecológicos que habían quedado libres, los anfibios abandonaron los mares y fueron a tierra.

Al principio del período carbonífero, enormes bosques cubrían grandes superficies de tierras bajas pantanosas. Cuando los árboles de estos bosques murieron y cayeron, sus restos, ricos en carbono almacenado por fotosíntesis, se conservaron en los pantanos, donde se apilaron en estratos de gran espesor que sufrieron luego la presión de las fuerzas geológicas y, después de un largo intervalo de tiempo, se convirtieron en carbón. Este período tuvo una duración que va desde hace 360 millones de años hasta hace aproximadamente 286 millones de años, aunque a veces se subdivide en el mississippiense (desde hace 360 hasta hace 320 millones de años) y el pennsylvaniense (desde hace 320 hasta hace 286 millones de años), llamados así por los depósitos de carbón que datan aproximadamente de esas épocas y se encuentran en la actualidad en las zonas de América del Norte que sus nombres indican.

Gran parte del carbón que se quema hoy en día en todo el mundo, y que es responsable en gran medida del agravamiento del efecto invernadero que estamos sufriendo actualmente, proviene de los depósitos que se generaron en el carbonífero. Hoy esto plantea un problema de calentamiento mundial, como ya comentaremos en el próximo capítulo. Sin embargo, en el carbonífero fue tan grande la cantidad de dióxido de carbono que las plantas tomaron del aire y fijaron como carbón, que debió de tener el efecto contrario, es decir, enfriar el planeta. Desgraciadamente, no sabemos qué efecto pudo haber tenido esto exactamente sobre los seres vivos, ya que desconocemos cuánto dióxido de carbono había en el aire al comienzo del carbonífero, ni hasta qué punto disminuyó el efecto invernadero.

No obstante, para entonces la actividad evolutiva, por lo que respecta a las líneas que conducen hasta nosotros, los humanos, se había trasladado definitivamente a tierra. Al carbonífero le siguió el período pérmico, que se prolongó hasta hace 248 millones de años, y fue aproximadamente a medio camino del pérmico, hace unos 270 millones de años, cuando entraron en escena los primeros animales de sangre caliente: los protomamíferos. Durante la última parte del pérmico, casi toda la corteza continental de la Tierra se encontraba reunida en el único continente que existía: la Pangea. Dicho continente se extendía desde un polo hasta el otro y fue el causante de otro período de frío. En estas condiciones, las criaturas de sangre caliente tuvieron mucho más éxito que sus rivales de sangre fría. Sin embargo, justo cuando parecía que estaban prosperando, la vida en la Tierra sufrió un desastre que trajo la mayor extinción que se puede ver reflejada en el conjunto del registro fósil. Fue tan espectacular que los geólogos la utilizan para marcar no sólo el final de un período (el pérmico), sino también el final de una era (el paleozoico).

En los mares, las extinciones que se produjeron a finales del pérmico (las «extinciones del pérmico terminal») hicieron desaparecer al 90 por 100 de todas las especies durante un período de, todo lo más, 10 millones de años, o quizá mucho

menos. En tierra es más difícil decir qué proporción de las especies resultó extinguida, pero está claro que se trataba de la primera, y mayor, extinción que afectaba a los vertebrados establecidos en tierra, y fue especialmente grave para los protomamíferos, incluidos nuestros antepasados directos. Este enfriamiento del globo que se relaciona con la formación de la Pangea debió de hacer que la vida fuera difícil durante millones de años; sin embargo, no debemos dejar de lado la conclusión de que alguna catástrofe adicional, casi con toda seguridad un golpe proveniente del espacio, remató una situación que ya era de por sí dura, convirtiéndola en un desastre para la vida en tierra firme.

Durante el período triásico, que siguió al pérmico y duró desde hace 248 millones de años hasta hace 213 millones de años, fueron los reptiles, no los mamíferos, quienes mejor se recuperaron de la crisis y se extendieron para llenar los nichos ecológicos disponibles. Hace aproximadamente 230 millones de años, los dinosaurios ya habían hecho su entrada en la escena evolutiva. Tuvieron tanto éxito (en parte porque probablemente también eran de sangre caliente) que dominaron dos períodos geológicos completos: el jurásico (desde hace 213 millones de años hasta hace 144 millones de años) y el cretácico (desde hace 144 millones de años hasta hace 65 millones de años). En números redondos, el «día» del dinosaurio duró 150 millones de años (como dato comparativo, utilizando la definición más amplia del término «humano», nosotros hemos estado por el mundo unos 5 millones de años). Durante la época de la dominación de los dinosaurios, Pangea se rompió en trozos, primero formando dos supercontinentes distintos, Laurasia y el continente Gondwana II, y luego con la ruptura, a su vez, de estos dos supercontinentes en fragmentos que se movieron a la deriva para ocupar las posiciones que disfrutaban actualmente. Pero ninguno de estos continentes pasó en su deriva por los polos, por lo que no hubo grandes períodos de glaciación, y además tampoco se produjeron grandes colisiones continentales. Fue un largo intervalo de estabilidad geológica relativa, que es la razón por la que los períodos jurásico y cretácico son unos intervalos tan largos dentro del registro geológico.

Las extinciones que marcan la frontera entre el jurásico y el cretácico se sitúan en la mitad de este prolongado lapso de calma geológica, y constituyen una especie de puzzle, en términos de deriva continental y tectónica de placas, ya que no existen razones obvias que justifiquen el hecho de que la actividad tectónica haya podido causar en aquella época cambios significativos en las condiciones medioambientales. Tanto las criaturas marinas como los dinosaurios (y otras criaturas establecidas en tierra) sufrieron ciertamente mucho al final del jurásico, y éste es uno de los motivos que se pueden barajar para pensar en un desastre provocado, aunque no sea totalmente, por un impacto proveniente del espacio. No obstante, aunque muchas especies de dinosaurios quedaron extinguidas y otras nuevas ocuparon su lugar, la

línea de los dinosaurios continuó dominando la fauna terrestre. Sobrevivieron en un mundo en continua transformación adaptándose y evolucionando. Entre los cambios a los que tuvieron que adaptarse figura la proliferación de hierba y plantas de flores por todo el terreno, lo cual sucedió en el cretácico medio, hace poco más de cien millones de años. Para entonces, los propios dinosaurios llevaban ya existiendo en el planeta mucho más de cien millones de años; el tiempo transcurrido entre la aparición de los dinosaurios y la proliferación de los pastos es tan amplio como el tiempo transcurrido entre la proliferación de los pastos y nosotros. Desde luego, ninguno de los dinosaurios del jurásico (incluidos algunos favoritos de los niños, tales como el diplodocus o el brontosaurio) vio nunca una flor, ni caminó sobre la hierba.

Incluso los dinosaurios fueron incapaces de sobrevivir a la catástrofe que afectó a la Tierra al final del cretácico, hace 65 millones de años. Algunos de los cambios que trajeron por aquella época las nuevas presiones del medio ambiente estuvieron, una vez más, relacionados con una geografía cambiante. Hablando en términos generales, a ambos lados del océano protoAtlántico, al final del cretácico, existían continentes en el sur (en particular América del Sur, África y la India) y otros continentes en el norte (principalmente África del Norte, Europa y Eurasia) separados por una extensión de océano abierto, el mar de Tetis, que rodeaba todo el planeta en latitudes bajas. Gran parte de sus aguas desaparecieron cuando los continentes se unieron (el Mediterráneo es un vestigio del mar de Tetis), cambiando las pautas de circulación oceánica, alterando el clima y dando quebraderos de cabeza a las criaturas marinas que habitaban en los bajíos.

Actualmente, sin embargo, se tienen pruebas irrefutables de que la catástrofe que afectó a la vida terrestre al final del cretácico, y especialmente a los dinosaurios, fue causada por el impacto de un objeto que venía del espacio y que chocó con la Tierra en algún punto de lo que es ahora la península de Yucatán, en México. Se ha identificado el cráter formado por el impacto, y se sabe que fue producido por un objeto de más de 10 kilómetros de diámetro, que viajaba a una velocidad diez veces más rápida que una bala de rifle y que cuando impactó emitió una cantidad de energía equivalente a al menos cien millones de megatones de TNT y su energía cinética se convirtió en calor.[32]

Los materiales de la corteza terrestre arrancados de la misma, como en una voladura, por efecto del impacto, probablemente fueron lanzados hacia arriba, al espacio, y luego volvieron a caer, de tal modo que, en todos los lugares donde los fragmentos chocaron, su energía cinética debió convertirse en calor. Como consecuencia, muchas de las criaturas de la superficie sólida de la Tierra —incluidos muchos dinosaurios— fueron asadas vivas, mientras el calor provocaba un incendio forestal general, el cual ardió con tal virulencia que dejó una capa de hollín visible aún en los estratos de roca que tienen 65 millones de años. Después del calor vino el

frío, ya que el humo y el polvo envolvieron el planeta e impidieron que el calor del Sol alcanzara el suelo. La vida vegetal, dependiente de la luz solar para realizar la fotosíntesis, fue gravemente afectada; los animales herbívoros no tenían nada para comer. Así, cuando los herbívoros desaparecieron, les tocó morir a su vez a los animales carnívoros. En total un 70 por 100 de todas las especies de la Tierra fue aniquilado en el suceso que tuvo lugar al final del cretácico, un desastre tan grave que los geólogos lo han elegido para indicar no sólo el final de este período, sino también el final de la era mesozoica y el principio de la era cenozoica.

Entre los supervivientes había algunos pequeños mamíferos, criaturas del tipo de las musarañas que habían vivido durante decenas de millones de años a la sombra de los dinosaurios, nuestros antepasados. El modo en que se extendieron y adaptaron después de las extinciones es otro ejemplo perfecto del funcionamiento de la evolución darwiniana, y además explica por qué estamos hoy aquí.

Durante los períodos geológicos que siguieron a la muerte de los dinosaurios (el paleógeno, desde hace 65 millones de años hasta hace 24 millones de años, y el neógeno, que comenzó hace 24 millones de años y aún no ha terminado), la geografía del globo ha cambiado sólo ligeramente, aunque esos cambios puedan haber desempeñado un papel importante en nuestra propia evolución. Es hora de dejar esta historia sobre el modo en que la parte sólida de la Tierra se ha transformado, para mirar los cambios mucho más rápidos que se producen en la capa de aire que rodea a nuestro planeta, que además tienen consecuencias relacionadas con las actividades humanas pasadas, presentes y futuras.

Vientos de cambio

La atmósfera de la Tierra es esencial para nuestro bienestar y sin ella probablemente no habría ningún tipo de vida en nuestro planeta. Ahora bien, en términos planetarios es apenas algo más que una idea geológica de menor importancia. El diámetro del volumen sólido de la Tierra mide 12.756 km. En cambio, definir el espesor exacto de la atmósfera es difícil, ya que ésta no tiene un borde preciso, sino que pierde densidad y acaba desvaneciéndose en el espacio, aunque la altitud del monte Everest sobre el nivel del mar, o sea 8'85 km, viene a ser la máxima altura a la que una persona puede subir y encontrar aún un aire respirable, mientras que la parte superior de la capa de la atmósfera, donde se deciden las situaciones meteorológicas (la troposfera), se encuentra por término medio a unos quince kilómetros de altitud sobre el nivel del mar. Si nuestro planeta tuviera el tamaño de una pelota de baloncesto, el espesor de la atmósfera respirable no sería más que un cuarto de milímetro, es decir, un poco de suciedad apenas perceptible sobre la superficie de la pelota. También se puede pensar de otra manera: si se tratara de viajar por la superficie de la Tierra, una distancia de diez kilómetros (justo algo más de seis millas) sería un viaje muy cortito. Saliendo de cualquier punto del planeta, si se recorren diez kilómetros en cualquier dirección, en la mayoría de los casos esperaríamos ver un entorno muy parecido al del punto de donde hemos salido. Sin embargo, si viajáramos diez kilómetros verticalmente hacia arriba y al llegar intentáramos respirar el aire que hay allí, moriríamos.

No obstante, esta delgada mancha de aire que cubre la superficie terrestre es absolutamente crucial para hacer que nuestro planeta sea un lugar adecuado para vivir. Todo esto empezó precisamente al principio, justo después de formarse la Tierra. No es mucho lo que sabemos sobre las condiciones que se daban entonces en la superficie terrestre, pero sí que conocemos, por las pruebas geológicas, que se estaban formando rocas sedimentarias hace 3.800 millones de años, existiendo vestigios de vida en dichas rocas desde hace 3.500 millones de años. Las rocas sedimentarias sólo se forman bajo el agua, por lo que está claro que había agua líquida en abundancia sobre la superficie de la Tierra hace 3.800 millones de años, es decir, menos de mil millones de años después de la formación del planeta. Esto resulta intrigante, pues los astrónomos calculan, a partir de sus modelos relativos al funcionamiento de las estrellas (véase el capítulo 10), que en aquella época el Sol únicamente producía alrededor del 75 por 100 del calor que genera hoy en día. Sin tener en cuenta otros aspectos, esto podría significar que la Tierra en su primera época era una pelota de hielo sin vida, con una temperatura en su superficie que estaría bastante por debajo del punto de congelación del agua. Además, dado que el hielo es un buen reflector, aunque el Sol con el tiempo se fue calentando, aquella Tierra primitiva congelada podía haber permanecido congelada para siempre,

reflejando cualquier aporte de calor, en vez de absorberlo y utilizarlo para fundir el hielo.

Así pues, ¿por qué estaba aquella Tierra de los primeros tiempos lo suficientemente caliente como para tener océanos de agua líquida en los que la vida evolucionaba? Casi con toda seguridad, esto sucedió gracias al efecto invernadero. El efecto invernadero es el proceso por el cual las atmósferas mantienen a los planetas más calientes de lo que estarían en otro caso. Recibe este nombre porque el aire en un invernadero está también más caliente de lo que lo estaría si el invernadero no existiera; aunque, paradójicamente, un invernadero lo hace de otra manera. Un invernadero real capta el calor porque su techo de cristal actúa como una barrera física. El aire caliente que se encuentra cerca del suelo se calienta porque el propio suelo (y todo lo que haya en el invernadero) se calienta debido a la entrada de rayos solares, y dicho aire caliente no puede elevarse por convección, ni tampoco escapar, porque el tejado se interpone en su camino. El «efecto invernadero» funciona de un modo bastante diferente. La energía procedente del Sol, en su mayor parte en forma de luz visible, pasa a través del aire casi sin obstáculos y calienta la superficie de la tierra o del mar. La superficie caliente también irradia energía, pero con longitudes de onda más largas (porque está más fría que la superficie del Sol), dentro de la parte del espectro correspondiente a los infrarrojos. Esta radiación de infrarrojos es captada en gran medida por los gases que están en la atmósfera, como el dióxido de carbono y el vapor de agua. Como resultado de esto, la parte inferior de la atmósfera se calienta, ocasiona una convección y pone en marcha todos los sistemas meteorológicos del planeta.

Justo después de formarse, la Tierra era un balón de rocas fundidas en el que no había aire. Sin embargo, cuando las rocas se enfriaron y se formó una corteza, los gases que escapaban de los volcanes y de las grietas de esa corteza empezaron a formar una atmósfera. Estos gases deben haber sido el mismo tipo de mezcla que sale actualmente de los volcanes. Contenían vapor de agua, que cayó en forma de lluvia y originó los océanos, nitrógeno y también amoníaco, que fue descompuesto por la luz solar dando hidrógeno y nitrógeno. La mayor parte del hidrógeno se fue al espacio, porque sus moléculas son muy ligeras; el nitrógeno es un gas relativamente pesado, estable y no reactivo, por lo que este primer nitrógeno se quedó formando parte de la atmósfera, donde constituye más del 78 por 100 del aire que respiramos hoy en día. Además, en la primitiva atmósfera de la Tierra también había una enorme cantidad de dióxido de carbono, que es un gas muy efectivo a la hora de ocasionar un efecto invernadero.

Sabemos qué cantidad de dióxido de carbono había, porque se ha conservado en forma de rocas carbonatadas, que se formaron en procesos asociados a la disolución en agua del dióxido de carbono que estaba en el aire; las estalactitas y estalagmitas de

las cuevas de calizas son un ejemplo de cómo funciona este proceso. La presión que la atmósfera ejerce actualmente al nivel del mar se define como un bar, es decir, una atmósfera de presión. Si todos los carbonatos que se encuentran en las rocas de la corteza terrestre se convirtieran de nuevo en dióxido de carbono, en la superficie se produciría una presión de 60 bares, es decir, sesenta atmósferas de presión. Desde luego, es evidente que todo el dióxido de carbono no estaba en la atmósfera al mismo tiempo. Pero había suficiente para justificar que la joven Tierra se mantuviera lo bastante caliente como para que el agua fluyera, debido al efecto invernadero, incluso sabiendo que el Sol hacía mucho tiempo que estaba más frío. Lo que ha hecho que la Tierra sea un lugar adecuado para la vida durante más de tres mil millones de años es la combinación de una atmósfera de dióxido de carbono que mantenía caliente a la Tierra y unos océanos en los que el dióxido de carbono se disolvía y se depositaba formando rocas, de modo que el efecto invernadero se debilitó cuando el Sol se calentó. Además, podemos ver cómo podrían haber sido las cosas sólo mirando a nuestros más próximos vecinos del espacio: la Luna, que carece de aire, el planeta Venus (el siguiente a la Tierra, si nos acercamos al Sol) y el planeta Marte (el siguiente a la Tierra, pero más alejado del Sol).

Un efecto producido por la capa de aire que rodea al planeta es transportar el calor alrededor del globo, suavizando las diferencias de temperatura. Esto no puede suceder en la Luna, que no tiene aire, por lo que allí la temperatura se eleva a más de 100° C en la parte bañada por el Sol y desciende a unos -150° C durante la noche. La temperatura media es de aproximadamente -18° C, a pesar de que la Luna se encuentra más o menos a la misma distancia del Sol que la Tierra y su superficie se calienta y devuelve luego la energía solar de una forma muy parecida a como lo hace la superficie terrestre. En cambio, la temperatura media en la superficie de la Tierra es de unos 15° C. Por lo tanto, en la superficie terrestre la temperatura es aproximadamente treinta y tres grados más alta en comparación con lo que sería si no tuviera una capa de aire y esto se debe por completo a la existencia del efecto invernadero, a pesar de que actualmente la atmósfera sólo contiene alrededor de un 0'035 por 100 de dióxido de carbono (esta proporción era solamente un 0'028 por 100 hace siglo y medio, antes de que se empezaran a quemar profusamente combustibles fósiles, devolviendo así al aire el dióxido de carbono del carbonífero). Es fácil calcular las dimensiones del efecto invernadero producido por este dióxido de carbono y por el vapor de agua contenido en el aire, y dicho cálculo, desde luego, «predice» que la temperatura en la superficie terrestre puede ser, de hecho, treinta y cinco grados más caliente en comparación con lo que sería si la Tierra no tuviera atmósfera, o si la atmósfera estuviera compuesta exclusivamente por nitrógeno, que no es un gas de los que producen efecto invernadero.

Entonces ¿cuánto calentamiento por efecto invernadero producirían 60 bares de

dióxido de carbono? Es casi una pregunta capciosa, porque la radiación de infrarrojos que sale de la superficie terrestre y puede ser absorbida es limitada. Un fotón que sea absorbido por una molécula de dióxido de carbono no será absorbido de nuevo, independientemente del número de moléculas de dióxido de carbono que haya. Una vez que todos los fotones de la radiación de infrarrojos hayan sido absorbidos en el efecto invernadero, no habría ninguna diferencia aunque se añadiera más dióxido de carbono. Mucho antes de que alcanzáramos los 60 bares, se produciría un efecto invernadero galopante, con lo que el planeta se calentaría tanto como fuera posible.

Esto es exactamente lo que sucedió en Venus. Aunque Venus es casi igual que la Tierra en tamaño, se encuentra más cerca del Sol y recibe de él más calor. Incluso sin atmósfera, la temperatura de su superficie sería de 87° C y con tan sólo un pequeño efecto invernadero su temperatura ascendería a 100° C. Venus nació exactamente del mismo modo que la Tierra, pero su mayor calentamiento por el Sol junto con el efecto invernadero se combinaron para garantizar que, incluso cuando Venus era aún joven, su temperatura nunca descendiera tanto como para que el agua se condensara y cayera de la atmósfera. Todos los gases de dióxido de carbono que se pudieron emitir en Venus permanecieron en el aire, porque no había océanos para disolverlos. Esto se sumó al efecto invernadero, y la temperatura de la superficie se elevó cuando el Sol se calentó. El resultado es que Venus actualmente tiene en su superficie una temperatura de más de 500° C, muy superior al punto de ebullición del agua. Entonces ¿cuánto dióxido de carbono hay en su espesa atmósfera? En números redondos, 60 bares, es decir que este planeta gemelo de la Tierra ha emitido la misma cantidad de gas invernadero que nuestro propio planeta, pero la diferencia es que, al no tener agua líquida para disolverlo, todo el dióxido de carbono se ha quedado en la atmósfera. Así pues, ésta es la razón por la cual Venus es un desierto abrasador, mientras que la Tierra es un lugar adecuado para la vida.

Marte, que es más pequeño que la Tierra y se encuentra más alejado del Sol, sólo posee actualmente una fina atmósfera de dióxido de carbono y es un frío desierto sin vida. En la mayoría de los lugares de Marte, la temperatura cae cada noche por debajo de -140° C y las temperaturas diurnas sólo rebasan la temperatura de congelación del agua durante un breve período del verano en el hemisferio sur del planeta. Sin embargo, su superficie está marcada por muchos trazos que parecen haber sido excavados por agua que fluyera hace mucho tiempo, cuando el planeta aún era joven. Todas las evidencias sugieren que Marte comenzó siguiendo el mismo camino que la Tierra, con un efecto invernadero lo bastante fuerte como para subir la temperatura en su superficie por encima del punto de congelación durante todo el año. Sin embargo, debido a que Marte es tan pequeño —su masa es sólo poco más de un décimo de la masa de la Tierra— la fuerza de atracción de la gravedad fue (y sigue siendo) demasiado débil para retener el dióxido de carbono, por lo que la mayor parte

de éste se ha perdido en el espacio. Una vez que se perdió, la atmósfera original no pudo ser reemplazada, debido a lo pequeño que es Marte. Se supone que el planeta perdió su calor interno de manera relativamente rápida, de tal modo que sus procesos tectónicos se detuvieron y ya no hubo volcanes activos que pusieran más dióxido de carbono en la atmósfera. Así, el efecto invernadero que existe en Marte actualmente es sólo un vago recuerdo de lo que fue. Ésta es la razón por la que Marte es hoy un desierto congelado, mientras que la Tierra es un bonito lugar para vivir.

Pero, si bien la naturaleza de la primitiva atmósfera de la Tierra tuvo un profundo efecto en cuanto a hacer el planeta habitable, una vez que la vida puso pie en dicho planeta, comenzó ella misma a cambiar la naturaleza de esa atmósfera. En la fotosíntesis hay una serie de reacciones químicas inducidas por la energía de la luz solar (es decir, van «hacia arriba» en el sentido relativo a la energía química que le hemos dado a la expresión anteriormente) que fabrican unos compuestos complejos —hidratos de carbono— a partir del carbono, el hidrógeno y el oxígeno.

La primera bacteria que aprovechó la fotosíntesis utilizó ácido sulfhídrico como fuente de hidrógeno (razón por la cual algunos piensan que la vida en la Tierra se originó en torno a chimeneas volcánicas calientes) y consiguió carbono y oxígeno a partir del dióxido de carbono. Dado que estas bacterias vivían en el agua, tomaron el dióxido de carbono de una solución, permitiendo así que se disolviera más dióxido de carbono procedente del aire y contribuyendo de esta manera a estabilizar el medio ambiente de la Tierra, sin riesgo de que se produjera un efecto invernadero galopante. Posteriormente, cuando algunas bacterias ya habían evolucionado, empezaron a utilizar una versión diferente de la fotosíntesis en la que intervenía la clorofila, es decir, la sustancia que hace que las plantas verdes sean verdes. En este proceso, la energía de la luz solar se emplea para obtener hidrógeno a partir del agua, en vez de a partir del ácido sulfhídrico. Debido a que los organismos consiguen todo el oxígeno que necesitan para la fotosíntesis obteniéndolo del dióxido de carbono, el oxígeno que queda cuando el hidrógeno se extrae del agua es, para ellos, un producto de desecho. Este oxígeno se libera y va a incorporarse a la atmósfera, por lo que, una vez que empezó a realizarse, este proceso trajo con el tiempo un dramático cambio en la composición del aire.

El oxígeno es una sustancia altamente reactiva, que se combina con facilidad con muchos otros elementos en un proceso conocido como oxidación (una llama ardiente es un ejemplo de oxidación rápida; la herrumbre es un ejemplo de oxidación lenta). Al principio hubo por todas partes muchas sustancias con las que se pudo combinar el oxígeno liberado por el nuevo tipo de organismos agentes de la fotosíntesis y parece ser que se deshicieron de este oxígeno peligrosamente reactivo capturándolo en el agua mediante iones de hierro (a los que se denomina habitualmente iones ferrosos, utilizando el latín, [33] en vez de iones de hierro, por razones obvias). Grandes

cantidades de un mineral de hierro conocido por el nombre de hematites, un óxido de hierro, se depositaron en los lechos de mares poco profundos donde vivían las primitivas formas de aquellos agentes de la fotosíntesis, entre hace tres mil y dos mil millones de años, y dicho mineral se extrae actualmente en lugares tales como el oeste de Australia, la península del Labrador y Ucrania.

Posteriormente, algunos de estos agentes de la fotosíntesis aprendieron a vivir con oxígeno libre, desarrollando enzimas que los protegían de la oxidación. Esto significaba que podían deshacerse del oxígeno enviándolo al medio ambiente, sin que existiera la preocupación de tener que encerrarlo en compuestos de hierro. Esto fue un desastre para muchas de las formas de vida que existían en aquella época (hace alrededor de dos mil millones de años), que no disponían de tal protección frente al poder reactivo del oxígeno y para las cuales este gas era un veneno. Aunque los blandos cuerpos de estos organismos unicelulares no han dejado ningún rastro en el registro fósil, tuvo que producirse una de las mayores extinciones de vida cuando los primeros miembros de esta tercera hornada de fotosintetizadores aprendieron cómo vivir con oxígeno libre, con el cual polucionaron el mundo entero, y este oxígeno libre tuvo afortunadamente (para ellos) un efecto secundario consistente en aniquilar a muchos de sus competidores. Además, el oxígeno también reacciona con gases tales como el metano que está en el aire, produciendo dióxido de carbono y agua, mientras acaba con todo el hidrógeno libre que emiten los volcanes, produciendo aún más agua. Cuando desaparecieron todas las sustancias con las que podía reaccionar, el oxígeno se integró en el aire hasta conseguir, hace aproximadamente mil millones de años, que la atmósfera empezara a parecerse mucho a la mezcla de gases que respiramos hoy: 78 por 100 de nitrógeno, 21 por 100 de oxígeno y sólo pequeños vestigios de todo lo demás, incluyendo el dióxido de carbono. [34]

Los animales aprovechan este aporte libre de oxígeno respirándolo y dejando que se combine con sustancias tales como el hidrógeno y el carbono en una especie de proceso de combustión muy lenta que proporciona la energía utilizada para vivir. Así, la evolución de las formas de vida animal, respirando oxígeno y empleando energía química en abundancia (si se compara con las plantas), empezó a suceder después de la difusión del oxígeno en la atmósfera.

La respiración, que es como se llama el proceso de combustión mediante oxígeno, aportó una ventaja tan enorme que, después de 1.500 millones de años en los que la vida en la Tierra se limitaba a organismos unicelulares, toda una variedad de seres animales pluricelulares irrumpió en el escenario de la evolución con gran rapidez, en un intervalo de unos cien millones de años, hace algo menos de dos mil millones de años. La razón de este éxito es que la energía química de que se dispone gracias a la respiración (combustión lenta) se obtiene con rapidez y de una manera sencilla, aunque hay que tener en cuenta que se dispone de ella porque la fotosíntesis realizada

en las plantas ha usado un procedimiento más lento y más complicado para tomar energía de la luz solar y utilizarla en la producción de oxígeno libre, entre otras cosas. Actualmente incluso las plantas usan también la respiración, además de la fotosíntesis, en sus procesos químicos, sin embargo, en un claro ejemplo del modo en que la evolución tiene que seguir modelando las formas existentes, en vez de volver al principio y comenzar de nuevo, continúan expulsando el oxígeno liberado en la fotosíntesis, y luego respiran oxígeno del aire durante la respiración.

El hecho de que formas de vida emergieran de los océanos y se trasladaran a tierra se debió también a la presencia de oxígeno en la atmósfera, aunque sucedió por otra razón bastante distinta. Antes de que hubiera oxígeno en el aire, toda la energía de la luz solar podía alcanzar la superficie de la Tierra. Gran parte de esta energía se encuentra en forma de radiación ultravioleta, y esta radiación es perjudicial para la vida; concretamente, los fotones de la luz ultravioleta tienen la cantidad de energía necesaria para romper algunas de las cadenas químicas que forman las moléculas de ADN (siendo ésta la razón por la que los rayos ultravioletas causan cáncer de piel). Pero los rayos ultravioletas también reaccionan con el oxígeno, en una serie de reacciones que producen una forma de oxígeno dotada de tres átomos del mismo en cada molécula, en vez de dos.[35] En esencia, lo que hacen los rayos ultravioletas es romper cada molécula de oxígeno, produciendo dos átomos libres que pueden integrarse cada uno de ellos en otra molécula de oxígeno. La forma triatómica del oxígeno se llama ozono, el cual, como resultado de estas reacciones, ha formado una capa en la atmósfera, situada aproximadamente entre los quince y los cincuenta kilómetros sobre el nivel del mar, y que se llama estratosfera (pero que no nos engañe este amplio intervalo de altitudes; el aire es tan poco denso en la estratosfera que si todo el ozono que contiene se llevara al nivel del mar, la presión de la atmósfera lo aplastaría reduciéndolo a una capa de sólo tres milímetros de espesor). También el propio ozono absorbe a su vez la radiación ultravioleta, por lo que la concentración de ozono de la estratosfera permanece equilibrada, ya que el ozono por una parte se crea y por otra se destruye en reacciones que utilizan la luz solar (se parece bastante a una bañera en la que los grifos están abiertos, pero el agua se va al mismo tiempo por el sumidero, de tal modo que el agua de la bañera permanece constante, aunque las moléculas de agua que contiene la bañera no sean nunca las mismas).

La vida emergió, trasladándose a tierra firme bajo la capa protectora de ozono, y son los propios seres vivos quienes mantienen esta capa protectora de ozono, ya que la fotosíntesis continúa aportando oxígeno al aire, mientras la respiración de los animales (y las plantas), así como otros procesos de oxidación (tales como los incendios forestales), retiran oxígeno del aire. Al final del siglo XX existe, con razón, una considerable preocupación por el daño que los gases producidos en los procesos industriales del ser humano, y que nunca habían existido anteriormente (los CFC),

están causando en la capa de ozono. Estos gases son extremadamente inertes, por lo que no se destruyen en ningún proceso químico o biológico que tenga lugar cerca del suelo. Finalmente, entran en la estratosfera, donde sus moléculas se rompen por la acción de los rayos ultravioletas (otro ejemplo de lo poderosos y dañinos que pueden ser los fotones de la radiación ultravioleta) y liberan cloro, que destruye el ozono. El famoso agujero de la capa de ozono situado sobre la Antártida y su contrapartida sobre el Ártico se han producido directamente como resultado de la presencia de los CFC en la atmósfera. Aunque se ha conseguido hacer más lenta la emisión de estos dañinos compuestos, hay tal cantidad de estas materias potencialmente perjudiciales que ya se encuentran en el aire o abriéndose camino hacia la capa de ozono, que pasarán décadas antes de que la capa de ozono se recupere, aunque no se emitieran ya más gases CFC.

Después de explicarlo, es realmente obvio el modo en que los organismos vivos interactúan con la atmósfera en un proceso de retroalimentación que mantiene aproximadamente constante la mezcla de gases que respiramos hoy en día. Algo más sorprendente, sin embargo, es el hecho de que esta clase de proceso de retroalimentación también relaciona la vida con la actividad tectónica que se produce en nuestro planeta. El dióxido de carbono del aire se disuelve en el agua de lluvia, formando anhídrido carbónico, que desgasta las rocas que contienen compuestos de calcio, silicio y oxígeno (silicatos de calcio). Esta actividad química libera iones de calcio e hidrogenocarbonatos (un ion de hidrogenocarbonato es, lógicamente, un ion de carbonato que lleva unido un átomo de hidrógeno), que finalmente van a parar al mar, donde son utilizados por organismos, tales como el plancton, para formar sus conchas calcáreas, hechas principalmente de carbonato cálcico. Cuando estas criaturas mueren, sus conchas caen al fondo del mar, constituyendo estratos de sedimentos ricos en carbonatos. En los primeros tiempos de la Tierra, este proceso tuvo que ser responsable de que la atmósfera de dióxido de carbono se volviera menos densa. Sin embargo, dado que la actividad geológica empuja la fina corteza del fondo marino hasta situarla bajo los bordes de la corteza continental, que es más espesa, los carbonatos son introducidos bajo los continentes y más profundamente en el interior de la Tierra, cuando se calientan y se funden. A presiones y temperaturas elevadas, cuando se forman nuevas rocas de silicatos, se libera dióxido de carbono y el gas se abre camino hacia la superficie y sale a la atmósfera durante las erupciones volcánicas.

Estos procesos geológicos necesitan mucho tiempo para desarrollar un ciclo completo del carbono. Sin embargo, una vez que el sistema está funcionando, la cantidad de dióxido de carbono que los volcanes liberan a la atmósfera cada milenio es aproximadamente constante, y a largo plazo se cuadra el balance con la cantidad de dióxido de carbono que se pierde por la combinación de los procesos biológicos y

geológicos. Pero ¿qué sucede si la temperatura del planeta cambia? Si desciende la temperatura, se evapora menos agua de los océanos, por lo que hay menos lluvia, menos erosión de las rocas, y se toma menos dióxido de carbono del aire. Como la acción de los volcanes continúa siendo constante, la concentración de dióxido de carbono aumenta, incrementando el efecto invernadero y calentando así la Tierra, con lo que aumentan las lluvias una vez más, hasta que se restablece el equilibrio. De manera similar, si el planeta se calienta por alguna razón (por ejemplo, un aumento de producción de calor en el Sol), habrá más lluvias y más erosión, con lo que se retira más dióxido de carbono del aire y se reduce así el efecto invernadero.

Esto es un ejemplo de retroalimentación negativa, es decir, una influencia estabilizadora en el medio ambiente de la Tierra. El aumento de temperatura del Sol reduce la fuerza del efecto invernadero producido por el dióxido de carbono. También resulta que los seres vivos desempeñan un papel en este proceso geológico de estabilización. El modo en que ha evolucionado el clima de nuestro planeta y el modo en que ha evolucionado la vida están tan íntimamente ligados que este proceso se denomina a veces «coevolución» de la vida y el clima.

En un momento de osadía, James Lovelock sugirió que la mejor forma de entender estos procesos es considerar todo el planeta, incluyendo las actividades biológica y geológica, como un único organismo vivo, que él llamó Gaia. Ésta es aún una teoría controvertida, aunque ha estado presente desde la década de los setenta. Pero, incluso sin aceptar la idea de que la Tierra está literalmente viva, algunos científicos planetarios han opinado que el planteamiento de Lovelock, que tiene en cuenta la contribución biológica en retroalimentaciones geológicas y climáticas de muchas clases, es un método fructífero para llegar a comprender cómo se ajusta el medio ambiente de nuestro planeta mediante un conjunto de controles y equilibrios naturales.

Ahora parece quedar fuera de duda que al menos es aceptable lo que se podría llamar una versión «moderada» de la teoría de Gaia: que la vida no sólo se ve afectada por el entorno físico, sino que también ella afecta al entorno físico, siquiera en escalas de tiempo geológicas, y a menudo (en realidad, habitualmente) de tal manera que se mantienen las condiciones adecuadas para la vida. Los distintos ciclos en los que interviene el dióxido de carbono están entre los aspectos más importantes de esta relación entre los seres vivos y el entorno físico, por lo que hemos de tener presente que extrayendo combustibles fósiles del suelo y quemándolos, así como talando bosques tropicales que toman dióxido de carbono del aire para la fotosíntesis, estamos alterando el equilibrio natural de un proceso que constituye un factor clave para regular la temperatura de la Tierra. [36] No obstante, antes de que analicemos un poco más detalladamente este efecto invernadero humano (o «antropogénico»), debemos echar un rápido vistazo a lo que sucedía con el clima antes de la

intervención del ser humano.

Ya hemos comentado cómo se ha visto afectado el clima de la Tierra por los largos y lentos procesos asociados con la deriva continental, cuando se producían épocas de glaciación, al pasar los continentes en su deriva por uno o ambos polos. Pero ahora queremos concentrarnos en analizar el estado en que se encuentra la Tierra actualmente y cómo ha estado durante varios millones de años anteriores al momento presente, durante el tiempo en que nuestra propia especie, el *Homo sapiens*, ha surgido y entrado en una etapa evolutiva.

La fuerza que hace funcionar toda la maquinaria meteorológica es, por supuesto, el Sol; pero el mayor aporte de energía solar no se recibe en la parte más alta de la atmósfera, sino en su parte inferior. La mayor parte de la energía procedente del Sol pasa libremente a través de la atmósfera y calienta la superficie de la Tierra. A su vez, es la superficie caliente de la Tierra (tierra o mar) la que calienta la parte inferior de la atmósfera. Así, la capa más caliente del aire se encuentra junto al suelo o al nivel del mar y, a medida que se asciende a través de la atmósfera, la temperatura desciende. La temperatura media al nivel del mar es de 15° C, pero desciende tan rápidamente que al llegar a la estratosfera, a aproximadamente quince kilómetros sobre el nivel del mar, se baja a una gélida temperatura de -60° C.

En la estratosfera (la capa de ozono) la temperatura vuelve a subir de nuevo, porque se absorbe energía en forma de rayos ultravioletas procedentes del Sol. Por consiguiente, en la parte más alta de la estratosfera, a unos cincuenta kilómetros sobre el nivel del mar, la temperatura del aire desciende al punto de congelación del agua, 0° C. Por encima de la estratosfera existe otra capa donde se produce enfriamiento. Se conoce como la mesosfera, que se extiende hasta una altitud de ochenta kilómetros. A altitudes aún mayores, los tenues gases que rodean la Tierra se ven afectados por partículas energéticas emitidas por el Sol, así como por rayos cósmicos que llegan desde las profundidades del espacio, y a duras penas se corresponden con lo que consideraríamos como aire. Lo que tiene importancia en cuanto a dirigir la evolución de la situación meteorológica es la troposfera, la capa en la que se decide el tiempo meteorológico y la estratosfera, que actúa como una tapadera de la troposfera.

La estratosfera actúa como una tapadera porque es una capa en la que la temperatura aumenta con la altitud. El viejo adagio «el aire caliente asciende» es verdadero sólo si el aire situado sobre el aire caliente es un aire más frío que el que asciende. Cuando el calor del Sol calienta la superficie de la Tierra, el aire caliente de la superficie se eleva a través de la troposfera, expandiéndose y enfriándose mientras sube. Pero no puede elevarse entrando en la estratosfera, ya que aquí la temperatura es más elevada que en la parte superior de la troposfera.[37]

Este calentamiento solar alcanza su máximo en los trópicos, donde el Sol a mediodía está casi vertical sobre la superficie terrestre. En latitudes más altas el Sol

queda más bajo sobre el horizonte, por lo que el efecto de calentamiento es menor, del mismo modo que, en cualquier latitud, por la mañana temprano o al final de la tarde es también menor que al mediodía. La razón de esto es que, cuando los rayos del Sol chocan con la superficie de la Tierra con un ángulo menor, se extienden por un área más amplia; una ejemplificación de ello es lo que sucede enfocando una linterna hacia un pedazo de cartulina. Si la cartulina se coloca perpendicular al haz de luz de la linterna, la luz forma sobre la superficie de la cartulina una mancha redonda, pequeña pero brillante. Sin embargo, si inclinamos la cartulina de modo que no esté perpendicular al haz, la luz se extiende formando una mancha alargada y más débil. Lo que sucede es que la misma cantidad de luz está cubriendo un área mayor, por lo que todas las zonas de la mancha alargada se iluminan más débilmente, a pesar de que la cantidad total de luz es la misma.

El aire caliente que se eleva en los trópicos es también húmedo, porque contiene vapor de agua evaporado del mar. Cuando el aire asciende y se enfría, esta humedad cae en forma de lluvia, siendo ésta la razón por la cual los trópicos son muy húmedos. El aire que se ha elevado desde la superficie da la vuelta en lo alto de la troposfera y se adentra a ambos lados del ecuador. Cuando ya se ha enfriado, vuelve cayendo hacia el suelo. El aire descendente se calienta de nuevo, porque está siendo aplastado por la presión que ejerce el aire que le sigue por detrás (el mismo efecto se produce en una bomba de bicicleta, porque el aire que está dentro de la bomba se calienta cuando lo comprimimos para empujarlo al interior del neumático). El aire caliente absorbe humedad, por lo que allí donde llega al suelo se forman franjas desérticas, que es lo que sucede a ambos lados del ecuador, por ejemplo, en el desierto del Sáhara. Parte de este aire seco fluye de nuevo hacia el ecuador a la altura de la superficie terrestre, ocasionando los vientos alisios y absorbiendo mientras tanto humedad; otra parte de este aire desarrolla ciclos más conectivos, que en conjunto se llevan aire caliente del ecuador y lo conducen hacia los polos, produciendo los vientos predominantes en latitudes más altas, incluidos los vientos del oeste que llegan del Atlántico y determinan el clima en las islas Británicas.

El clima es una especie de promedio del tiempo meteorológico, producido por los efectos combinados de todos estos procesos a lo largo de muchos años. Sin embargo, el clima siempre está cambiando, dentro de una diversidad de escalas cronológicas, porque las distintas aportaciones que recibe el tiempo meteorológico continuamente están variando. Incluso cambian a lo largo del año, aunque estamos tan acostumbrados a ello que no lo consideramos un cambio climático.

La Tierra es esencialmente una bola de roca, cubierta por una fina capa de atmósfera y océano, que gira sobre su eje una vez cada veinticuatro horas y describe una órbita alrededor del Sol una vez al año. Pero el eje en torno al cual rota cada veinticuatro horas (una línea imaginaria que une el polo Norte y el polo Sur) está

inclinado con respecto al plano en el cual nuestro planeta describe la órbita alrededor del Sol. Esta inclinación se mantiene igual a lo largo de todo el año, apuntando siempre a la misma dirección; pero cuando nos movemos en torno al Sol, unas veces estamos a un lado del astro y otras (seis meses antes o después) estamos en el lado opuesto. Así pues, el polo Norte está a veces inclinado hacia el Sol y otras veces se inclina hacia el lado contrario, mientras el polo Sur siempre está inclinado en sentido opuesto al de la inclinación del polo Norte. Exactamente el mismo proceso que hace que las latitudes más altas sean más frías que las latitudes más bajas la dispersión de la energía solar entrante en un área más amplia cuando el Sol está bajo en el cielo hace que el hemisferio que está inclinado hacia el lado contrario al Sol esté más frío que el hemisferio que está inclinado hacia el Sol. Es por esta razón por lo que tenemos el ciclo de las estaciones y por lo que el Sol se eleva a mayor altura en el cielo durante el verano y a menor altura durante el invierno.

Aunque la inclinación de la Tierra se puede considerar constante durante un intervalo de tiempo tan breve como es un año, este planeta en rotación es realmente como un giroscopio oscilante, y cualquiera que haya visto la peonza de un niño en acción sabrá cómo puede oscilar un giroscopio cuando está rotando. La primera oscilación de la Tierra describe una trayectoria circular, en gran parte debida a la influencia gravitatoria del Sol y de la Luna que tira de nuestro planeta mientras rota; sin embargo, la oscilación es un proceso muy lento comparado con el ciclo de las estaciones. Dicha oscilación describe una línea imaginaria: apuntando en línea recta hacia arriba sobre el polo Norte, describe un círculo en el cielo, siguiendo un ciclo que varía en longitud pero tarda en completarse unos veinte mil años. Este cambio en la dirección hacia la cual apunta el polo (llamado precesión) causa una aparente deriva de las estrellas a través del cielo, cuando son observadas desde la Tierra. Una consecuencia de esta oscilación es que la pauta que siguen las estaciones cambia lentamente a medida que transcurren los milenios. Otra consecuencia es que hace diez mil años la estrella polar actual no habría sido una buena guía para dirigirse hacia el norte.

Al mismo tiempo que se produce esta precesión, el ángulo de inclinación de la Tierra cambia, balanceándose ligeramente hacia arriba y hacia abajo en el transcurso de los milenios. El modo en que se desarrolla este balanceo se puede calcular muy fácilmente aplicando las leyes de Newton. Varía entre una separación de 21'8 grados y una separación de 24'4 grados con respecto a la vertical, durante un ciclo de alrededor de 41.000 años. Actualmente la inclinación es de 23'4 grados y está disminuyendo. Si no intervienen otros factores, significa que las diferencias entre el verano y el invierno son hoy en día menos extremas que hace unos pocos miles de años. Los veranos son un poco más frescos y los inviernos un poco más templados de lo que solían ser.

Un tercer factor astronómico que influye en el clima es la red de interacciones de las fuerzas gravitatorias entre el Sol y todos los planetas del sistema solar; este conjunto de interacciones estira ligeramente la órbita de la Tierra, haciéndola variar de más circular a más elíptica, y posteriormente la vuelve a comprimir para hacerla circular, todo ello durante un período de poco más de cien mil años. Cuando la órbita es más o menos circular, el planeta recibe cada día la misma cantidad de calor del Sol. Cuando la órbita es más elíptica, durante una parte del año la Tierra está más cerca del Sol y recibe más calor, mientras que en otras épocas del año está más lejos y recibe menos calor. Pero lo más importante, entre todas estas influencias astronómicas sobre el clima, es que no cambian la cantidad total de calor que el conjunto de todas las zonas de la Tierra reciben durante un año completo. Todo lo que pueden hacer es reordenar la distribución de calor entre las estaciones.

Resulta que estos efectos, aunque son pequeños, ejercen hoy en día una gran influencia sobre el clima a causa de la geografía actual del globo. Hay hielo sobre ambos polos, lo que es en sí mismo inusual. El hielo que cubre el polo Sur es algo que ya ha sucedido anteriormente en la historia de la Tierra, por la existencia de un continente (la Antártida) que se encuentra situado exactamente sobre el polo y bloquea el flujo de agua caliente hacia latitudes altas del hemisferio sur. Sin embargo, el hielo que cubre el polo Norte es el resultado de una distribución geográfica mucho más rara (posiblemente única en la historia de nuestro planeta) y es también mucho más interesante.

La razón por la que el Ártico está cubierto de hielo es también que el agua caliente no puede penetrar fácilmente en latitudes altas. Pero esta vez se debe a que el océano Ártico está casi totalmente rodeado de tierra. El polo está cubierto por las aguas y el hielo flota en la superficie del océano Ártico. Como consecuencia, todo esto hace que el hemisferio norte sea muy sensible a los ritmos astronómicos asociados a la oscilación de esta peonza llamada Tierra.

Estas teorías relativas a la influencia astronómica sobre el clima fueron formuladas por primera vez por el escocés James Croll en la década de 1860, y perfeccionadas por el astrónomo yugoslavo Milutin Milankovitch durante la primera mitad del siglo XX. A menudo reciben el nombre de Modelo de Milankovitch. Sin embargo, no se llegó a demostrar que éste fuera un buen modelo hasta la década de 1970, cuando una combinación de, por una parte, ordenadores electrónicos, utilizados para calcular con exactitud cómo han cambiado las influencias astronómicas durante los últimos cientos de miles de años y, por otra parte, el estudio de muestras obtenidas mediante perforaciones realizadas en niveles marinos profundos, mostró que el modelo encajaba casi perfectamente. Como todos los buenos modelos, el Modelo de Milankovitch concuerda con los resultados experimentales.

Lo que nos dice esta combinación de observaciones y teoría es que, dada la

geografía actual del planeta, las épocas de glaciación se producen cuando los veranos son especialmente fríos en el hemisferio norte. Esto parece extraño a primera vista, porque los veranos fríos, como ya hemos visto, suelen ir acompañados de inviernos relativamente cálidos. Se podría pensar que, para hacer que el hielo se extienda desbordando la región del Ártico e invadiendo las tierras que lo rodean, se necesitarían inviernos fríos para generar mucha nieve. Pero es preciso recordar que la cantidad media de calor que recibe la Tierra cada año es siempre la misma, por lo que unos inviernos fríos irían acompañados de veranos calurosos, y éstos son lo más indicado para fundir la nieve. La clave está en que la Tierra actualmente está siempre lo bastante fría para que caiga nieve en invierno sobre las tierras que rodean el helado océano Ártico. Para que el hielo se extienda, lo que hace falta es que las temperaturas del verano sean tan frescas que una parte de esta nieve no se derrita y se vaya acumulando de año en año, hasta formar finalmente nuevas capas de hielo.

Sin embargo, por una vez, «finalmente» puede ser dentro de no mucho tiempo. Con tan sólo una delgada capa de nieve que dure todo el verano, hay suficiente para producir un fuerte impacto en el clima local, porque refleja el calor solar que llega, el cual en otro caso habría calentado el suelo que está cubierto por la nieve. Al comienzo del invierno siguiente el suelo está más frío de lo que estaría sin la nieve, por lo que es fácil que la próxima nevada se quede allí sin derretirse. De esta manera, el equilibrio natural cambia rápidamente, produciéndose una época de glaciación plena, si las condiciones contempladas en el Modelo de Milankovitch son correctas; el cambio tarda en realizarse seguramente menos de un milenio, quizá sólo un par de siglos. Sin embargo, lo que resulta mucho más difícil es calentar el planeta para que salga de un período de glaciación, debido al modo en que el hielo y la nieve reflejan, rechazándolo, el calor que llega del Sol. Se necesitan varios miles de años para cambiar unas condiciones de máximo de glaciación por las de máximo interglacial.

En realidad, es totalmente incorrecto el modo en que hemos contado esta historia, aunque resulte lógico desde el punto de vista de la gente que está viviendo en un período relativamente caliente. Dada la geografía actual de la Tierra, el estado natural del planeta es que se encuentra al borde de un máximo de glaciación. Sólo se funde una parte suficiente del hielo en el hemisferio norte, para dar así una tregua en el avance del hielo, cuando todos los ciclos de Milankovitch actúan al mismo tiempo para aportar la máxima cantidad de calentamiento estival. Pero, como ya hemos dicho, actualmente los veranos se están volviendo cada vez más fríos.

Si hay tres ciclos diferentes actuando al mismo tiempo para producir un cambio de este tipo en el clima, y cada uno de ellos está sujeto a ligeras variaciones, la pauta no se repite exactamente. Sin embargo, dicho de una manera aproximada, la pauta de cambio climático durante más o menos los últimos cinco millones de años ha sido de tal modo que los máximos de glaciación (cada uno con una duración algo superior a

los cien mil años) han estado separados por períodos más breves relativamente cálidos, de unos diez o veinte mil años de duración cada uno. Estos períodos cálidos se llaman interglaciales, y estamos viviendo ahora en un interglacial que comenzó hace unos quince mil años. De todas formas la Tierra está actualmente más fría que lo que haya podido estar durante la mayor parte de su historia, ya que en una escala de tiempo geológica es rara la existencia de hielo en ambos polos. Toda la civilización humana se ha desarrollado durante este interglacial en concreto, y es posible que nosotros estemos aquí gracias a este ciclo de épocas glaciales e interglaciales.

No fue tanto el frío, sino la sequía asociada al máximo de glaciación, casi con toda seguridad, lo que hizo a nuestros antepasados salir de los bosques del África oriental y los puso en el camino de llegar a convertirse en seres humanos. La sequía acompaña siempre al aumento de las superficies cubiertas de hielo, por un lado debido a que el agua está retenida en el hielo, y por otro lado también porque cuando el planeta está frío se evapora menos agua de los océanos, por lo que hay menos agua que pueda caer en forma de lluvia.

En los tiempos en que llegó a establecerse un ritmo repetitivo de las glaciaciones y de los períodos interglaciales, nuestros antepasados estaban llevando una vida bastante cómoda y afortunada como habitantes de los bosques y pastos de África oriental, de lo cual da testimonio una abundante cantidad de restos fósiles. Pero, en aquella época, la Tierra estaba sumida en la sucesión de glaciaciones y períodos interglaciales que se ha estado produciendo desde entonces, y que parece haber sido única en la larga historia de nuestro planeta. Cada vez que sobrevenía una glaciación, los bosques se veían afectados por la sequía y comenzaban a declinar, mientras las llanuras se volvían más áridas. Éstas son precisamente las condiciones duras que hacen funcionar la rueda de la evolución, ya que las especies que se adaptan peor son las que más sufren, mientras prosperan aquellas que son capaces de resistir las condiciones cambiantes. Sin embargo, durante los interglaciales, cuando el planeta se calentaba, las condiciones meteorológicas volvían a ser las adecuadas para que los árboles y la hierba (y la vida vegetal en general) florecieran, con lo que aquellos individuos y especies que habían sobrevivido a la sequía tenían una oportunidad para multiplicarse y gozaban de una auténtica explosión demográfica.

Cada vez que había una sequía, sólo sobrevivían los que eran más capaces de adaptarse; cada vez que volvían las lluvias, los descendientes de estos individuos adaptables proliferaban. Se pueden distinguir dos modos según los cuales pudo realizarse la adaptación evolutiva en tales condiciones, partiendo de lo que, a falta de un término más adecuado, llamaríamos el «hombre mono». Un modo de salir bien parado pudo ser adaptarse aún mejor a los bosques que quedaban, retirándose a la parte más interna de estos bosques cada vez que se producía la sequía. Este desarrollo evolutivo conduciría al moderno chimpancé y al gorila actual. La otra manera de

sobrevivir sería evolucionar para adaptarse cada vez mejor a las llanuras secas que irían apareciendo al reducirse el bosque. Como ya ha sucedido muy frecuentemente en el pasado de la evolución, serían las criaturas menos afortunadas las que se verían obligadas a conformarse con vivir en las zonas marginales; en este caso, los monos menos afortunados se verían obligados a salir de los bosques que se iban encogiendo y a defenderse lo mejor que pudieran en las llanuras. Si una sequía hubiera durado un millón de años, puede que aquellos que, a su pesar, habían tenido que habitar las llanuras se hubieran extinguido por completo; sin embargo, después de cada período de unos cien mil años de sequía, al volver las lluvias los supervivientes tuvieron una oportunidad para reagruparse y aumentar su número, casi como un ejército al que se permite una retirada del frente para descansar y tener un poco de esparcimiento antes de volver a la batalla. Está claro que en estas condiciones de tener que soportar un entorno medioambiental cambiante, la inteligencia supondría una ventaja evolutiva especial.

No es solamente el registro fósil el que apoya esta teoría que, aunque especulativa, es sin embargo plausible. En números redondos, el ADN humano coincide en más de un 98 por 100 con el ADN de los chimpancés y de los gorilas (tengamos en cuenta que dos seres humanos comparten poco más del 99'8 por 100 de su ADN). Somos sólo un 1 por 100 humanos, y aproximadamente en el 99 por 100 coincidimos con los monos. Los biólogos moleculares pueden medir a qué velocidad se acumulan los cambios en el ADN al sucederse las generaciones, comparando el ADN de muchas especies vivientes con restos fósiles de la época en que se produjo el fraccionamiento en diferentes líneas a partir de un antepasado común.

El calibre de la diferencia entre nosotros mismos y los otros monos africanos apunta a un fraccionamiento en tres líneas a partir de la línea evolutiva de un antepasado común que vivió hace alrededor de cinco millones de años. Este fraccionamiento se produjo exactamente cuando empezó el modelo distintivo de cambios climáticos que consideramos como normal. Dicho fraccionamiento es tan reciente que parece probable que ese antepasado común de nosotros mismos, de los chimpancés y de los gorilas fuera más parecido al ser humano que a los monos; concretamente, ya había desarrollado una posición erecta. Fue la presión de las repetidas sequías y la necesidad de adaptarse aún mejor a la vida en los bosques lo que convirtió al chimpancé y al gorila en lo que son actualmente, de tal modo que, dando la vuelta a la expresión vulgar habitual, más que decir que el género humano descende de los monos, tendría más sentido decir que los monos descienden del hombre, o al menos del homínido.[38]

Debemos nuestra existencia a una especie de efecto de trinquete, resultante de estos ritmos naturales del clima. La tendencia natural sería, según los ritmos de Milankovitch, llevar la situación del planeta hacia atrás, sumergiéndolo en un

máximo de glaciación dentro de unos pocos miles de años. Sin embargo, dado que las actividades del ser humano son ahora una parte importante de la maquinaria meteorológica, parece probable que, como resultado, la próxima glaciación se pueda posponer indefinidamente, sea por accidente o a propósito.

La concentración de dióxido de carbono, el principal gas causante del efecto invernadero, es en la atmósfera de sólo un 0'03 por 100, aproximadamente, es decir, 300 partes por millón (ppm). Sin embargo, desde la revolución industrial del siglo XIX, ha aumentado de 280 ppm a 350 ppm, y esto es el resultado únicamente de la actividad humana, a saber, de quemar combustibles fósiles y destruir bosques. Esta variación constituye un aumento relativo del 25 por 100 y ya está cambiando la meteorología del planeta. Éste se ha calentado en conjunto un grado (Celsius) desde la década de 1880, y se sigue vertiendo dióxido de carbono a la atmósfera a un ritmo cada vez más rápido. Hasta dónde puede llegar este calentamiento depende de la cantidad de dióxido de carbono que aún se vaya a incorporar al aire y de la velocidad a la que esto se lleve a cabo; sin embargo, los cálculos más moderados sugieren que se producirá un calentamiento de un grado más durante, más o menos, los próximos veinte años, después de los cuales las previsiones serían ya más un asunto de adivinos que de científicos. No obstante, incluso con este calentamiento aparentemente moderado se alcanzarían temperaturas por encima de cualquiera de las que se han dado durante el interglacial actual y, por consiguiente (puesto que es obvio que las temperaturas eran más bajas durante la última glaciación), más elevadas que cualquiera de las que se han dado en la Tierra durante más de cien mil años.

El tiempo meteorológico del siglo XXI será distinto de cualquiera de los que se han dado en la Tierra desde que empezó la civilización, teniendo en cuenta que durante la primera mitad de ese siglo el planeta se estará calentando a una velocidad cincuenta veces mayor, aproximadamente, que al final de la última glaciación, cuando el cambio desde el máximo de glaciación hasta las condiciones de máximo interglacial supusieron un calentamiento de unos 8° C durante más o menos cinco mil años. Es imposible predecir qué efecto puede ocasionar en el mundo de los seres vivos un cambio tan dramático del entorno físico. Por lo tanto, escapándonos como los cobardes, no vamos a intentar predecir el futuro clima de nuestro planeta, sino que, en vez de eso, nuestro próximo paso será entrar en el universo en toda su amplitud, desde donde podemos abordar problemas mucho más sencillos, tales como explicar qué es lo que hace que el Sol y las estrellas brillen, y de dónde sacó el Sol su familia de planetas.

El sol y su familia

El Sol es una estrella. Es una estrella de lo más corriente, ni especialmente grande, ni especialmente pequeña, ni especialmente brillante, ni especialmente apagada, y que se encuentra aproximadamente en la mitad de su ciclo vital. La única razón por la que parece diferente de otras estrellas del cielo es porque estamos muy cerca de ella; la Tierra describe una órbita alrededor del Sol a una distancia de sólo 150 millones de kilómetros y tarda un año en realizar el viaje completo. La mayoría de los libros de astronomía (y muchos maestros de escuela) nos dirían que en la familia del Sol hay nueve planetas, incluyendo la Tierra. Sin embargo, esto no es exacto, ya que el más distante (con respecto al Sol) de estos nueve objetos, Plutón, es claramente un objeto de otro tipo, diferente de los otros ocho planetas. Se podría definir más bien como un trozo bastante grande de detritos cósmicos, más parecido a los cometas y a los asteroides que ensucian el sistema solar, que a un planeta auténtico.

Una de las características más obvias de la peculiaridad de Plutón en comparación con otros planetas es que, aunque por término medio sea, desde luego, el planeta más alejado del Sol en el conjunto de los nueve planetas reconocidos, tiene una órbita más elíptica que la de cualquiera de los otros planetas, por lo que a veces (por ejemplo, entre 1979 y 1999) está realmente más cerca del Sol que Neptuno, el más distante de los otros planetas con respecto al Sol. Ningún otro planeta cruza la órbita de ninguno de sus vecinos.

Los astrónomos miden las distancias en el sistema solar utilizando como referencia la distancia media entre la Tierra y el Sol, que se define como una unidad astronómica (1 UA); la distancia media de Plutón al Sol es algo menos de 40 UA, pero la distancia real oscila entre 30 UA y 50 UA en distintos puntos de su órbita. Así, en el punto más distante del Sol, Plutón se encuentra cincuenta veces más alejado que nosotros. Tiene un diámetro de sólo 2.320 km (dos tercios del diámetro de nuestra Luna) y su masa es sólo un 0'3 por 100 de la masa de la Tierra. Plutón va acompañado por una luna llamada Caronte, cuyo diámetro es de 1.300 km (más de la mitad del diámetro de dicho planeta) y describe una órbita alrededor de Plutón a una distancia de tan sólo 19.400 km. Ambos objetos espaciales están constituidos en su mayor parte por agua helada y metano congelado, siendo su densidad media menos del doble de la densidad del agua. Sus superficies están a 50 grados Kelvin (por debajo de menos 220 grados Celsius) y tardan 248 años en describir una órbita completa alrededor del Sol.

Dejando a Plutón a un lado, podemos concentrarnos en los ocho auténticos planetas que son los miembros más evidentes de la familia del Sol, así como en el estudio más detallado de los detritos cósmicos del sistema solar. La familia del Sol se divide en dos grupos de cuatro: por una parte, cuatro planetas rocosos, pequeños, que

describen sus órbitas en la parte interior del sistema solar y, por la otra, cuatro grandes planetas gaseosos, que orbitan en la parte exterior del sistema solar. Estos dos cuartetos están separados por una franja de detritos (entre las órbitas de Marte y Júpiter) conocida como cinturón de asteroides.

Todas las estrellas brillantes relucen, como el Sol, porque generan calor en su interior, mediante procesos de fusión nuclear. Sin embargo, los planetas son visibles sólo porque reflejan la luz proveniente del Sol. Por consiguiente, son unos objetos mucho menos luminosos, siendo éste el motivo por el cual hasta hace poco no ha habido pruebas directas de la existencia de planetas que orbiten en torno a otras estrellas. Los astrónomos estaban seguros de que debían estar allí, pero los planetas eran demasiado pálidos para ser vistos. No obstante, en la década de 1990 se hallaron pruebas directas de la existencia de planetas que describen órbitas en torno a otras estrellas, gracias a mediciones extremadamente precisas del modo en que dichas estrellas se balancean en el cielo. El balanceo se interpreta como el efecto de la gravedad de un gran planeta que ejerce una tracción sobre la estrella cuando orbita alrededor de ella, tirando de ésta primero en un sentido y luego en el otro. La técnica sólo revela la existencia de planetas grandes (como Júpiter, el planeta de mayor tamaño en nuestro sistema solar), pero resulta una extrapolación razonable inferir que si se ven planetas como Júpiter en órbita alrededor de estrellas como el Sol, entonces habrá probablemente planetas como la Tierra que describan órbitas alrededor de al menos algunas de esas estrellas.

Otro logro de la década de 1990 ha sido la detección directa de nubes de materia en polvo en un grueso disco que rodea a algunas estrellas jóvenes. Incluso antes de que dichos discos fueran identificados (y fotografiados directamente mediante el telescopio espacial Hubble y otros instrumentos), los astrónomos parecían tener ya un buen modelo del modo en que los planetas del sistema solar pudieron haberse formado a partir de un disco de polvo que rodeaba al Sol cuando era joven. La existencia de unos discos que son exactamente del tipo requerido por este modelo y que describen órbitas en torno a estrellas exactamente iguales al joven Sol (según los modelos astrofísicos, que discutiremos con más detalle en el próximo capítulo) deja poco margen para dudar de que realmente sí comprendemos cómo se formaron el Sol y su familia. Sin embargo, esto no es lo mismo que una comprobación de los modelos mediante experimentos llevados a cabo en los laboratorios. Está claro que no podemos comprobar nuestro modelo de la formación del sistema solar haciendo otro sistema solar.

Esto pone de manifiesto una diferencia importante entre la astronomía y la mayoría de las ciencias de las que ya hemos hablado con anterioridad en este libro, y significa que, hasta cierto punto, los modelos astronómicos son siempre menos satisfactorios que los mejores modelos de que disponemos para procesos que se

desarrollan aquí en la Tierra. Sin embargo esta diferencia de calidad puede ser muy pequeña en algunos casos y esperamos poder dejar claro que esto ciertamente no significa que en general los modelos astronómicos no sean más que especulaciones arbitrarias (algunos lo son, pero de éstos no vamos a tratar aquí). Los mejores de estos modelos se verifican por comparación con el modo en que se desarrollan las cosas en el universo real, o mediante simulación por ordenador y, en muchos casos (incluido, por ejemplo, el modo en que las estrellas generan calor), utilizando datos de experimentos realizados en laboratorios situados en la Tierra, en los que se mimetizan al menos algunos de los procesos clave que intervienen en aquellos fenómenos astrofísicos que son de interés.

El mejor modelo que tenemos para explicar cómo se formaron el Sol y su familia de planetas (y otros fragmentos de detritos) relaciona estos acontecimientos con la estructura de toda la galaxia de estrellas (la Vía Láctea) en la que vivimos. El tema de las estrellas se comenta con más detalle en el capítulo 10, pero las características más importantes son que el Sol es una estrella entre unos pocos cientos de miles de millones de estrellas que juntas forman un sistema aplanado, en forma de disco, que tiene aproximadamente un diámetro de cien mil años luz y un espesor de un par de miles de años luz.[39]

El sistema solar se encuentra a alrededor de dos tercios de la distancia que hay entre el centro y el borde de este disco y, como el resto de la población del disco, describe una órbita alrededor del centro del mismo. Si viajáramos a una velocidad de 250 kilómetros por segundo, tardaríamos unos 225 millones de años en recorrer una órbita completa, un intervalo que se denomina a veces «año cósmico». Como muchas de las otras galaxias en forma de disco, nuestra Vía Láctea se caracteriza por dos rasgos distintivos conocidos como brazos espirales, que se enrollan alejándose del centro. Estos brazos se han puesto de manifiesto por la emisión de radiaciones procedentes de las nubes de gas hidrógeno que contienen.[40] Las formas en espiral son ondas de presión, por lo que todo lo que orbita alrededor de la galaxia se aplasta cuando pasa a través de estos brazos. Los brazos equivalentes de otras galaxias se ven brillantes porque contienen muchas estrellas jóvenes calientes y, además, la razón por la que contienen dicho tipo de estrellas es que las nubes de gas y polvo que orbitan alrededor de la galaxia resultan aplastadas por la onda de presión cuando pasan a través de un brazo espiral. Por otra parte, las estrellas de mayor masa que surcan los brazos espirales desarrollan muy rápidamente sus ciclos de vida y explotan, emitiendo ondas de choque que producen un aplastamiento añadido en todas las nubes de gas y polvo que se encuentren próximas.

Las estrellas del disco están distribuidas en grupos y muchas han nacido al mismo tiempo a partir de una única gran nube de gas y polvo que se colapsaba. En primer lugar forman lo que se conoce como un cluster (o enjambre) abierto de estrellas,

existiendo más de setecientos de estos clusters abiertos dentro de un radio de ocho mil años luz en torno al Sol. Sin embargo, la cohesión interna de estos clusters por efecto de la gravedad no es lo suficientemente fuerte como para mantener su identidad, por lo que las estrellas individuales se dispersan y se distancian una de otra para seguir las órbitas que describen por la galaxia, de tal manera que después de unos pocos cientos de millones de años no hay ya ningún modo de decir qué estrellas nacieron juntas de una misma nube.

Nuestro Sol se formó de esta manera hace unos cinco mil millones de años y desde entonces ha estado orbitando por la Vía Láctea, habiendo realizado unas 20 vueltas aproximadamente durante toda su vida hasta la fecha actual. La nube de materia a partir de la cual se formaron el Sol y su familia (junto con las otras estrellas de un cluster abierto que ahora aparece ampliamente disperso), estaba constituida casi en su totalidad por hidrógeno (alrededor del 75 por 100) y helio (alrededor del 25 por 100), unos gases que quedaron después del Big Bang en que nació el universo (véase el capítulo 11). Pero esta nube de materia también estaba salpicada de una pequeña cantidad (alrededor del 1 por 100) de elementos más pesados, fabricados todos ellos en el interior de unas estrellas (tal como explicamos en el próximo capítulo) y que se diseminaron por el espacio al morir estas estrellas. Cuando el fragmento de nube interestelar que se iba a convertir en el sistema solar empezó a contraerse, colapsándose bajo su propio peso, su núcleo se calentó, porque al empequeñecerse la nube se liberó energía gravitatoria; si dos objetos se atraen mutuamente por el efecto de la gravedad, obviamente hay que aplicarles una cantidad de energía para separarlos; cuando se aproximan el uno al otro, esa misma cantidad de energía se libera, siendo esto aplicable a cualquier molécula de gas que formara parte de la nube que estaba contrayéndose. Finalmente, su interior llegó a estar tan caliente (a unos 15 millones de grados) que los núcleos de hidrógeno empezaron a convertirse en núcleos de helio, liberándose energía durante este proceso.

En un proceso de pasos múltiples, cuatro protones (núcleos de hidrógeno) se combinan para formar una partícula alfa (un núcleo de helio). La masa total de la partícula alfa es precisamente un 0'7 por 100 menos que la masa de cuatro protones juntos, por lo que esta diferencia de masa se convierte en energía pura cada vez que se lleva a cabo todo este proceso. Con el fin de estabilizar el Sol e impedir que siga colapsándose en la actualidad, cinco millones de toneladas de masa se convierten de esta manera en energía pura cada segundo (aproximadamente el equivalente a convertir un millón de elefantes en energía pura cada segundo). Incluso después de cinco mil millones de años de producir energía a esta velocidad prodigiosa, el Sol hasta ahora únicamente ha utilizado cerca del 4 por 100 de su reserva inicial de hidrógeno, siendo sólo el 0'7 por 100 de este 4 por 100 lo que realmente se ha convertido en radiación y se ha perdido en el espacio. La masa equivalente a toda esta

energía emitida por el Sol en todo su tiempo de vida hasta la fecha es alrededor de cien veces la masa de la Tierra.

En unos cinco mil millones de años más, el Sol empezará a tener problemas porque habrá utilizado todo el hidrógeno que tiene en su núcleo. Habrá todavía hidrógeno en abundancia en las capas exteriores de esta estrella, pero el núcleo caliente estará compuesto casi enteramente por helio, es decir, la «ceniza» de su vida de combustión nuclear. En ese momento de su vida, el núcleo del Sol se reducirá y llegará a estar aún más caliente, lo cual permitirá que tengan lugar otras reacciones de fusión nuclear, que convertirán los núcleos de helio en núcleos de carbono a una temperatura de aproximadamente 100 millones de grados Celsius. El calor suplementario generado en el núcleo en este estadio de su vida hará que se dilaten las capas exteriores del Sol, convirtiéndolo en un tipo de estrella conocido como gigante roja, que se tragará al planeta más próximo, Mercurio. Finalmente, después de otros mil millones de años, más o menos, cuando también se haya agotado su reserva de helio, el Sol dejará de generar energía en su interior y se apagará, convirtiéndose en escoria refrigerante, con un tamaño no mayor que el de la Tierra, y recibirá el nombre de enana blanca. Otras estrellas de mayor masa avanzan por sus ciclos vitales más rápidamente, porque tienen que quemar combustible a un ritmo mayor para mantenerse frente al tirón centrípeto de la gravedad; estas estrellas también hacen cosas más interesantes al final de sus vidas, tal como se explica en el capítulo 10.

Sin embargo, desde el punto de vista de la familia de planetas del Sol (por tanto, de nosotros mismos) lo que ahora importa es que durante un período completo de diez mil millones de años una estrella como el Sol genera energía a una velocidad más o menos constante, mientras los planetas giran en rueda en torno a él y se desarrollan cada uno en su modo particular, incluyendo en estas peculiaridades (al menos en uno de estos planetas) el desarrollo de la vida. Pero ¿por qué son todos los planetas tan diferentes uno de otro?

La naturaleza de cada uno de los planetas del sistema solar, cuando se condensó a partir de una nube de gas y polvo que se colapsaba, quedó determinada en primer lugar por la rotación y después por el calor que enviaba el propio Sol una vez que se formó. Cualquier nube de materia que exista en el espacio se ve obligada a estar rotando; es insignificante la probabilidad de que quede en equilibrio en un estado estacionario. Cuando la nube empezaba a colapsarse hacia adentro, es posible que rotara más rápido, del mismo modo que un patinador sobre hielo puede aumentar su velocidad de rotación recogiendo sus brazos hacia adentro. Esto se debe a la conservación de lo que se denomina momento angular. El momento angular de una masa que gira en círculo depende de la cantidad de esta masa, de su distancia al centro del círculo y de la velocidad a la que se mueve. Así, si esa misma masa se acerca más hacia el centro, ha de moverse más rápido para conservar su momento

angular. La mayor parte de la masa de la nube que se condensó para formar el sistema solar se asentó en forma de esfera en el centro, constituyendo el Sol.[41] Sin embargo, esto sólo fue posible debido a que el momento angular de la nube se transfirió en gran medida a un disco de materia que quedó situado en torno a la esfera central de gas. Al girar más rápido, y también al estar más alejado del centro, este disco pudo almacenar la mayor parte del momento angular original que tenía aquella materia que se convirtió en el Sol. El Sol se quedó con la mayor parte de la masa, pero el disco se llevó la mayor parte del momento angular.

Los planetas y lunas del sistema solar se formaron en este disco en el que se hacían remolinos, y conservaron su momento angular. Todos los planetas orbitan en torno al Sol en la misma dirección, y asimismo casi todas las lunas describen órbitas en torno a estos planetas en la misma dirección. Incluso el giro de los planetas en torno a sí mismos, con la excepción de Venus y Urano (que parecen haber recibido importantes impactos cósmicos), se produce en la misma dirección, almacenando un poco más del momento angular original, y el propio Sol, que rota una vez cada 25'3 días, gira también en la misma dirección. Ésta es una prueba clara de que el Sol y los planetas se formaron a partir de una única nube de gas que se encontraba en rotación, pero también de que el Sol, por ejemplo, no fue captando de uno en uno los planetas cuando iba orbitando por la galaxia. Si éste hubiera sido el caso, las órbitas de estos planetas tendrían una inclinación aleatoria, no estarían en un único disco, y la dirección de los planetas en estas órbitas sería aleatoria también.

Al igual que sucede en los discos que rodean a algunas estrellas jóvenes, según sabemos por pruebas directas que tenemos actualmente, los dos planetas más grandes del sistema solar, Júpiter y Saturno, son en sí mismos como sistemas solares en miniatura, escoltados por familias de lunas que describen órbitas en torno a ellos de la misma forma que los planetas que giran en órbita alrededor del Sol. Está claro que estos planetas gigantes se formaron del mismo modo que el Sol, aunque a menor escala, generando discos de materia en torno a ellos cuando se contraían, y formando tanto lunas como anillos, a partir de unos detritos en los que se almacenaba el momento angular.

Se supone que el proceso de formación de los planetas había comenzado ya antes de que la bola central de gas que iba a convertirse en el Sol hubiera llegado a estar lo suficientemente caliente como para desencadenar la fusión nuclear.

Unas diminutas pizcas de polvo que estaban en la nube original se habrían adherido entre sí para formar pequeños granos esponjados de unos pocos milímetros de diámetro, y luego estos granos habrían colisionado unos con otros para adherirse y formar a su vez granos aún mayores. En las primeras etapas de este proceso, se supone que los granos habrían estado inmersos en gas, bombardeados constantemente por moléculas de gas dentro de la nube que se colapsaba, de tal forma que estas

colisiones garantizaran que se compartía el mismo momento angular, depositándose la materia en un disco en torno al protoSol. Esta concentración de materia en un disco habría hecho más probables las colisiones entre partículas y así, incluso aunque el gas remanente salió fuera del sistema solar cuando el Sol empezó a calentarse (probablemente se perdió así tanto gas como el que aún permanece ahora en el Sol), los supergranos que se habían formado en la nube serían todavía capaces de llevar a cabo interacciones recíprocas.

El proceso de acreción continuó poco más o menos del mismo modo hasta que se formaron unos objetos que tenían el tamaño actual de los asteroides, es decir, unos bloques rocosos que podían tener un diámetro de alrededor de un kilómetro. Sin embargo, para entonces la gravedad había empezado ya a ser importante y atraía a estos bloques de rocas uniéndolos en enjambres donde podían chocar uno con otro, adhiriéndose para formar bloques más grandes. Los de mayor tamaño, que ejercían el tirón gravitacional más fuerte, atraían más materia, incorporándola a su propio bloque, y aumentaban así aún más su masa y su fuerza de atracción gravitatoria, creciendo de esta manera hasta convertirse en planetas. En este punto, es de suponer que el calor generado por los impactos de sucesivas oleadas de rocas que golpeaban el protoplaneta lo había fundido completamente, haciendo que el hierro y otros metales existentes se instalaran en el núcleo y produciendo, cuando el planeta se enfrió, el tipo de estructura en capas que vemos hoy en la Tierra.

Con todo esto es fácil de explicar por qué existen actualmente dos tipos de planetas y una gran cantidad de detritos cósmicos en el sistema solar. Cerca del Sol, el calor que emitía esta joven estrella habría liberado gases ligeros y materiales que se podían vaporizar fácilmente. Es probable que los granos que podían sobrevivir con este calor fueran ricos en materiales que no se vaporizaban con facilidad, tales como hierro y silicatos. Estos materiales formaron los bloques que constituyeron los planetas interiores, que son pequeños y rocosos, y tienen tan sólo una modesta capa de atmósfera.

A mayor distancia del joven Sol, los granos a partir de los cuales se formaron los planetas seguramente mantuvieron un recubrimiento de agua helada, metano congelado y amoníaco sólido (todas ellas sustancias que, como sabemos por estudios espectroscópicos, existen en las nubes interestelares). Además, los gases más ligeros, hidrógeno y helio, que salieron de la parte más interna del sistema solar, estarían disponibles para ser atraídos por la gravedad de cualquier planeta que se formara en esas zonas más frías. Así, los planetas exteriores están constituidos casi en su totalidad por gases y no tienen más que unos núcleos de rocas relativamente pequeños, aunque podían haber crecido inicialmente por efecto de la atracción gravitatoria de bloques rocosos que se formaron del mismo modo que los planetas interiores.

Una característica importante de este proceso de formación de los planetas es que debe haber unas grandes cantidades de escombros cósmicos que quedaron sobrantes después de formarse dichos planetas. Mucha de esta basura permanece aún hoy en día en el cinturón de asteroides situado entre las órbitas de Marte y Júpiter. Allí no ha podido agruparse formando un planeta debido al efecto de la atracción gravitatoria del propio Júpiter. En zonas aún más alejadas del Sol hacía —y hace— suficiente frío como para que los bloques helados de detritos permanezcan en forma de cometas.

Cuando el sistema solar era joven, durante unos mil millones de años posteriores a la formación de los planetas, había detritos cósmicos por todas partes, y era frecuente que sus impactos dejaran cicatrices en las superficies de los jóvenes planetas. Aunque los grandes choques con los planetas interiores acabaron hace unos cuatro mil millones de años, existen trozos perdidos de basura cósmica que todavía chocan con los planetas de vez en cuando, como se demostró dramáticamente cuando los fragmentos del cometa Shoemaker-Levy 9 impactaron en Júpiter en 1994. Como ya hemos dicho, es casi seguro que algunos impactos semejantes que se produjeron sobre la Tierra hayan afectado al curso de la evolución y, en particular, que uno de esos impactos contribuyera a la extinción de los dinosaurios hace 65 millones de años. Incluso ahora, el sistema solar está lejos de gozar de una total tranquilidad; sin embargo, se ha asentado en una situación muy estable, y hay que tener en cuenta que el Sol y todos los miembros de su familia poseen sus propias características individuales distintivas.

El Sol es indiscutiblemente el miembro dominante del sistema solar. Contiene el 99'86 por 100 de la masa de dicho sistema y mantiene a todos los demás objetos girando en órbitas a su alrededor, capturados en su dominio gravitatorio, tal como explicó Newton. Un solo planeta, Júpiter, contiene dos tercios de la masa restante, lo cual, hablando en sentido estricto, sitúa incluso a la Tierra en la categoría de «otros trozos y fragmentos». Pero es natural, sin embargo, desde nuestra perspectiva humana, describir incluso el Sol, hasta donde sea posible, por comparación con nuestro propio planeta.

En números redondos, la masa del Sol es 330.000 veces la masa de la Tierra y tiene un diámetro de 1'4 millones de kilómetros, es decir, 109 veces el diámetro de la Tierra. Esto significa que podrían situarse 109 Tierras una al lado de la otra a lo largo de un único diámetro que cruce el Sol; [42] pero, dado que el volumen de la esfera es proporcional al cubo de su radio (o al cubo de su diámetro), el volumen del Sol es igual a 109 al cubo multiplicado por el volumen de la Tierra, es decir, bastante más de un millón de veces el volumen de nuestro planeta. Haciendo un promedio de todo el Sol, su densidad es sólo un tercio de la densidad media de la Tierra, con lo cual el Sol tiene aproximadamente 1'4 veces la densidad del agua. Sin embargo, en el centro del Sol, donde las reacciones nucleares están generando energía, la densidad es doce

veces la densidad del plomo sólido, siendo la temperatura de unos 15 millones de grados centígrados.

Conocemos, por una serie de estudios combinados, las condiciones existentes en las profundidades del interior del Sol. En primer lugar, los físicos han podido averiguar (utilizando elementos muy sencillos y básicos de la física) cómo debe ser su interior para radiar tanta energía y para mantener su configuración frente a la gravedad. Para ratificar este planteamiento, existen experimentos, realizados aquí en la Tierra mediante aceleradores de partículas y combinados con la teoría cuántica, que nos dicen cómo se genera la energía en el interior del Sol, siendo en todo esto un ingrediente absolutamente decisivo el modo en que los efectos cuánticos permiten que los protones se fusionen entre sí a pesar de que las temperaturas son de «sólo» 15 millones de grados. Los modelos estelares (veremos más sobre ellos en el próximo capítulo), combinados con la física de las partículas, concretan sólo una gama limitada de posibilidades en cuanto a propiedades tales como la densidad y la temperatura a distintas profundidades en el interior del Sol. Y lo mejor de todo es que en décadas recientes los astrónomos han conseguido determinar pequeños ripples (ondulaciones) que aparecen en la superficie del Sol y que serían el equivalente solar de los terremotos. Utilizando la sismología solar, han examinado el interior del Sol de un modo muy parecido al que emplean los geofísicos cuando utilizan ondas sísmicas para examinar el interior de la Tierra, y han averiguado que la estructura interna concuerda realmente con los modelos estelares. Por lo tanto, lo que estamos contando aquí son auténticos hechos científicos, comprobados mediante experimentos.

Bajo las condiciones extremas que se dan en el interior del Sol, los electrones son expulsados de sus átomos, con lo que quedan simples núcleos de hidrógeno y helio (protones y partículas alfa, respectivamente). Debido a que los núcleos son mucho más pequeños que los átomos, el núcleo central del Sol actúa como un gas perfecto, en el que hay núcleos que rebotan unos contra otros produciendo colisiones de alta energía. Esta zona central del Sol, que sólo representa el 1'5 por 100 de su volumen, contiene la mitad de su masa.

La energía producida en el núcleo aparece en su mayor parte en forma de fotones de alta energía, inicialmente rayos gamma, los cuales, bajo esas condiciones de densidad extrema, sólo pueden viajar una corta distancia antes de encontrarse e interactuar con una partícula cargada (electrón, protón o partícula alfa). Aunque estas interacciones degradan poco a poco los rayos gamma convirtiéndolos en unos rayos X ligeramente menos energéticos, cada fotón salta por todas partes en el interior del Sol, rebotando de una partícula cargada a otra, como una bola en una máquina de billar loca. Si bien los fotones viajan a la velocidad de la luz, siguen un sendero tan tortuoso y zigzagueante que tardan, por término medio, diez millones de años en encontrar el camino de salida hacia las capas superficiales. Si viajaran

directamente hacia el exterior a la velocidad de la luz, el viaje desde el centro del Sol hasta la superficie sólo duraría 2'5 segundos; pero la luz recorre en realidad diez millones de años luz en zigzag para llevar a cabo el viaje de 2'5 segundos luz.

Una consecuencia de esto es que en la actualidad el aspecto global del Sol es realmente el resultado ya terminado de todo lo que ha estado sucediendo en su interior durante aproximadamente los últimos diez millones de años. La observación directa de la superficie del Sol nos dice más sobre lo que sucedía en su núcleo hace diez millones de años, que sobre lo que estaba sucediendo en su núcleo ayer. Sin embargo, las ondas sonoras que producen ripples en la superficie del Sol viajan directamente a través de él en minutos, por lo que la helio-sismología nos dice cómo es hoy la estructura interior del Sol, lo cual hace que esta ciencia sea doblemente válida como comprobación de la exactitud de los modelos estelares.

La zona de radiación llega hasta un millón de kilómetros de distancia desde el centro del Sol, lo que representa alrededor del 85 por 100 del camino hasta la superficie. En esa zona la temperatura ha bajado a 500.000° C y la densidad es sólo el uno por ciento de la densidad del agua. En estas condiciones algunos núcleos son capaces de captar electrones y los propios fotones han sido degradados a longitudes de onda más larga y niveles de energía más bajos, como resultado de sus reiteradas colisiones con partículas cargadas en su viaje de diez millones de años. El resultado en general es que el gas parcialmente ionizado que se encuentra en esta zona puede absorber energía de la radiación. La materia caliente producida de este modo se eleva por convección, transportando energía al exterior durante el último 15 por 100 de su trayecto desde el centro del Sol hasta su superficie (unos 150.000 kilómetros, es decir, la mitad de la distancia que hay entre la Tierra y la Luna). La brillante superficie visible del Sol, que tiene una temperatura de aproximadamente 5.500° C, es el lugar donde los átomos liberan energía en forma de fotones de luz, que posteriormente tardan sólo 8'3 minutos en cruzar los restantes 150 millones de kilómetros hasta la Tierra.

Toda la luz que vemos partir del Sol procede de una capa que se encuentra a tan sólo 500 kilómetros de profundidad, siendo dicha capa la zona más exterior y el 0'1 por 100 del Sol. Pero la influencia de este astro se extiende más lejos aún en el espacio, a través de una especie de atmósfera solar, que se conoce como cromosfera y termina en una zona llamada la corona, siendo esta última una región que se extiende millones de kilómetros en el espacio y produce una corriente de materiales tenues que sopla desde el Sol y es conocida como viento solar.

El planeta más cercano al Sol es Mercurio, que describe su órbita a una distancia de 0'39 unidades astronómicas y tarda 87'97 días de los nuestros en realizar un viaje completo alrededor del Sol. Debido a que este planeta rota sobre su eje una vez cada 58'64 días terrestres, resulta que en Mercurio dos años duran tres días. Aunque

Mercurio es visible a simple vista (ya que brilla porque refleja la luz solar) y fue uno de los planetas que conocieron los antiguos, es muy difícil verlo a la luz del Sol, por lo que la mayor parte de la información que poseemos acerca de su superficie procede de la sonda espacial Mariner 10, que pasó tres veces por este planeta en 1974 y 1975. El Mariner 10 envió imágenes que mostraban una superficie abundantemente sembrada de cráteres y que recordaba mucho a la superficie de la Luna. Esto resultó una sorpresa total para los astrónomos, pero actualmente es algo que ya se ha incorporado al modelo estándar del modo en que los planetas se crearon, con un fuerte bombardeo de asteroides que continuó durante cientos de millones de años después de que los planetas logaran más o menos su tamaño actual.

Básicamente no hay atmósfera en Mercurio; las temperaturas cambian desde 190°C cuando el Sol da de lleno, hasta -180°C en el lado donde es de noche. Tiene un diámetro de 4.880 km, con lo cual Mercurio presenta un tamaño intermedio entre los de la Tierra y la Luna, y tiene una masa que es cerca del 5 por 100 de la masa de la Tierra.

Como ya hemos mencionado en el capítulo anterior, Venus, el segundo planeta contando desde el Sol, está muy cerca de ser un gemelo físico de la Tierra. Su masa es el 82 por 100 de la masa de nuestro planeta y su diámetro mide en el ecuador 12.104 km, que es una longitud parecida a los 12.756 km del radio de la Tierra. Dado que la superficie de Venus está totalmente cubierta de nubes, ni los mejores telescopios pueden apreciar desde la Tierra ninguna de las características de su superficie. La similitud entre las superficies de Venus y la Tierra ha dado pie a los escritores de ciencia ficción (e incluso a algunos científicos) para especular diciendo que esas nubes podrían ocultar una húmeda jungla llena de vida. Sin embargo, como ya hemos visto anteriormente, el galopante efecto invernadero ha hecho que Venus sea en realidad un abrasador desierto, con unas temperaturas en su superficie que superan los 500°C , una presión atmosférica que llega a ser 90 veces la presión que tenemos en la superficie de la Tierra y una lluvia altamente ácida que cae de sus nubes, arrastradas por fuertes vientos. Hemos tenido amplia información relativa a estas condiciones gracias a la serie de sondas espaciales rusas Venera, algunas de las cuales descendieron atravesando la atmósfera de Venus a finales de los sesenta y en los setenta; dos de estas sondas llegaron a la superficie y realizaron un breve envío de datos antes de ser destruidas por las duras condiciones del planeta. La atmósfera de Venus contiene aproximadamente un 98 por 100 de dióxido de carbono y un 2 por 100 de nitrógeno, así como vestigios de otros gases.

A pesar de su completa cobertura de nubes, la superficie de Venus se ha cartografiado con mucho detalle utilizando radares desde varios satélites, incluidos algunos de los satélites Venera. El último y mejor de estos reconocimientos fue llevado a cabo por la sonda Magellan de la NASA, que se puso en órbita alre dedor

de Venus en agosto de 1990 y cartografió casi la totalidad de su superficie. Aunque ésta aparece también llena de abundantes cráteres, presenta mucha más variación que la superficie de Mercurio, existiendo una gran llanura que cubre casi los dos tercios del planeta (como un fondo marino seco) y una masa de tierra que se alza sobre esta llanura como si fuera un continente de los que hay en la Tierra. Las montañas más altas de Venus llegan a elevarse ocho kilómetros sobre la superficie y, además de los numerosos cráteres producidos por impactos, existen también volcanes, sistemas de valles y torrentes de lava. Sin embargo, teniendo en cuenta que el área de Venus es mayor, hay muchos menos cráteres en la superficie de Venus que en las superficies de Mercurio o de la Luna. Comparando la «densidad» de cráteres en los tres cuerpos celestes, los astrónomos deducen que toda la superficie de Venus se renovó hace unos 600 millones de años en algún cataclismo que hizo que la lava fluyera hacia el exterior desde la parte interior del planeta. Puede ser que la superficie de Venus se haya renovado así varias veces durante los más o menos cuatro mil millones de años transcurridos desde que se formó el planeta, y está claro que su actividad tectónica es distinta de la actividad que produce la deriva continental en la Tierra.

Existe otra peculiaridad relativa a este planeta. Realiza su rotación muy despacio y además en sentido opuesto a la rotación del Sol y de la mayoría de los planetas, siendo la duración de una rotación completa igual a 243 días de los nuestros. Quizá sea esto el resultado de un importante impacto ocasionado por un asteroide al final del bombardeo masivo que tuvo lugar cuando el sistema solar era joven. Sea cual sea la causa, significa que, dado que Venus tarda 225 días terrestres en describir una órbita completa alrededor del Sol, la combinación de su giro retrógrado y su movimiento orbital hace que, desde cualquier punto del ecuador de Venus, el tiempo transcurrido desde el mediodía de un día hasta el mediodía del día siguiente sea equivalente a 116'8 días de los nuestros y que en cada año de Venus haya algo menos de dos días de los suyos.

Debido a su posición especial como nuestro propio lugar de residencia, hemos dicho ya muchas cosas sobre el planeta Tierra, el tercer planeta de nuestro sistema solar en orden creciente de distancias al Sol. Pero la Tierra —o más bien el sistema Tierra-Luna— es también única, en otro sentido, entre los planetas del sistema solar. Nuestra Luna tiene aproximadamente un cuarto del tamaño de la Tierra, siendo por ello mucho más grande (en proporción al planeta en torno al cual orbita) que cualquier otra de las lunas del sistema solar (excepto Caronte, pero, en todo caso, no consideramos que Plutón sea un auténtico planeta). Tiene un diámetro de 3.476 km y describe una órbita en torno a la Tierra manteniendo una distancia media de 384.400 km (una distancia que la luz tarda 1'3 segundos en cubrir). Dentro del sistema solar la Luna se parece a Mercurio más que a ningún otro cuerpo celeste, siendo el tamaño de este planeta un 38 por 100 del tamaño de la Tierra y teniendo en cuenta que Mercurio

es un planeta de pleno derecho. Desde la perspectiva de cualquiera que no sea un habitante de la Tierra, tiene mucho más sentido considerar el sistema Tierra-Luna como un planeta doble, que se debió de formar de un modo especial cuando el sistema solar era aún joven. Este carácter único que posee el sistema Tierra-Luna es todavía más chocante si tenemos en cuenta que ni Mercurio ni Venus tienen ninguna luna y que el cuarto planeta interior, Marte, tiene dos pequeñas lunas, que son en realidad asteroides capturados, y no primordialmente compañeros de viaje del planeta rojo.

Así pues, ¿cómo se formó la Luna? Mediante simulaciones por ordenador y el análisis de rocas traídas del satélite por los astronautas del Apolo, los astrónomos han desarrollado un riguroso modelo que resume todo lo que sucedió hasta que se formó un único planeta doble. La Luna, al parecer, fue arrancada de la Tierra en las últimas etapas de formación de este planeta por el impacto de un objeto que tuvo que ser al menos tan grande como Marte. Pero este proceso no fue como arrancar un trozo de roca sólida de un bloque de granito. Los astrónomos denominan a este acontecimiento el Big Splash («gran salpicadura»), una expresión que da idea de lo que fue.

Según este modelo, el impacto de un objeto del tamaño de Marte contra la joven Tierra generó una cantidad de calor suficiente para fundir la totalidad de la superficie del planeta hasta una profundidad de aproximadamente mil kilómetros. Dicho objeto quedó totalmente deshecho por efecto del impacto, se fundió y se fusionó en el océano de roca líquida que se había formado. Si este bloque de materia que llegó del espacio tenía un pesado núcleo metálico, es de suponer que este metal fundido se hundiría a través de las capas de roca fundida para fusionarse con el núcleo de hierro de la Tierra; pero los trozos de roca de este superasteroide seguramente se fusionaron haciéndose indistinguibles de las rocas fundidas pertenecientes a la corteza terrestre y se supone que parte de la mezcla resultante fue arrojada fuera y quedó en órbita alrededor de la Tierra por efecto del Big Splash.

Estos detritos calientes formaron probablemente un anillo alrededor de la Tierra, del cual se evaporó toda el agua y otras sustancias volátiles, que se perdieron en el espacio. Sin embargo, cuando estos materiales se enfriaron, lo lógico es que quedaran soldados, uniéndose así para formar la Luna, exactamente del mismo modo que la materia que inicialmente estaba en un anillo alrededor del Sol se unió en bloques para formar los planetas. Los efectos colaterales de la colisión dejaron para siempre su impronta en el planeta; probablemente, este impacto es la razón por la cual la Tierra rota tan rápido en torno a su eje, de tal forma que el día dura exactamente 24 horas, y también el hecho de sufrir un impacto descentrado aporta una explicación posible para entender la inclinación de la Tierra, que es la causa, tanto del ciclo anual de las estaciones, como de los ciclos de glaciación de Milankovitch. Cuatro mil millones de

años después de que se produjera, aún estamos directamente afectados por el impacto. Además, si los ciclos de Milankovitch fueran realmente un factor decisivo en la evolución del hombre mono para llegar al ser humano, deberíamos nuestra propia existencia a este impacto.

No existe ningún modo de demostrar la exactitud de este modelo, salvo que tomemos el planeta Marte y lo lancemos contra Venus para ver lo que pasa. Sin embargo, existen poderosas pruebas circunstanciales a favor de dicho modelo —así como la convincente prueba que ofrecen las simulaciones por ordenador de un impacto de este tipo— que parten del hecho de que la Luna no parece contener ni rastro de agua o de otras sustancias volátiles y es el único de los cinco grandes objetos de roca del sistema solar interno (Mercurio, Venus, la Tierra, la Luna y Marte) que no tiene un núcleo de hierro. El hielo encontrado recientemente en la superficie de la Luna no estaba allí primitivamente, sino que fueron cometas los que lo depositaron allí.

Marte es el primer planeta que nos encontramos después del sistema Tierra-Luna cuando nos alejamos del Sol. Aunque la distancia de la Tierra a Marte varía según la posición en que se encuentren los dos planetas en sus respectivas órbitas alrededor del Sol, en su posición más cercana Marte está a 56 millones de kilómetros de la Tierra y ya ha sido visitado por toda una flota de naves espaciales que han enviado una gran cantidad de información que se suma a la información sobre este planeta obtenida en estudios realizados mediante telescopios desde la Tierra.[43]

Marte describe una órbita completa alrededor del Sol cada 686'98 días terrestres a una distancia que oscila entre 1'38 y 1'67 UA. El día en Marte dura 24 horas 37 minutos y 23 segundos, casi lo mismo que el día en la Tierra, pero sólo tiene una delgada atmósfera (siendo la presión un 0'7 por 100 de la presión de la atmósfera de la Tierra al nivel del mar), formada en su mayor parte por dióxido de carbono y las temperaturas en su superficie varían desde -140° C hasta (en raras ocasiones) justo por encima del punto de congelación, pero en la mayoría de sus zonas las temperaturas nunca se elevan por encima del punto de congelación. El diámetro de Marte es de 6.795 km (aproximadamente la mitad del radio de la Tierra); su masa es sólo un poco más de un décimo de la masa de la Tierra. Como el resto de los planetas interiores, su superficie está marcada por gran cantidad de cráteres. En muchos lugares la superficie aparece también excavada por las cuencas de ríos desaparecidos hace tiempo, presentando una serie de cañones y valles, pero hace al menos varios cientos de millones de años que Marte perdió su atmósfera original y el planeta se congeló cuando el efecto invernadero se debilitó. Cualquier resto de agua que se encuentre aún en Marte está en forma de hielo perpetuo, encerrado bajo la superficie. En ocasiones, debido al calor generado por el impacto de algún asteroide, se podría liberar temporalmente parte de su agua congelada y sería posible que ésta fluyera en

algunos lugares concretos.

Marte se parece mucho a una imitación del planeta Tierra y está claro que habría tenido océanos de agua líquida actualmente si fuera tan grande como nuestro planeta. Del mismo modo, Venus habría sido también muy parecido a la Tierra solamente con que hubiera comenzado su existencia un poco más alejado del Sol (o el Sol hubiera estado un poco más frío), ya que en tal caso se habría formado agua líquida y ésta habría disuelto parte de su espesa atmósfera de dióxido de carbono. Se da la intrigante consecuencia de que la «zona con vida» en torno a una estrella como el Sol, la región en la que se podrían formar planetas habitables, se extiende desde un poco más allá de Venus hasta un poco antes de llegar a la órbita de Marte. En vez de considerar que el sistema solar ha sido afortunado por tener un planeta como la Tierra en su familia, podría resultar, cuando consigamos explorar otros sistemas planetarios, que es un infortunio no tener dos planetas como la Tierra.

Sin embargo, en nuestro sistema, Marte es significativamente distinto de la Tierra y, desde luego, también distinto de Venus. Hay tantos cráteres en Marte que resulta claro que lo que vemos hoy es con mucho su superficie primitiva, y que, a diferencia de Venus, dicha superficie no ha sido renovada por ningún cataclismo volcánico durante los últimos cuatro mil millones de años. Incluso así, Marte ha estado geológicamente activo durante todo el tiempo de su existencia y puede estarlo aún. El mayor volcán de Marte es el Olympus Mons, que se eleva hasta los 23 km sobre las planicies que lo rodean y tiene un diámetro de 500 km. En comparación, el mayor volcán de la Tierra, el Mauna Loa en Hawai, se eleva sólo 9 km sobre el fondo marino y su diámetro no mide más de 200 km.

Como ya hemos dicho, a Marte lo acompañan dos diminutas «lunas». Se trata de dos bultos con forma de patata. Fobos, la mayor de las dos, tiene un tamaño de unos 28 km por 20 km y describe una órbita en torno a Marte cada 0'3 días a una distancia de 9.380 km; la otra, Deimos, viene a ser de unos 16 km por 12 km y describe una órbita en torno a Marte cada 1'3 días a una distancia de 23.460 km. Ambas están llenas de cráteres (el mayor cráter de Fobos tiene un diámetro de 10 km, en un satélite cuya anchura máxima es de 28 km) y no pueden ser sino fragmentos de detritos del cercano cinturón de asteroides capturados por la gravedad de Marte. Esto nos lleva directamente a tratar sobre otra característica del sistema solar: el cinturón de asteroides.

Los asteroides son bloques pequeños y rocosos, formados por detritos, mucho más pequeños que los planetas. Muchos de ellos describen órbitas en una franja situada entre las órbitas de Marte y Júpiter. Debido a que tienen un tamaño tan pequeño, los asteroides reflejan muy poca luz solar, por lo que el cinturón de asteroides no fue descubierto hasta el siglo XIX; ahora bien, se han identificado y catalogado más de 2.500 de estos objetos en órbitas que se encuentran a distancias de

entre 2'2 UA y 3'3 UA del Sol. Otros cientos de asteroides se han visto intermitentemente, pero nunca durante el tiempo preciso para calcular sus órbitas con exactitud; en el cinturón de asteroides puede haber alrededor de medio millón de objetos lo suficientemente grandes como para ser fotografiados por el telescopio de 200 pulgadas situado en Monte Palomar. Pero, entre todos estos bloques de escombros, sólo unos 250 tienen más de 100 km de anchura máxima. La mayoría de los asteroides conocidos miden alrededor de un kilómetro en su parte más ancha. Sumando la masa total de todos los objetos que se encuentran en el cinturón de asteroides sólo se obtiene algo más de un cuarto de la masa de la Luna; tardan entre tres y seis años en describir una órbita alrededor del Sol, dependiendo de su distancia exacta a esta estrella.

La razón por la cual todos estos detritos no llegaron a adherirse entre sí para formar un único objeto de gran tamaño es que están afectados constantemente por la atracción gravitatoria de Júpiter, que arrastró los primitivos fragmentos hasta situarlos en una confusión de órbitas en las que era lo más probable que chocasen violentamente entre sí, en vez de aproximarse para adherirse unos a otros. En la zona que ahora ocupa el cinturón de asteroides, los modelos sugieren que podía haber habido materia prima suficiente como para hacer cuatro planetas rocosos del tamaño de la Tierra (o un planeta que tuviera cuatro veces la masa de la Tierra). Algunos de los meteoritos que alcanzan la Tierra, y que, según se cree, proceden del cinturón de asteroides, contienen materiales metálicos, lo que demuestra que alguna vez estuvieron en el núcleo de cuerpos celestes mayores.

Uno de los modelos actuales favoritos (por desgracia no hay forma de probarlo) afirma que la materia original que hay en el cinturón de asteroides constituía en realidad ocho objetos del tamaño de Marte, pero la mayoría de ellos se deshicieron, en aproximadamente cien millones de años, como resultado de las colisiones causadas por la influencia perturbadora de Júpiter. Uno de los superasteroides fue afectado de tal manera que cayó en el sistema solar interno, colisionando con la Tierra y dando lugar a la formación de la Luna en el Big Splash. Uno de los objetos está aún ahí: el propio Marte. Los demás se fragmentaron al sufrir colisiones y se dispersaron de tal modo que la mayor parte de los detritos producidos fue arrojada fuera del cinturón de asteroides, unas veces para describir órbitas que tenían un final abrasador en el Sol, y otras veces para salir del sistema solar.

Es evidente la importancia que tiene todo esto. Aunque los asteroides que quedan describen órbitas relativamente estables (porque todos los que estaban en órbitas inestables hace tiempo que fueron expulsados), Júpiter todavía funciona, ejerciendo su atracción gravitatoria sobre ellos. Aún se producen colisiones entre asteroides y hay fragmentos originados en estas colisiones que a veces se ven arrastrados a órbitas que les hacen cruzar la órbita de la Tierra. La lluvia de detritos cósmicos que marcó

las superficies de los planetas interiores cuando el sistema solar era joven no se ha detenido todavía, sino que tan sólo se ha reducido, pero hay que tener en cuenta que el impacto de un bloque de detritos de unos diez kilómetros con la Tierra fue suficiente para poner fin a la era de los dinosaurios. Por fortuna, tales acontecimientos son raros actualmente; sin embargo, a muy largo plazo, si se desea que la civilización sobreviva en la Tierra, habrá que encontrar un modo de protegerse de los impactos cósmicos.

El paso para ir más allá del cinturón de asteroides hasta el siguiente miembro de la familia del Sol ofrece el mayor contraste que se puede observar en el sistema solar. De los guijarros planetarios de más o menos un kilómetro de amplitud, se pasa al mayor planeta del sistema solar, Júpiter, cuyo diámetro es de unos 143.000 km (una décima parte del diámetro del Sol) y cuya masa es 318 veces la masa de la Tierra (un 0'1 por 100 de la masa del Sol). Es esta masa, que para ser un planeta es enorme, la que hace a Júpiter tan importante como fuerza gravitatoria que actúa sobre el resto del sistema solar, y debe su masa al hecho de que se formó lo suficientemente lejos del Sol como para retener grandes cantidades de los gases originales: en un 90 por 100 Júpiter está constituido por hidrógeno, en un 10 por 100 por helio, y sólo tiene algunos vestigios de otros gases tales como el metano y el amoníaco. Recorre su órbita a una distancia media del Sol de 5'2 UA y tarda 11'86 años en describir una órbita completa.

Todo lo que se refiere a Júpiter se expresa mediante superlativos. Aparece rayado por bandas de colores producidas por la circulación en su atmósfera, el equivalente a las estelas de los aviones que vuelan por la Tierra a gran altitud. Sin embargo, una única tormenta en Júpiter, la Gran Mancha Roja, ha estado produciendo remolinos por todo el planeta durante los últimos trescientos años y es lo bastante grande como para tragar a toda la Tierra.

Pero la característica más impresionante de Júpiter es que tiene su propia familia de lunas, como un sistema solar en miniatura. Hay cuatro grandes satélites jovianos (Ío, Europa, Ganimedes y Calisto), que fueron descubiertos por Galileo a principios del siglo XVII. Este descubrimiento de satélites en órbita alrededor de un planeta de la familia contribuyó a desarrollar la idea de que la Tierra recorre una órbita en torno al Sol de una manera similar, lo cual hizo añicos el dogma que proponía que la Tierra ocupaba el centro del universo. Júpiter posee también al menos una docena de satélites menores (su número sigue creciendo a medida que se descubren más), muchos de los cuales son asteroides capturados.

Los cuatro satélites galileanos son objetos que en sí mismos resultan extraordinarios. Las sondas espaciales han mostrado que Ío es una bola brillante de un material sulfuroso rojo y anaranjado que brota de los volcanes activos que puntean toda la superficie de este satélite. El calor que da potencia a estos volcanes proviene

de las fuerzas de atracción de Júpiter, las cuales ejercen presión sobre el interior de este satélite. Como contraste, Europa aparece enteramente cubierto de hielo, que presenta una maraña de oscuras grietas. Como también se calienta por la presión que produce la atracción de Júpiter, Europa posee un mar de agua líquida bajo su corteza de hielo y podría ofrecer un lugar para albergar vida. Sin embargo, Calisto está demasiado alejado de Júpiter, por lo que no puede calentarse mucho por efecto de la presión que produce la atracción del planeta, y en consecuencia está cubierto por una espesa capa de hielo sólido, fuertemente marcada por cráteres, como no se ha visto en ningún otro objeto del sistema solar. Finalmente, Ganimedes, el mayor satélite del sistema solar, tiene una superficie de hielo que está parcialmente marcada por cráteres, como Calisto, pero en otras partes es lisa, cubierta en apariencia por hielo fresco que se debe de haber extendido por la superficie en algún cataclismo relativamente reciente.

Después de Júpiter, el resto de los planetas del sistema solar son algo así como un anticlímax. Saturno es un Júpiter más pequeño, con un diámetro que sólo es 9'4 veces el de la Tierra y una masa que es 95 veces la de nuestro planeta. Describe una órbita alrededor del Sol cada 29'46 años a una distancia que varía entre 9 UA y 10 UA. Pero el sistema de Saturno se distingue de otros sobre todo por dos características: los famosos anillos, que en las fotografías astronómicas le confieren el aspecto más hermoso en comparación con el resto de los planetas exteriores, y el poseer una familia de satélites que incluye a Titán, la luna más intrigante del sistema solar. Titán tiene un diámetro de 5.150 km, lo que le hace ser un poco más pequeño que Ganimedes (diámetro 5.262 km). Sin embargo, el rasgo especial de Titán es que posee una espesa atmósfera, compuesta principalmente por nitrógeno, pero también rica en metano. La presión en la superficie de Titán es l'6 veces la presión atmosférica al nivel del mar en la Tierra, y la temperatura que hay allí es de -180° C. Puede haber lagos u océanos de metano líquido en su superficie y es posible que caiga lluvia de metano de las nubes. Titán es como una segunda versión de la Tierra en sus primeros tiempos, pero helada y de menor tamaño. Cuando el Sol se acerque al final de su vida y se infle hasta convertirse en un gigante rojo, los planetas interiores, incluida la Tierra, se freirán como churros, pero Titán podría calentarse lo suficiente como para que la vida tuviera una segunda oportunidad dentro del sistema solar (o podría ser una tercera oportunidad, si el satélite Europa, de Júpiter, es realmente otro lugar adecuado para albergar vida). Pero todo esto queda muy lejos en el futuro. Como hecho más inmediato, podemos mencionar que la nave espacial Cassini, enviada a finales del año 1997, dejará caer un módulo en la atmósfera de Titán a principios del siglo XXI, con la esperanza de que el estudio de su atmósfera helada pueda aportar claves para saber cómo era la atmósfera de la Tierra hace mucho tiempo, y quizá también algunas ideas sobre el origen de la vida.

Más allá de Saturno hay otros dos gigantes de gas. Urano describe una órbita alrededor del Sol cada 84'01 años, a una distancia que oscila entre 18'31 UA y 20'07 UA. Sin embargo, tiene una masa que sólo es 14'5 veces la masa de la Tierra y un diámetro que es cuatro veces el de nuestro planeta. Neptuno, el planeta real más distante dentro de la familia del Sol, describe una órbita en torno a éste cada 164'79 años, a una distancia de 30'06 UA. Su masa es un poco mayor que la de Urano, 17'2 veces la masa de la Tierra, pero su diámetro es ligeramente menor (alrededor del 0'1 por 100) que el de Urano. El último de los planetas gigantes está sólo unas treinta veces más lejos del Sol que nosotros y, en comparación con las distancias entre las estrellas, toda la familia de planetas se apiña muy cerca en torno a su centro, el fuego solar. Utilizando como medida el tiempo que tarda la luz en viajar por el espacio, la distancia de Neptuno al Sol no llega a 4'2 horas luz; sin embargo, la distancia desde el Sol a las más próximas de las demás estrellas es más de 4'2 años luz (casi nueve mil veces más). No obstante, hay algo entre Neptuno y la estrella más próxima, y ese algo sigue siendo justamente parte de la familia del Sol.

Más allá de la órbita de Júpiter, el calor del Sol es tan débil que puede haber aún detritos que quedaron como restos de la formación del sistema solar. Éstos existirían en forma de bloques de hielo, y no precisamente como asteroides de rocas (aunque estos bloques de hielo sí que pueden contener rocas empotradas en ellos). Estos icebergs cósmicos contienen mucho más que agua helada. Dióxido de carbono sólido, metano y amoníaco son sustancias que están presentes en todos ellos. A veces, sufren perturbaciones a causa de las cuales acaban describiendo órbitas que los llevan cerca del Sol, hasta que dan la vuelta a este astro y vuelven a las profundidades del espacio. Cuando se acercan al Sol y se calientan, parte del hielo se evapora y forma una larga cola, que brilla por el efecto de la luz solar que refleja; el bloque de hielo se convierte así en un corneta. Pero cuando sale del sistema solar interior, la cola se desvanece a medida que se hace más débil el calor que reciben del Sol y el núcleo del corneta se convierte de nuevo en un inerte bloque de hielo.

Los estudios de las órbitas de los cometas muestran que su origen último es una nube esférica de icebergs que se encuentra muy alejada en las profundidades del espacio, rodeando el Sol, literalmente a medio camino entre éste y la estrella más próxima, a una distancia de aproximadamente 100.000 UA, o un par de años luz. Un típico núcleo de cometa de los que se encuentran en esta nube describe una órbita alrededor del Sol a unos cien metros por segundo y puede haber estado allí durante miles de millones de años, desde que se formó el sistema solar. Sin embargo, ocasionalmente puede llegar alguna influencia del exterior, por ejemplo la atracción gravitatoria de una estrella que pasa, que hace que algunos de estos bloques de hielo caigan en la parte interior del sistema solar, acelerándose todo el tiempo, pero pasando millones de años en el viaje hasta que dan una vuelta alrededor del Sol y

vuelven a tomar el camino hacia el espacio. Algunos de estos visitantes del sistema solar interior son capturados por la gravedad de Júpiter y quedan describiendo órbitas más cortas, pero alargadas, y, como el cometa Halley, hacen varias pasadas junto al Sol cada pocas décadas o cada pocos siglos, antes de evaporarse por completo, dejando algún que otro bloque de roca, granos de arena y polvo diseminados por sus órbitas.

Cuando la Tierra atraviesa uno de estos rastros de cometas, el cielo queda iluminado por los brillantes haces de luz de los meteoros, siendo la causa de cada uno de estos haces un trocito de polvo cósmico, no mayor que un grano de arena, que arde en la atmósfera. Sin embargo, si un cometa choca contra la Tierra, puede hacer tanto daño como un asteroide de roca, o probablemente más, ya que es muy posible que viaje a mucha mayor velocidad, dado que ha caído desde una distancia tan grande, y traerá por consiguiente más energía cinética que un asteroide que tenga la misma masa. Desde luego, lo más probable es que el impacto que ocasionó la desaparición de los dinosaurios fuera un cometa y no un asteroide.

En gran medida hay que agradecer a Júpiter que los cometas se quedaran donde se encuentran hoy en día, es decir, en los márgenes exteriores del sistema solar. Cuando los planetas se formaron, tuvo que haber enormes cantidades de estos icebergs cósmicos en la zona situada entre Júpiter y Neptuno, pero, al igual que sucedió con los detritos que fueron origen del cinturón de asteroides, bajo la influencia gravitatoria de Júpiter y otros planetas gigantes, debieron de recibir perturbaciones que los situaron en unas órbitas, bien para llevarlos a su desaparición en el Sol, o bien para sacarlos hacia los márgenes del espacio interestelar. Aún existe un cinturón de este tipo de detritos de hielo más allá de la órbita de Neptuno. Los más modernos telescopios han conseguido ya empezar a detectarlos, pero, según todos los indicios obtenidos hasta ahora, parece que puede haber unos mil millones de cometas en este Cinturón de Kuiper, que se ensancha hacia afuera (en su sección transversal, como una gigantesca trompeta), llegando a conectarse con una nube esférica de cometas, la nube Oort, en zonas remotas del espacio, y se calcula que podría haber en total alrededor de diez billones de cometas. El total de la masa de todos los cometas llegaría a ser sólo unas pocas veces la masa de la Tierra, pero los icebergs cósmicos más grandes que se han detectado hasta este momento entre las órbitas de los planetas exteriores tienen un diámetro de un par de cientos de kilómetros (alrededor de una décima parte del tamaño de Plutón; con esto, Plutón parece ser solamente un ejemplo grande de este tipo de iceberg).

Cualquier objeto de esta área del sistema solar sufrirá finalmente las perturbaciones que ocasiona la influencia gravitatoria de Júpiter, bien trasladándose a nuestra parte del sistema solar, o bien escapando hacia las profundidades del espacio (y no puede haber estado en esa órbita durante más que unos pocos millones de años,

siendo presumible que haya entrado en ella viniendo de una zona más exterior del sistema solar). Si se traslada hacia el interior, es probable que al calentarse se fragmente y los gases que salgan de él hirviendo rompan el hielo, produciéndose así un enjambre de cometas que viajan juntos por el sistema solar interior. Aunque ninguno de ellos fuera a chocar con la Tierra, la cantidad de polvo fino que se esparciría por todo el sistema solar interior cuando los cometas se evaporaran podría ser suficiente para reducir la cantidad de calor que llega a nuestro planeta desde el Sol. Algunos astrónomos creen que ésta podría ser la causa de algunas glaciaciones que se producirían en la Tierra, algo así como un invierno cósmico. Si sucediera esto, el curso de la evolución y la civilización podría resultar influenciado directamente por las estrellas, aunque no en el modo en que piensan los astrólogos. Una estrella que pasara cerca podría hacer que se liberara un supercometa de la nube Oort, que se dirigiría hacia el interior para cruzar dando vueltas por el sistema solar y terminar deshaciéndose al llegar cerca del Sol. Esto sería la causa de una glaciación en la Tierra millones de años más tarde.

Cualquiera de estas sugerencias no puede ser sino una especulación, y es posible que nunca lleguemos a saber qué exactitud pueden tener estas descripciones. Sin embargo, sirven para recordarnos que el Sol y su familia no existen de una manera aislada, sino que son parte de un sistema estelar mucho mayor, la galaxia conocida como Vía Láctea. Para situar el sistema solar en perspectiva, necesitamos examinar las vidas de las propias estrellas. Dado que nos iremos a unas condiciones de temperatura y presión mucho más extremas, esto significa que la física nos va a simplificar mucho el camino. Ya verá el lector que es preciso acostumbrarse a las escalas de tiempo y distancia que aparecen en esta parte de la historia.

La vida de las estrellas

Una de las afirmaciones más fantásticas que se pueden oír en el mundo de las ciencias en su conjunto es que tenemos buenos modelos para explicar lo que sucede en el interior de las estrellas —cómo nacen, cómo viven y cómo mueren—. Las estrellas aparecen ante nosotros simplemente como puntos de luz en el cielo, a distancias tan enormes que la luz tarda cientos y hasta miles de años en viajar desde ellas hasta nosotros (incluso el hecho de que conozcamos con exactitud las distancias a las estrellas es un logro que habría llenado de asombro a los astrónomos hace menos de doscientos años). Aunque los átomos y las partículas subatómicas son, en cierto sentido, tan remotos como las estrellas dentro del marco de la experiencia cotidiana, porque no podemos verlos a simple vista, sin embargo, existen al menos en los laboratorios aquí en la Tierra y se pueden estudiar directamente comprobando las diferentes teorías al respecto mediante experimentos; no hay modo, sin embargo, de comprobar nuestras teorías sobre las estrellas realizando experimentos en objetos tan distantes.

Incluso si pudiéramos realizar estas comprobaciones, no podríamos vivir lo suficiente como para ver el resultado de tales experimentos, no porque fuera peligroso, sino porque nos haríamos muy viejos antes de lograrlo. Los modelos estelares nos dicen que en muchos casos las estrellas viven miles de millones de años. El intervalo de vida del ser humano suele ser de menos de cien años y el tiempo que lleva existiendo toda nuestra civilización es más breve que cien intervalos de vida humana, es decir, menos de 10.000 años. ¿Cómo podría nadie afirmar con seriedad que sabe de qué manera nació una estrella, hace miles de millones de años, y de qué manera terminará su vida, dentro de miles de millones de años?

De hecho, los modelos que describen la estructura y la evolución de las estrellas[44] figuran entre los grandes triunfos de la ciencia moderna y se han comprobado con gran precisión. Incluyen una combinación de observaciones de las propias estrellas (utilizando técnicas espectroscópicas), simulaciones por ordenador (basadas en las leyes físicas conocidas) de lo que sucede en el interior de las estrellas e incluso experimentos reales, realizados aquí en la Tierra, para comprobar algunos aspectos de los modelos, especialmente la velocidad a la que se desarrollan ciertas reacciones nucleares en condiciones que se corresponden con las que se supone que existen, según proponen otros modelos, en el centro de la estrella en cuestión. La totalidad del conjunto de los modelos es científicamente coherente. Pero el ingrediente clave, aquel sin el cual la astronomía no sería más que una especie de partida de mus, es la espectroscopia.

Como ya mencionamos en el capítulo 2, la espectroscopia permite a los astrofísicos identificar los diferentes elementos químicos presentes en la superficie de

una estrella, analizando la luz que llega de dicha estrella. Esto permitió a Norman Lockyer incluso identificar un elemento que hasta entonces era desconocido, el helio, a partir de su sintonía espectral en la luz del Sol. La distribución de la luz de una estrella entre fotones de diferentes energías (diferentes colores) nos dice también la temperatura de su superficie, utilizando la famosa curva de la energía del cuerpo negro que fue tan importante porque puso a Max Planck en el camino hacia la física cuántica.

Si conocemos la temperatura de la superficie de una estrella y sabemos cuál es su masa, entonces las leyes básicas de la física (las relaciones entre propiedades tales como la temperatura y la presión de un gas) y los modelos desarrollados por ordenador nos dicen qué temperatura debe tener la estrella en su interior, cuáles han de ser la presión y la densidad, etc. (ya explicaremos en breve cómo se pueden conocer las masas de las estrellas). Además, si conocemos la temperatura y la presión que hay en el núcleo de una estrella, y también sabemos de qué está hecha la estrella, entonces sabremos qué reacciones nucleares se están produciendo en el interior de la estrella. Luego, unos experimentos realizados en laboratorios aquí en la Tierra nos dirán cuánta energía deben estar generando dichas reacciones. Esto se puede comparar con la cantidad de energía que una estrella emite en realidad al espacio y los modelos se pueden confrontar para hacer que la teoría y la observación se compaginen mejor entre sí. Todo coincide maravillosamente, aplicando una gran cantidad de conceptos y leyes de la física, desde cosas tan corrientes que ningún físico suele pensar mucho en ellas (por ejemplo, la relación entre presión y temperatura), hasta los más sofisticados experimentos de la física nuclear. El éxito de la astrofísica representa en muchos aspectos la culminación del método científico, confirmando que todas las pequeñas parcelas de la física que se descubrieron de forma separada, cuando se aplican en conjunto, funcionan realmente del modo que sugieren nuestros modelos. Por supuesto, la manera más fácil de ver esto es hacerlo desde la estrella más próxima: el Sol.

El nacimiento de la astrofísica se puede datar con bastante precisión, situándolo en la conferencia dada en 1920 por el pionero de esta disciplina científica Arthur Eddington en la reunión anual de la British Association for the Advancement of Science (Asociación Británica para el Avance de la Ciencia), que tuvo lugar aquel año en el mes de agosto en Cardiff. Hasta el comienzo del siglo XX, había sido para los astrónomos un rompecabezas el explicar de dónde obtiene el Sol su energía. Las pruebas que ofrecía la geología, y también la teoría de Darwin sobre la evolución por selección natural, implicaban una historia muy larga de la Tierra y, en consecuencia, del Sol. Sin embargo, ninguna de las formas conocidas de la energía química (tales como, por ejemplo, la combustión del carbón) eran capaces de explicar cómo podía el Sol haber estado caliente durante el tiempo suficiente para que la geología y la

evolución llevaran a cabo sus tareas.

El descubrimiento de la radiactividad, que incluía la cuestión de la energía liberada desde los núcleos de los átomos, empezó a cambiar el panorama. Luego, Albert Einstein cuantificó la enormidad de esta transformación con su famosa fórmula que nos dice que la propia materia se puede convertir en energía. Sin embargo, al principio la gente fue reacia a aceptar todas las implicaciones de estas nuevas teorías. Una cosa era utilizar el concepto de reacciones subatómicas (es decir, nucleares) para explicar por qué un trozo de radio resulta caliente al tacto, pero otra cosa era dar el salto al acto de fe necesario para admitir que el mismo tipo de procedimiento podría explicar la enorme emisión de energía que realiza el Sol, la cual es equivalente (según la fórmula de Einstein) a convertir cinco millones de toneladas de materia en pura energía cada segundo.

Durante algún tiempo, la gente persistió en aceptar una teoría decimonónica, según la cual una estrella como el Sol podría mantenerse caliente durante el tiempo suficiente como para explicar los acontecimientos acaecidos en la Tierra, sencillamente contrayéndose muy despacio bajo su propio peso y convirtiendo mientras tanto su energía potencial gravitatoria en calor. Sin embargo, incluso con esto, sólo podría mantener su resplandor durante unas pocas decenas de millones de años. Fue Eddington quien finalmente acabó con esta teoría y situó a la astrofísica en su línea correcta, diciendo en Cardiff a sus colegas que:

Sólo la inercia de la tradición mantiene viva la hipótesis de la contracción; o más bien, no viva, sino como un cadáver sin enterrar. Pero, si nos decidimos a dar sepultura a su cadáver, reconozcamos francamente la posición en que nos quedamos. Una estrella está utilizando algún enorme depósito de energía de un modo que nosotros desconocemos. Este depósito difícilmente puede ser otro que la energía subatómica que, como se sabe, existe abundantemente en toda la materia; a veces soñamos que algún día el ser humano aprenderá a liberarla y a utilizarla para su propio provecho. El almacén es casi inagotable; sólo hace falta que se pueda explotar. En el Sol hay suficiente energía subatómica como para mantener su producción de calor durante quince mil millones de años...

Transcurridos unos pocos años, Eddington demostró cómo, incluso mediante la aplicación de las leyes más sencillas de la física, se podía descubrir profundos secretos relativos a la naturaleza de las estrellas. Si una estrella está en equilibrio —es decir, no se está contrayendo ni expandiendo— la fuerza de la gravedad que se ejerce hacia el interior debe ser compensada por la presión hacia el exterior de la materia caliente contenida en el interior de la estrella. Las leyes de la física nos dicen que una

esfera de gas que contenga una cierta cantidad de materia y que se mantenga por su presión interna, debe tener un determinado tamaño, una determinada temperatura central y ha de radiar una determinada cantidad de energía, venga de donde venga esta energía. La masa del Sol se ha conocido con bastante exactitud a partir del modo en que su atracción gravitatoria influye en las órbitas de los planetas. En el caso de muchas otras estrellas, sus masas se pueden calcular de una manera similar porque gran cantidad de ellas tienen a otras estrellas como compañeras, formando sistemas binarios. Si las estrellas binarias están lo bastante cerca de nosotros como para que podamos medir las propiedades orbitales de las dos estrellas cuando se mueven una en torno a la otra, entonces sus masas se pueden también calcular (la espectroscopia interviene aquí de nuevo, porque la variación de la frecuencia por el efecto Doppler en las líneas espectrales de las estrellas nos dice a qué velocidad se están moviendo en sus órbitas).

La cantidad de energía que produce una estrella puede también calcularse, midiendo su brillo aparente en el cielo y teniendo en cuenta a qué distancia está. Aquí es donde aparece la necesidad de medir distancias.

No fue sino al final de la década de 1830 cuando se midieron las primeras distancias exactas a las estrellas más cercanas. La técnica utilizada fue la triangulación, exactamente la misma técnica que utilizan los topógrafos en la Tierra, cuando miran un objeto distante desde dos puntos de una larga línea base cuidadosamente medida y se utilizan los ángulos que forma esta línea con las dos visuales para calcular la distancia al objeto a partir de nuestro conocimiento de las propiedades de los triángulos. La línea base utilizada en este tipo de topografía astronómica es el diámetro de la órbita que describe la Tierra en torno al Sol. [45]

Las observaciones se realizan con un intervalo de seis meses, cuando la Tierra se encuentra en lados opuestos de la órbita y unas pocas estrellas cercanas están lo bastante próximas como para que parezca que, durante este intervalo, se mueven ligeramente con respecto al trasfondo de otras estrellas distantes. Este efecto se conoce como paralaje y es exactamente igual al que se produce cuando extendemos un brazo horizontalmente, levantamos un dedo y lo miramos, cerrando primero un ojo y luego el otro: parece que en ese intervalo de tiempo el dedo se mueve en relación con los objetos que están mucho más distantes. Dado que los ángulos se miden en segundos de arco, de esta técnica resultan distancias medidas en términos del cociente entre segundos de arco y paralaje, expresándose el resultado en parsecs. Un parsec es algo más de 3'25 años luz, es decir, bastante más que 206.000 veces la distancia de la Tierra al Sol.

Las primeras estrellas cuyas distancias se midieron, hace algo más de siglo y medio, fueron 61 del Cisne (a una distancia de 3'4 parsecs o algo más de 11 años luz), de la Lira (8'3 parsecs o 27 años luz) y del Centauro, de la que se sabe actualmente

que es un sistema múltiple de estrellas en el que está incluida la estrella más cercana al Sol, situada a una distancia de 1'3 parsecs (4'3 años luz). El sistema estelar más próximo al Sol se halla a una distancia que es siete mil veces la distancia media de Plutón al Sol. Durante la pasada década, se midieron con una precisión exquisita las distancias a muchas estrellas algo menos cercanas, y se hizo utilizando esta técnica, por medio del satélite Hipparcos, que recorre una órbita libre del oscurecimiento causado por la atmósfera terrestre. Aprovechando el modo en que algunas estrellas todavía más lejanas se mueven por el cielo, se han utilizado a lo largo de los años otras técnicas geométricas, con las que se han calculado las distancias a diversas estrellas y clusters (enjambres) de estrellas. Todas estas técnicas en conjunto han demostrado ser adecuadas para proporcionarnos información sobre las distancias a clusters de estrellas que ofrecen los más importantes descubrimientos de la astronomía: las estrellas conocidas como cefeidas variables.

Lo extraordinario de las cefeidas es que su brillo varía de una forma muy regular y que para cada cefeida en particular el período de la variación (el tiempo que tarda en brillar, palidecer y brillar de nuevo) depende del brillo medio real, o luminosidad, de la estrella. Así, si medimos el período de una cefeida y su brillo aparente en el cielo, podemos calcular a qué distancia está, comparando el brillo aparente con el brillo intrínseco determinado por la relación período-luminosidad, del mismo modo que se puede determinar la distancia a una bombilla de 100 vatios midiendo su brillo aparente. Cuánto más débil sea éste, más lejos ha de estar.

Esta técnica sencillamente funciona; se llegó a conocer las distancias a, literalmente, un puñado de cefeidas antes de que se dispusiera de los datos proporcionados por el satélite Hipparcos, y fueron estas pocas estrellas las que se utilizaron para calibrar la relación período-luminosidad. Una vez que se calibró dicha relación, fue posible aplicarla para obtener las distancias a las cefeidas que aparecían formando clusters por toda la Vía Láctea, e incluso en nuestras galaxias vecinas más cercanas (se verá más sobre este tema en el próximo capítulo). Así es como sabemos que la galaxia llamada Vía Láctea es un disco aplanado de aproximadamente 4 kiloparsecs (trece mil años luz) de espesor en su parte central, unos 30 kiloparsecs (98.000 años luz) de longitud máxima, empotrado en un halo esférico de clusters globulares de estrellas que se extiende por un volumen de espacio cuyo diámetro es de 150 kiloparsecs (490.000 años luz). El Sol es una estrella muy corriente, situada en una zona corriente de esta galaxia, a unos 9 kiloparsecs del centro del disco, aproximadamente dos tercios del recorrido que hay desde el centro de la Vía Láctea hasta el borde de este disco, que tiene un radio de 15 kiloparsecs.

Hemos llegado a comprender las vidas de las estrellas de este disco porque podemos (al menos en algunos casos) medir sus distancias y su brillo, determinar sus masas y estudiar su composición utilizando la espectroscopia. También disponemos

de otro instrumento de valor incalculable: la estadística. Hay tantas estrellas para estudiar que podemos verlas en todas las fases de sus ciclos vitales y podemos comparar las propiedades observadas en ellas con las predicciones de nuestros modelos por ordenador, el equivalente astronómico de comprobar modelos mediante experimentos. Exactamente del mismo modo, se puede averiguar cómo es el ciclo vital de un árbol estudiando un único bosque durante unas cuantas semanas e investigando los árboles que están en las distintas etapas de su vida, sin tener que estar observando todo el tiempo para ver cómo crece un plantón hasta convertirse en un árbol maduro que a su vez produce plantones.

Las leyes de la física que han de aplicarse para describir lo que sucede bajo las condiciones existentes en el interior de una estrella son particularmente sencillas (mucho más sencillas que las que describen lo que sucede en el interior de un árbol), porque, como ya hemos explicado, los electrones son lanzados fuera de sus átomos bajo estas condiciones de altas temperaturas y presiones elevadas, de tal forma que los núcleos individuales de los átomos se comportan como los componentes de un gas perfecto. Cuando una bola de gas contiene masa suficiente para que suceda esto, formando el llamado plasma de partículas cargadas, entonces las partículas de movimiento rápido convierten parte de su energía cinética en radiación, y esta radiación produce a su vez una interacción con otras partículas cargadas, contribuyendo a crear la presión que mantiene el volumen de la bola de gas, contrarrestando la fuerza de su propio peso. Una estrella incandescente y estable se mantiene por el efecto de una combinación de presión del gas y presión de la radiación. Pero si la bola de gas tiene más de una cierta cantidad de masa, las condiciones en su parte central llegan a ser tan extremas que las partículas de movimiento rápido irradian enormes cantidades de radiación electromagnética, una energía que hace que la estrella estalle en pedazos.

Eddington constató que existen tres, y sólo tres, destinos posibles para una bola de gas en el espacio, cuando se colapsa bajo su propio peso. Si es demasiado pequeña para que se forme un plasma en su interior cuando los electrones son arrancados de los átomos, se convertirá en un globo frío que se mantendrá únicamente gracias a la presión del gas (algo así como Saturno). Si es de un tamaño bastante mayor, puede convertirse en una estrella incandescente, que se mantendrá por una combinación de presión del gas y presión de la radiación. Y, si su tamaño sobrepasa unas dimensiones determinadas, brillará brevemente como un globo de gas extremadamente caliente, antes de estallar en pedazos por efecto de la presión de la radiación. A partir de estas ideas tan sencillas de la física, Eddington dedujo entre qué valores oscilan las masas que permiten a las estrellas estables continuar su existencia. Los números exactos que se obtienen de estos cálculos depende de la composición real que se suponga que tiene la estrella en cuestión, porque el número de electrones disponibles depende de

qué átomos (núcleos) estén presentes (un electrón por cada átomo de hidrógeno, ocho por cada núcleo de oxígeno, etc.). Por esta razón, las cifras dadas por Eddington no se corresponden demasiado bien con las versiones modernas de estos cálculos. Sin embargo, los aspectos más amplios siguen siendo válidos, a pesar de que hoy en día, en relación con la composición de las estrellas, tenemos ideas mejores que las que tuvo Eddington en los años veinte.

El astrónomo británico nos pedía que imagináramos una serie de globos de gas de varios tamaños, primero de 10 gramos, luego de 100 gramos, el siguiente de 1.000 gramos, y así sucesivamente. El n ésimo globo tiene una masa igual a 10^n gramos. Resulta que, según las leyes básicas de la física, los únicos globos que pueden tener una existencia estable como estrellas incandescentes son los que ocupan los lugares comprendidos entre 32 y 35 dentro de esta serie. Entonces ¿qué pasa si cotejamos esta predicción de las leyes básicas de la física con las observaciones realizadas en el universo real? El globo número 31 de la sucesión anterior tiene una masa de 1031 gramos, que es alrededor de cinco veces la masa de Júpiter. Desde luego, con bastante seguridad, Júpiter es una bola de gas que se mantiene por la presión del mismo, no una estrella incandescente. El globo número 32 de la sucesión anterior tiene una masa de 1032 gramos, que es aproximadamente una décima parte de la masa del Sol. Por lo tanto, una estrella no puede ponerse incandescente hasta ser varias veces mayor que Júpiter y tener alrededor de una décima parte de la masa del Sol; por cierto que el Sol está en el lado correcto de la línea divisoria.[46] Y, en el otro extremo, las masas reales de las mayores estrellas conocidas no son más que unas cuantas veces 1035 gramos, es decir, alrededor de 100 veces la masa del Sol. El globo número 35 es realmente el mayor globo que puede formar una estrella estable.

El otro factor clave que surgió de la investigación de Eddington es que todas las estrellas, con independencia de su masa y su brillo, deben tener aproximadamente la misma temperatura en sus núcleos. Eddington calculó que esta temperatura ha de ser de unos 40 millones de grados Celsius, porque se permitió aceptar la presencia de una cantidad bastante excesiva de electrones; la versión moderna de este cálculo reduce esta estimación a 15-20 millones de grados. En su gran obra *The Internal Constitution of Stars* (CUP, 1926) Eddington recopila todas estas cuestiones, justo un poco antes de que se llegara a establecer la nueva física cuántica. Utilizando ejemplos de estrellas específicas cuya masa y brillo se conocían, afirmó que aplicando las leyes más sencillas de la física:

Sin entrar en discusiones... tanto si se necesita un aporte [de energía] de 680 ergios por gramo (V Puppis), como si se necesita un aporte de 0'08 ergios por gramo (Krueger 60), la estrella ha de alcanzar los 40.000.000° para conseguirlo. A esta temperatura obtiene un aporte ilimitado.

La razón por la cual todas las estrellas tienen más o menos la misma temperatura interna se debe a la retroalimentación. Si una estrella (cualquier estrella) se contrae un poco, se calentará más en su interior, ya que se libera más energía gravitatoria. Esto hará que se expanda, para restaurar su equilibrio. O bien, supongamos que la estrella en todo caso se expande; este proceso necesitaría extraer energía del núcleo de la estrella, haciendo que se enfriara un poco, de tal modo que la presión caería y la estrella se contraería otra vez. Las estrellas llevan incorporado un termostato que mantiene sus núcleos justo a la temperatura necesaria para que se libere energía subatómica (nuclear).

Dicho sea de paso, esta retroalimentación ayuda a explicar una pregunta con la que generaciones de profesores de astronomía han puesto a prueba a sus estudiantes. La pregunta es sobre el efecto que producen las reacciones nucleares en la temperatura del núcleo de una estrella. La respuesta obvia es que dichas reacciones mantienen caliente a la estrella; sin embargo, esta respuesta está equivocada. Sin estas reacciones nucleares que sirven para producir energía, y con ello una presión hacia el exterior para resistir el tirón de la gravedad hacia el interior, la estrella se contraería y así se calentaría aún más. La función de las reacciones nucleares es mantener el núcleo de la estrella fresco, en términos relativos.

Al principio de la década de 1920, los físicos descartaron el conjunto de las teorías de Eddington, calificándolas de absurdas. La razón era que sabían cuánta energía cinética necesitarían dos protones para chocar con la fuerza suficiente y poder así superar su mutua repulsión eléctrica y fusionarse, produciendo helio, y sabían también que esta energía requeriría temperaturas mucho más elevadas que la que Eddington proponía como temperatura del núcleo de una estrella. Es famoso el hecho de que Eddington se defendió argumentando que «el helio que manejamos tuvo que ser producido en algún momento y en algún lugar» y continuó diciendo, irónicamente, «no estamos discutiendo con el crítico que preconiza que las estrellas no están suficientemente calientes para que este proceso pueda realizarse; lo que hacemos es decirle que vaya y encuentre un lugar más caliente». Esto se suele interpretar como el modo que tenía Eddington de decir a sus críticos que se fueran al infierno. Pero, incluso aunque el libro en el que aparecen estas palabras se estaba ya imprimiendo, la física estaba experimentando una transformación debida a la revolución cuántica.

En un par de años, la aplicación que hizo George Gamow de la incertidumbre cuántica para describir la fusión de los núcleos de los átomos dejó explicado ya como podían los protones realmente fusionarse entre sí a las temperaturas existentes en los núcleos de las estrellas. Fue un triunfo para la astrofísica y gracias a él quedó establecida esta rama de la ciencia, que se convirtió en una de las áreas de investigación más interesantes durante los cuarenta años siguientes. No entraremos en

todos los detalles de esta investigación; sin embargo, esperamos que el ejemplo del poder de la física que hemos puesto al explicar lo que sucede en el interior de algo tan sencillo como una estrella convencerá al lector de que los astrofísicos saben realmente de qué están hablando cuando describen las vidas de las estrellas. Entonces, examinemos con detalle este ciclo vital, basando nuestra explicación en lo mejor de la información que se ha obtenido desde los días de Eddington, así como también de las últimas y mejores simulaciones por ordenador.

En una galaxia como es actualmente la Vía Láctea, la formación de estrellas es un proceso continuo en el que interviene la materia reciclada de las nubes de gas y polvo que están en el espacio. Las jóvenes estrellas calientes se encuentran introducidas dentro de nubes de materia y tienen discos de polvo (en los que se formarán los planetas) que giran alrededor de ellas. De dichas estrellas surgen chorros de materia, que son lanzados hacia el exterior desde las regiones polares por efecto de la presión de la radiación que emite la propia estrella joven. Las nubes a partir de las cuales se forman las estrellas están a una temperatura lo bastante fría como para que haya en ellas moléculas estables, entre las que hay sustancias tales como el monóxido de carbono. Estas nubes se denominan nubes moleculares y se encuentran en el plano de la Vía Láctea a distancias de unos pocos años luz. Sus densidades no son superiores a valores entre 1.000 y 10.000 moléculas por centímetro cúbico. Estas nubes están compuestas casi en su totalidad por hidrógeno (75 por 100) y helio (25 por 100), con sólo una pequeñísima cantidad de elementos más pesados. La formación de una estrella tiene lugar cuando una de estas nubes sufre una presión desde el exterior y empieza a colapsarse, hasta tal punto que su propio peso le hace continuar colapsándose, fragmentándose para formar estrellas separadas, que pueden ser estrellas binarias o sistemas más complicados. Este colapso, y todo el proceso actual de formación de una estrella, parece estar íntimamente ligado a la estructura en espiral de nuestra galaxia (existe otra familia de galaxias, llamadas elípticas, que no tienen esta estructura en espiral, y en las cuales no parece que se esté produciendo actualmente ninguna formación de estrellas).

Las estrellas del disco de la Vía Láctea, como en muchas otras galaxias que tienen forma de disco, componen un diseño en espiral, bastante similar al dibujo que forma la nata líquida al verterla en una taza de café. La causa de este diseño es una onda de compresión que se mueve en torno a la galaxia, bastante parecida a una onda sonora. Una onda sonora, al desplazarse, ejerce presión sobre el aire, pero, después de pasar, deja cada molécula del aire más o menos en el lugar en que estaba. Este proceso de ejercer presión desencadena por sí solo el colapso de algunas nubes de materia interestelar, y las más grandes entre las estrellas que se forman recorren sus respectivos ciclos vitales muy rápidamente (en no más de unos pocos millones de años) y explotan como supernovas (más adelante daremos más detalles sobre estas

estrellas). Debido a que estas estrellas grandes tienen una vida muy corta, no les da tiempo a alejarse del lugar donde han nacido, con lo que estas explosiones se producen cerca de la zona donde produce su efecto la onda galáctica. La onda de choque de estas estrellas que explotan provoca el colapso de más nubes en esa parte de la galaxia. Por lo tanto, aunque la onda de densidad que se mueve por la galaxia es en sí misma invisible, está bordeada por una frontera visible de jóvenes estrellas brillantes y calientes, nacidas como resultado del paso de la onda.

El ejemplo clásico de uno de estos viveros de estrellas, visible desde la Tierra en el cielo nocturno, es la constelación Orión. En esta constelación hay una nube de gas y polvo, conocida como la Nebulosa de Orión, iluminada por la luz de las estrellas jóvenes que están en su interior. En nuestra parte de la Vía Láctea, las estrellas están separadas por distancias de algunos parsecs y existe aproximadamente una estrella por cada cinco parsecs cúbicos de espacio; en el vivero de Orión las estrellas están separadas por una distancia de un décimo de parsec y existen diez mil estrellas por parsec cúbico. Este es un caso típico del modo en que nacen las estrellas, aunque se diseminan cuando cada una de ellas se desplaza en su propia órbita alrededor del centro de la Vía Láctea. Ahora vamos a describir lo que le sucede a una estrella después de nacer.

El principal problema al que se ha de enfrentar una nube que se está colapsando, si va a formar una estrella, es librarse del momento angular. Ésta es la razón por la cual se forman discos de materia en torno a las estrellas jóvenes, y por la que luego se forman planetas en dichos discos. El momento angular también desaparece cuando la materia es lanzada hacia el espacio. A menudo, el núcleo de una nube colapsante se parte en dos componentes, dos protoestrellas que describen órbitas una alrededor de la otra a una distancia respetable; esto también acumula momento angular de un modo que no cesa por el hecho de que las estrellas se colapsen. Y tampoco es en absoluto algo inusual que cada uno de estos dos núcleos se parta a su vez, formando un doble sistema binario.

Una estrella como el Sol empieza a funcionar como un reactor de fusión nuclear cuando la temperatura de su núcleo se eleva a unos quince millones de grados. Luego, mientras en su núcleo haya combustible en forma de hidrógeno que se va a convertir en helio, las reacciones nucleares impiden que la estrella siga colapsándose y que se caliente más en su interior. El tiempo de vida de una estrella en su fase estable de utilizar el hidrógeno como combustible (conocida dicha fase como «secuencia principal») depende de su masa; no obstante, cuanto más grande sea una estrella, menos tiempo pasa en dicha secuencia principal, porque tiene que quemar su combustible más activamente para mantenerse contra su propio peso. Esto lo consigue, no calentándose mucho más en su interior, sino quemando su combustible mucho más rápidamente con el mismo nivel de temperatura. Una estrella que tenga

veinticinco veces la masa del Sol pasa sólo tres millones de años en este estado; el propio Sol será una estrella en la fase de secuencia principal durante un total de unos 10.000 millones de años; y una estrella que tenga la mitad de la masa del Sol puede estar tranquilamente convirtiendo hidrógeno en helio durante 200.000 millones de años. El Sol se encuentra ahora casi exactamente en la mitad de su vida como estrella en fase de secuencia principal.

Cuando se haya convertido en helio todo el hidrógeno que contiene el núcleo del Sol, éste seguirá estando rodeado por una «atmósfera» de hidrógeno que contendrá más o menos la mitad de la masa original del Sol. Cuando se haya agotado su reserva de energía, el núcleo de la estrella empezará a contraerse, generando calor a medida que se libera la energía gravitatoria. Este calor hará que se expanda la parte exterior de la estrella, convirtiéndola en una estrella del tipo gigante rojo, y también hará que se quemé hidrógeno nuclear en la base de la atmósfera, favoreciendo la expansión. En este estadio de la vida de una estrella es mucha la materia que se pierde en el espacio, lanzada por efecto de este proceso. Sin embargo, incluso teniendo en cuenta la cantidad total de materia que el Sol perderá cuando le suceda esto, todavía seguirá hinchándose hasta engullir la órbita de Mercurio, llegando casi hasta la actual órbita de Venus. No obstante, el Sol perderá tanta masa en este proceso, que las órbitas de los planetas se expandirán, ya que el Sol perderá parte de su fuerza gravitatoria, de tal forma que Venus no se verá engullido, sino que simplemente se freirá.

Antes de que transcurra mucho tiempo, mientras esté sucediendo todo esto, la temperatura en el núcleo colapsante de helio de la estrella se elevará a unos 100 millones de grados Celsius y comenzará una nueva fase de combustión nuclear. En este proceso, el helio se convierte en carbono y se libera energía. Así, la estrella se estabiliza de nuevo como una gigante roja. Esta fase de su vida es muy breve: en el caso del Sol durará sólo unos mil millones de años.

Al finalizar esta pausa de descanso, todo el helio estará agotado y el núcleo de carbono empezará a colapsarse, liberando calor suficiente para hacer que otra oleada de hidrógeno procedente del centro entre en combustión, mientras el astro se expande hasta alcanzar la órbita actual de la Tierra. En esta fase de su ciclo vital, el Sol se convierte en una estrella muy inestable, hinchándose su atmósfera y perdiendo materia en el espacio, para luego contraerse de nuevo, repetidamente. Esto contribuye a sembrar las nubes interestelares de elementos tales como el carbono y el nitrógeno. Finalmente, todo lo que queda es el núcleo interno, que se está enfriando, y que también está compuesto sobre todo por carbono. En este momento, el núcleo interno ya no se está colapsando, sino que se mantiene por la presión física de los núcleos y electrones, que se empujan chocando unos con otros.

Por último, se alcanza la densidad límite y el colapso cesa, cuando las fuerzas cuánticas impiden que los electrones sean empujados aumentando aún más la

proximidad entre ellos (recuérdese que dos electrones no pueden ocupar el mismo estado cuántico, y es precisamente este mismo fenómeno de exclusión cuántica que impide que los átomos se colapsen lo que mantiene a las llamadas «estrellas degeneradas», protegiéndolas del efecto de su propio peso). Esto sólo sucede cuando el remanente estelar, generalmente con dos tercios de la masa que tiene actualmente el Sol, se ha contraído hasta tener el tamaño de la Tierra, convirtiéndose en lo que se conoce como enana blanca.[47] Un centímetro cúbico de la materia de una enana blanca tendría una masa aproximada de una tonelada, es decir, tendría un millón de veces la densidad del agua.

Las estrellas que comienzan su vida teniendo bastante más masa que el Sol pueden atravesar más estadios de combustión nuclear, a temperaturas sucesivamente más elevadas, produciendo elementos tales como el oxígeno y el neón a partir del carbono, y diseminándolos después en el interior de las nubes interestelares. Excepto el hidrógeno y el helio, todos los elementos del universo, incluido todo elemento que se halle en la Tierra y en nuestros cuerpos, han sido fabricados en el interior de una estrella. Sin embargo, una estrella que comience a existir con más de unas once veces la masa del Sol se queda aún, incluso después de toda esta pérdida de masa, con una cantidad de materia en su rescoldo residual que supera la masa solar. En este punto, las cosas empiezan a ser todavía más interesantes.

Si una estrella muerta, que ya no se mantiene mediante la combustión nuclear, tiene más de aproximadamente 1'4 veces la masa del Sol, entonces ni siquiera los procesos cuánticos pueden detener un estadio posterior de compresión. Lo que sucede es que los electrones que están errantes entre los núcleos atómicos en una enana blanca se ven obligados a combinarse con los protones de estos núcleos para producir neutrones. Se trata de una desintegración beta inversa, el proceso inverso del que tiene lugar actualmente en la Tierra cuando un neutrón aislado se convierte de manera espontánea en un electrón y un protón (más un antineutrino). Esto reduce la escoria estelar a algo que no es más que una bola de neutrones, que se comprime de la forma más eficiente en que se puede comprimir la materia, dando lugar fundamentalmente a un único núcleo atómico que posee la masa de una estrella como el Sol. Mientras una enana blanca que tenga una masa como la del Sol es del tamaño de la Tierra, una estrella neutrón con una masa que es una vez y media la del Sol sólo mide diez kilómetros de un extremo a otro, es decir, tiene el mismo tamaño que una montaña terrestre grande. El colapso de un núcleo estelar pasando del estado de enana blanca (o incluso un estado menos compacto) al de estrella neutrón liberaría una enorme cantidad de energía gravitatoria, siendo llevada una gran parte de esta energía por los neutrinos, ya que se libera un neutrino cada vez que un protón y un neutrón se fusionan para formar un neutrón. Cada centímetro cúbico de materia de una estrella neutrón pesaría alrededor de mil millones de toneladas.

Tampoco esto es el final de la historia de un colapso estelar. Son los procesos cuánticos los que mantienen a una estrella neutrón frente al empuje hacia el interior que produce la gravedad, pero dichos procesos cuánticos son limitados y la gravedad no. Una bola compacta de materia cuya masa sea tres o más veces la masa del Sol, y que no esté ya sostenida por la energía liberada en la combustión nuclear, no puede mantenerse por sí misma. Ninguna estrella neutrón puede existir si tiene una masa mayor que la que acabamos de mencionar (según la teoría, pero lo que resulta reconfortante para los teóricos es que las observaciones de las estrellas reales todavía no han mostrado la existencia de ninguna estrella neutrón en esas condiciones, del mismo modo que aún tienen que mostrar la existencia de alguna enana blanca que tenga más de 1'4 veces la masa del Sol). Si hubiera un objeto de estas características, la gravedad vencería completamente incluso a las fuerzas cuánticas y la escoria de la estrella se contraería indefinidamente, hasta convertirse en un punto de densidad infinita, algo sumamente extraño. En fases intermedias adquiriría una densidad tan enorme y, en consecuencia, su campo gravitatorio llegaría a ser tan fuerte que nada podría escapar de ella, ni siquiera la luz. Se convertiría, por lo tanto, en un agujero negro. Las estrellas neutrón y los agujeros negros se forman en nuestra galaxia actualmente cuando las estrellas cuya masa de materia es más de once veces la del Sol llegan al final de su existencia. Sucede de este modo.

El problema de utilizar la fusión nuclear con el fin de proporcionar energía para mantener una estrella es que lo que la fusión puede hacer tiene límites. Concretamente, una vez que se han formado núcleos de elementos de la familia del hierro (el propio hierro, el níquel y otros parecidos) se precisa más energía para hacer núcleos todavía más pesados, pero ya no se puede conseguir más energía mediante el proceso de fusión. Se ha de introducir energía para fabricar los elementos realmente pesados, tales como el platino, el mercurio, el oro o el uranio. Esto es así porque los núcleos de hierro-56 constituyen la más estable, energéticamente, de todas las configuraciones posibles para los protones y los neutrones en forma de núcleos atómicos.

Utilizando una analogía que nos resulta familiar, es como si los núcleos de la familia del hierro se colocaran en la parte más baja de un valle, situándose los elementos más ligeros sobre una ladera de dicho valle y los más pesados sobre la otra ladera. Todos los elementos más ligeros «querrían» estar en la parte más baja del valle, en un estado de baja energía. Los más ligeros pueden conseguirlo, en principio, mediante la fusión (siempre que estos núcleos tengan energía térmica suficiente para vencer su repulsión eléctrica mutua); los más pesados pueden conseguirlo mediante la fisión (aunque una gran cantidad de elementos pesados son relativamente estables una vez que se han formado, como si estuvieran en profundas cuevas de la ladera del valle). Sin embargo, la verdad es que los elementos más pesados se forman en primer

lugar en el interior de las estrellas, en procesos mucho más energéticos que cualquiera de los que tienen lugar en el interior de una estrella como el Sol.

Ya hemos mencionado que la combustión nuclear en el interior de las estrellas comienza con la conversión de cuatro protones (por una vía algo indirecta) en una partícula alfa, es decir, un núcleo de helio, formado por dos protones y dos neutrones (a propósito, dos de los protones originales se convierten en neutrones por la desintegración beta inversa). El núcleo de helio es en sí una unidad especialmente estable y otros estadios posteriores de fusión nuclear que se realizan en el interior de las estrellas consisten básicamente en unir partículas alfa.

Sucede que el primer paso de este proceso, aunque parezca obvio, consistente en unir dos partículas alfa para hacer un núcleo de berilio-8 (4 protones y 4 neutrones) no funciona, puesto que el berilio-8 es extremadamente inestable y se rompe al momento en dos partículas alfa. Por este motivo, la «ceniza» producida por la combustión del helio está compuesta de núcleos de carbono, cada uno de los cuales está formado por tres partículas alfa. Los núcleos del carbono-12 se forman mediante unas interacciones relativamente raras en las que participan, al mismo tiempo, tres núcleos de helio. Sin embargo, a partir de este momento, si una estrella tiene la masa suficiente y consigue estar suficientemente caliente en su interior en cada estadio de su colapso, los elementos se forman en gran medida adhiriendo partículas alfa, una cada vez, a los núcleos ya existentes. Ésta es la razón por la cual elementos corrientes tales como el carbono (6 protones y 6 neutrones, siendo 12 nucleones en total) y el oxígeno (16 nucleones) tienen la estructura nuclear que tienen. A veces se produce a continuación la absorción de otro protón (o dos), y quizá también la expulsión de un positrón, convirtiendo uno de estos protones en un neutrón y consiguiendo núcleos tales como el nitrógeno (7 protones y 7 neutrones).

Si una estrella tiene masa suficiente, hacia el final de su vida estará constituida por capas de materia, con elementos del grupo del hierro en su núcleo más interno que está rodeado por una concha rica en elementos tales como el silicio, otra concha rica en núcleos de carbono, oxígeno, neón y magnesio, otra concha más de helio y una atmósfera de hidrógeno y helio. La combustión nuclear puede estar realizándose en todas estas conchas a la vez (pero no en el núcleo interno) en los últimos estadios de actividad de la estrella, aunque no durante mucho tiempo.

Los astrónomos depositan una gran confianza en sus modelos relativos a lo que sucede en el interior de una estrella de gran masa antes de que explote en una gloriosa llamarada final. La razón de esta gran confianza es que en 1987 consiguieron estudiar los estertores de agonía de una estrella de este tipo. Aquel año, se vio explotar como una supernova [48] a una estrella de la cercana Gran Nube de Magallanes, una pequeña galaxia satélite asociada a la Vía Láctea. Los detalles observados en esta explosión encajaban a la perfección con las predicciones sobre el comportamiento de

las supernovas que habían hecho los modelos por ordenador. Y lo que es aún mejor, resultó que la estrella progenitora había sido fotografiada en el curso de exploraciones astronómicas realizadas antes de que explotara, de tal modo que pudimos ver qué aspecto tenía anteriormente.

Esta combinación de teoría y observación nos revela que la estrella que explotó comenzó su vida hace sólo alrededor de once millones de años, con una masa de materia igual a unas dieciocho veces la masa solar (principalmente hidrógeno y helio). Con el fin de mantenerse contrarrestando su propio peso, tenía que producir tanta energía que se veía 40.000 veces más brillante que el Sol y gastó todo su combustible de hidrógeno en sólo 10 millones de años. Entonces, la combustión del helio la mantuvo durante otro millón de años antes de que empezara a verse en serias dificultades. Cuando el núcleo comenzó a contraerse más, el colapso de la estrella se detuvo temporalmente convirtiendo carbono en una mezcla de neón, magnesio y oxígeno, aunque sólo durante unos 12.000 años. Los procesos de fusión que utilizaban helio mantuvieron a esta estrella durante aproximadamente doce años más; luego, la combustión del oxígeno fue un apaño durante sólo cinco años y, como último recurso para realizar procesos de fusión, la combustión del silicio proporcionó calor suficiente como para retrasar el colapso durante más o menos una semana, pero ése fue el final del viaje.

Al no tener más aporte interno de calor, el núcleo interno de la estrella se colapsó súbitamente, contrayéndose hasta convertirse en una bola de sólo unas pocas decenas de kilómetros de diámetro a una velocidad de más o menos un tercio de la velocidad de la luz. Dado que se liberó energía gravitatoria, la temperatura se elevó por encima de los diez mil millones de grados, y los protones y los electrones se vieron obligados a unirse para formar neutrones. Entre tanto, las capas exteriores de la estrella — materia equivalente a varias masas solares—se quedaron sin nada que las mantuviera y empezaron a caer al interior hacia el núcleo. El propio núcleo, que estaba colapsándose, generó una onda de choque en su interior e intentó recuperarse inflándose de repente hacia afuera, como una pelota de golf que es aplastada con fuerza y luego se deja suelta; pero esta onda de choque dio contra la materia que estaba cayendo hacia adentro e intentó empujar hacia afuera toda esa masa, la totalidad de la zona exterior de la estrella. La única razón por la cual la onda de choque consiguió arrastrar la masa de la estrella original es que hubo un gran número de neutrinos, producidos en el núcleo por toda la desintegración progresiva beta inversa, que chocaron contra esa masa y la empujaron en su dirección.

Esto nos da una cierta idea de la densidad de la materia que se encuentra dentro de la onda de choque de una supernova. Los neutrinos son tan reacios a producir interacciones con la materia corriente que cerca de setenta mil millones de ellos producidos por reacciones nucleares en el interior del Sol pasan a través de cada

centímetro cuadrado de la Tierra (y de nosotros mismos) cada segundo, sin que nada les afecte. Para un neutrino, el plomo es tan transparente como el cristal para un fotón. Sin embargo, la densidad que hay dentro de la onda de choque es tan alta (aunque no nos lo podamos imaginar, es cien billones de veces la densidad del agua) que resulta tan impermeable para los neutrinos como una pared de ladrillos para una pelota de tenis de mesa.

El resultado es que las partes exteriores de la estrella reciben un empuje hacia el exterior a una velocidad de unos 10.000 kilómetros por segundo, llevando con ellas enormes cantidades de elementos pesados producidos, no sólo por reacciones de fusión que tienen lugar durante el ciclo vital de la estrella, sino también por reacciones forzadas por estas condiciones extremas «hacia arriba de la ladera» en el valle de hierro, haciendo que elementos más pesados que el hierro se sumen a la mezcla. Resulta bastante asombroso pensar que elementos tales como el carbono y el oxígeno que se encuentran en nuestro propio cuerpo hayan sido fabricados en el interior de unas estrellas normales, no mucho mayores que el Sol; pero es aún más asombroso pensar que elementos tales como el oro, que mucha gente luce en sus dedos, no se hicieron en ninguna estrella corriente, sino en la caldera de una supernova, donde una única estrella puede brillar fugazmente con tanto resplandor como toda una galaxia del tipo de la Vía Láctea.

Cuando las capas exteriores de la estrella fueron arrastradas al espacio y la barrera de materia se volvió más fina, los neutrinos pudieron salir a través de los detritos y continuar su camino a una velocidad cercana a la de la luz. Uno de los aspectos más espectaculares de la historia de la supernova de 1987 es que un total de diecinueve neutrinos procedentes de la explosión se detectaron mediante instrumentos situados aquí en la Tierra, después de haber tardado 160.000 años en su viaje desde la Gran Nube de Magallanes. Para ver esto con su propia perspectiva, los astrofísicos calcularon que en la explosión se produjeron unos 1058 neutrinos. Dado que éstos se dispersaron uniformemente en todas las direcciones, a una distancia de 160.000 años luz «sólo» podía haber alrededor de 300 billones de neutrinos de la supernova pasando a través de uno de los detectores, un tanque que contenía más de 2.000 toneladas de agua enterrado profundamente en Japón, donde justo once de estos neutrinos produjeron una interacción con electrones, de tal manera que dejaron su huella. Otros ocho neutrinos de la supernova fueron «vistos» por otro detector en Estados Unidos.[49] Para situar dentro de una perspectiva humana la reticencia de los neutrinos a la hora de producir interacciones con la materia que habitualmente conocemos, basta decir que unos diez mil millones de neutrinos de la supernova pasaron por cada uno de nosotros, por nuestro propio cuerpo, en febrero de 1987 y no nos dimos cuenta de nada.

Aunque es muy pequeño el número de neutrinos detectados, se corresponden

exactamente con la sensibilidad esperada en los detectores con respecto a los neutrinos de la supernova, y este encaje entre la teoría y la observación constituye uno de los grandes triunfos de la teoría y los modelos de la astrofísica. Y no sólo de la astrofísica, ya que el modelo completo del funcionamiento de una supernova también incluye a la física clásica (en lo que atañe a las fórmulas que describen las variaciones de la presión y la temperatura dentro de una estrella) y a la física nuclear y la física cuántica (para describir las interacciones productoras de la energía que parte la estrella). Se podría decir que casi la totalidad de la física interviene a la hora de hacer un modelo para una supernova, por lo que el éxito de estos modelos representa un triunfo para toda la física.

Ciento sesenta mil años antes de que esos diecinueve neutrinos se detectaran en la Tierra, en la Gran Nube de Magallanes, mientras estas partículas y sus miles de millones de compañeras participaban en la ruptura de las capas exteriores de la estrella, se producían reacciones nucleares frenéticas dentro de la materia comprimida, justo fuera del núcleo interno de la estrella, donde la temperatura se elevaba hasta los doscientos mil millones de grados, forzada a subir por efecto de la energía gravitatoria liberada durante el colapso. Todos estos núcleos, que se habían formado originalmente por uniones de partículas alfa, y que por consiguiente tenían el mismo número de protones y neutrones, se fusionaron para formar enormes cantidades (aproximadamente tanto como la masa del Sol) de níquel-56, que posee 28 protones y 28 neutrones en cada núcleo.

Pero el níquel-56 es inestable y se desintegra progresivamente de forma espontánea, liberando cada núcleo un positrón cuando transforma un protón en un neutrón, y se convierte él mismo en un núcleo de cobalto-56, que tiene 26 protones y 30 neutrones en cada núcleo y es estable. Es la energía liberada en este proceso —o sea, en la desintegración progresiva de elementos radiactivos para producir hierro estable— la que hace que una supernova siga brillando durante semanas después de la explosión inicial en la que se formó el níquel-56. Este elemento inestable se produjo por el aporte de energía gravitatoria efectuado durante el colapso. La luz brillante que deja tras de sí la estrella después de la explosión es parte de esa energía gravitatoria almacenada que está escapando. Alrededor de una décima parte del hierro fabricado de este modo se lanza al espacio interestelar, yéndose una porción de él a formar parte de la materia de la que se hacen nuevas estrellas y nuevos planetas (y, al menos en un planeta, forma parte del acero con que se fabrican navajas y carrocerías de automóviles).

Mientras tanto, el núcleo interno de la estrella se ha colapsado del todo hasta que sólo queda una estrella neutrón; y si ese núcleo interno contiene una cantidad de materia que es aún más de tres veces la masa solar (que no es el caso de la explosión de una estrella tan pequeña como la que vimos convertida en supernova en 1987),

continuará colapsándose hasta llegar al estado último de agujero negro.

Existe otro modo de hacer una supernova —ligeramente menos espectacular, pero también importante en cuanto a sembrar la Vía Láctea con elementos pesados—. Muchas estrellas se encuentran formando sistemas binarios, y a veces una de las estrellas de uno de estos sistemas habrá atravesado su ciclo vital y se habrá convertido en una enana blanca, mientras que la otra estará aún en el estadio de gigante roja. En estas circunstancias, la compacta enana blanca puede atraer, por efecto de la gravedad, parte de la materia de la extensa atmósfera de su compañera, ganando así en masa lentamente; no obstante, debe recordarse que la masa máxima para que una enana blanca sea estable es sólo 1/4 veces la masa del Sol. Si la enana blanca comienza estando justo debajo de este límite, puede atraer masa suficiente para pasar del límite, con lo que provocaría el colapso y pasaría al estadio de estrella neutrón, liberando casi tanta energía gravitatoria como se libera en el tipo de supernova que acabamos de describir.

Incluso si la enana blanca de un sistema binario de este tipo es demasiado pequeña para que esto suceda, el hidrógeno de la atmósfera de su compañera se puede acumular en su superficie hasta que haya suficiente para desencadenar una ráfaga de fusión nuclear, que haría que la estrella brillara deslumbrante durante un breve momento y enviaría otra carga de elementos pesados al espacio. A diferencia de la explosión de una supernova, este tipo de explosión menor —llamado una nova— no modifica el sistema en el que sucede, de modo que el proceso se puede repetir una y otra vez.

Todo este reciclaje de materia enriquece el medio interestelar y suministra materia prima para estrellas como el Sol, planetas como la Tierra, y gente como nosotros. No existiríamos si no hubieran existido generaciones previas de estrellas que estuvieron activas de este modo. La materia prima de la cual se formaron las primeras estrellas (hace al menos diez mil millones de años) contenía únicamente un 75 por 100 de hidrógeno, un 25 por 100 de helio y tan sólo algunos vestigios del siguiente elemento ligero, el litio, producido en el Big Bang, en el cual nació el universo (véase el capítulo 11). El Sol nació hace sólo cinco mil millones de años, de una nube de materia que ya estaba tan enriquecida que casi un átomo de cada mil era de un elemento más pesado que el litio.

Las estrellas más antiguas de nuestra galaxia tienen, por ejemplo, unas cantidades de hierro que son solamente una cienmilésima de la cantidad de hierro que posee el Sol. Sin embargo, todavía quedan posibilidades de mejorar nuestros conocimientos sobre las estrellas, ya que nadie ha encontrado aún una que no contenga elementos más pesados que el litio. Las primeras estrellas pioneras en las que se procesaron elementos pesados por primera vez parecen haber desaparecido de la escena hace mucho tiempo y no podemos saber con seguridad qué fue lo que puso en marcha el

funcionamiento de todo el proceso estelar. La mejor hipótesis es que, cuando el universo era joven, se formaron unas superestrellas de masa muy grande que, en consecuencia, atravesaron sus ciclos vitales muy rápido y sembraron las nubes de las que se formó la Vía Láctea con elementos pesados en cantidad suficiente como para poner en marcha los procesos que hemos descrito en este capítulo.

Esto nos lleva al universo a gran escala, más allá de la Vía Láctea y retrocediendo en el tiempo hacia el nacimiento del universo. Nuestra historia está ya casi completa; sin embargo, el final resulta ser el comienzo, y lo grande y lo pequeño resultan estar inextricablemente ligados, como Ouroberos, la serpiente que se muerde su propia cola.

Lo grande y lo pequeño

Hasta la década de 1920 no comenzó la exploración del universo a gran escala, porque, hasta entonces, nadie sabía que había un universo a gran escala.' La imagen que se aceptaba dentro de la cosmología era que el sistema de la Vía Láctea constituía todo el universo y que, aunque algunas estrellas podían nacer y morir, el patrón general de esta galaxia era eterno e invariable. Algunos habían discrepado de esta idea; así, ya en 1755 Immanuel Kant había sugerido que algunas de las borrosas manchas de luz (conocidas como nebulosas) que se veían en el cielo con ayuda de telescopios podrían ser otros «universos islas» como la Vía Láctea; pero muy pocos tomaron esta idea en serio hasta bien entrado el siglo XX.

El primer paso hacia una comprensión de la relación existente entre la Vía Láctea y el universo a gran escala se dio cuando se cartografió por primera vez la Vía Láctea con cierto detalle, al final de la segunda década del siglo XX. El cartógrafo fue Harlow Shapley, un joven investigador del observatorio de Mount Wilson en California, donde tenía acceso a lo que era entonces el mejor telescopio astronómico del mundo, un nuevo reflector de 60 pulgadas de diámetro. Utilizando este instrumento, Shapley pudo analizar la luz de las cefeidas variables a través de la Vía Láctea y determinar así sus distancias. Las cifras reales a las que llegó resultan ser ligeramente excesivas, porque no tuvo en cuenta el modo en que el polvo que hay en el espacio oscurece la luz que proviene de estrellas distantes; sin embargo, fue Shapley quien mostró, en una serie de estudios publicados en 1918 y 1919, que el Sol y el sistema solar se encuentran lejos del centro de la Vía Láctea y no ocupan un lugar especial ni siquiera en nuestro universo isla.

Debido a que había sobrestimado el tamaño de la Vía Láctea, Shapley pensó que las nubes de Magallanes eran sistemas relativamente pequeños dentro de la Vía Láctea, y no sistemas estelares independientes. Esto le animó a pensar que todas las nebulosas sembradas de estrellas[49] eran sistemas satélites, enormes enjambres de estrellas que se movían en órbitas en torno a la Vía Láctea.

No obstante, otros astrónomos discreparon. A medida que se descubrieron y estudiaron más nebulosas de éstas, usando nuevos telescopios como el de 60 pulgadas, y dado que se observó que muchas de ellas tenían un aspecto aplanado, con forma de disco y una estructura en espiral, ganó fuerza el argumento de que debían ser sistemas como la Vía Láctea, situados a distancias muy grandes. En 1920, todo esto desembocó en un debate formal, organizado en Washington por la Academia Nacional de Ciencias de Estados Unidos, entre Shapley (que argumentaba que las nebulosas eran satélites de la Vía Láctea) y Herber Curtis, del observatorio Lick, quien argumentaba que las nebulosas espirales eran cada una de ellas un gran sistema estelar ampliamente similar a nuestra Vía Láctea. El debate resultó poco convincente,

pero marcó un punto de inflexión en la cosmología, ya que marcó el momento en que llegó a ser respetable la idea de que existen galaxias más allá de la Vía Láctea. En diez años, sería demostrada la certeza de la postura defendida por Curtis, con ayuda del último de los grandes telescopios de Mount Wilson, el reflector de 100 pulgadas.

La prueba llegó con los trabajos realizados por Edwin Hubble, que pensó que el nuevo telescopio era lo bastante potente como para poder captar cefeidas variables en algunas de las nebulosas más cercanas, a las que llamaremos a partir de ahora galaxias. Con el fin de medir cómo se hace más brillante y cómo oscurece una cefeida, se necesita, desde luego, realizar una serie de fotografías durante un período de tiempo. Así pues, para obtener al menos una fotografía de la calidad requerida, Hubble tuvo que pasar horas ajustando el objetivo a las vistas del telescopio mientras la placa fotográfica estaba expuesta. Le llevó más de dos años (1923 y 1924) obtener cincuenta fotografías de calidad aceptable de una mancha borrosa de luz que se veía en el cielo (y que ahora se sabe que era una galaxia irregular); pero el esfuerzo valió la pena. La técnica de las cefeidas mostró que esta mancha borrosa estaba siete veces más lejos de nosotros que la Nube Pequeña de Magallanes y no podía ser un satélite de la Vía Láctea. Casi al mismo tiempo, Hubble fotografió cefeidas de la gran nebulosa situada en la constelación de Andrómeda, obteniendo una distancia de poco menos de un millón de años luz (esta estimación se ha corregido posteriormente, cuando nuestros conocimientos sobre las cefeidas han aumentado, incrementando la cifra hasta más de dos millones de años luz). Se demostró que la «nebulosa» Andrómeda es una galaxia espiral muy similar a nuestra Vía Láctea.

Al principio, no eran patentes todas las asombrosas consecuencias de estos descubrimientos. Aunque quedó claro que la Vía Láctea no es más que una galaxia entre muchas otras, dado que la escala de distancias a las cefeidas se había calibrado de manera incorrecta, todas las demás galaxias parecían estar bastante cerca, y en consecuencia se pensó que eran bastante más pequeñas que lo que los cálculos modernos sugerirían después. Hasta los años cincuenta, pareen este capítulo sólo se habla de nebulosas externas que son sistemas de estrellas situados en el exterior de la Vía Láctea.

cía que la Vía Láctea era un gigante entre las galaxias —posiblemente la mayor de todas ellas— a pesar de que hay muchas otras. Pero las sucesivas mejoras en las mediciones de las distancias han dejado claro que las otras galaxias están más lejos y son más grandes de lo que se creía. Esto ha reducido la importancia que creíamos tener en el universo.

Un estudio en el que participé en 1997, utilizando observaciones de cefeidas en muchas galaxias cercanas, que habían sido realizadas mediante el telescopio espacial Hubble, reveló que, desde luego, nuestra Vía Láctea es una espiral de tamaño mediano, comparada con todas las espirales que se encuentran lo bastante cerca como

para haber medido sus distancias mediante la técnica de las cefeidas. En todo caso, nuestra galaxia es ligeramente menor que una galaxia de tamaño medio del mismo tipo. Se calcula también que, en principio, al menos cien mil millones de galaxias son visibles mediante el telescopio espacial Hubble, aunque sólo se han estudiado con todo detalle unos pocos miles. La Tierra recorre su órbita en torno a una estrella corriente que no es más que una entre los varios cientos de miles de millones de estrellas que hay en la Vía Láctea, y esta galaxia es sólo un tipo medio de espiral entre los cien mil millones, o más, de galaxias que hay en el universo visible. Sin embargo, el descubrimiento más dramático de todos —realizado también por Hubble— es que el universo no es eterno ni inmutable. Al contrario, cambia y evoluciona a medida que pasa el tiempo.

Para un cosmólogo, incluso una galaxia como la Vía Láctea, que contiene varios cientos de miles de millones de estrellas, no es más que una «partícula de prueba» cuyo comportamiento se puede utilizar como guía para llegar a saber cómo está cambiando el universo en conjunto. Hubble fue el primero en darse cuenta de que existen diferentes variedades de galaxias. Las galaxias en forma de disco, como nuestra propia Vía Láctea (aplanada por los efectos de la rotación), constituyen alrededor de un 30 por 100 de la población de galaxias del universo, mientras que el 60 por ciento de todas las galaxias tiene formas elípticas (son, tal como sugiere el nombre, una elipse tridimensional). Algunas de las galaxias elípticas son mucho menores que nuestra galaxia, pero otras son mucho más grandes —las galaxias más grandes del universo conocido son elípticas gigantes que contienen billones de estrellas—. Además de las espirales y las elípticas, el otro 10 por 100 está constituido por galaxias irregulares, es decir, montones de estrellas que no tienen una forma definida. Para los cosmólogos todo esto no es más que una serie de meros detalles; en efecto, el comportamiento de las galaxias que más interesa a los cosmólogos es el modo en que parecen moverse, que se pone de manifiesto mediante el famoso desplazamiento hacia el rojo de las rayas espectrales.

Antes de que Hubble entrara en escena los astrónomos observaron que la luz proveniente de muchas «nebulosas» mostraba un desplazamiento hacia el rojo. Esto se interpretó como un efecto Doppler, lo que significaba que estos objetos se estaban alejando de nosotros rápidamente a través del espacio, pero todo ello causó gran desconcierto a mediados de los años veinte, porque algunos de esos cuerpos se estaban moviendo a unas velocidades que en aquella época parecían extraordinarias: 600 kilómetros por segundo, o más.

La medición de los desplazamientos hacia el rojo resulta fácil, con tal de que la galaxia sea lo bastante brillante como para que su luz se pueda analizar espectroscópicamente. El gran adelanto de Hubble consistió, no en medir los desplazamientos hacia el rojo, sino en medir distancias, para luego aunar los

desplazamientos hacia el rojo y las distancias. Las primeras mediciones de distancias a las galaxias que efectuó Hubble utilizando la técnica de las cefeidas sólo funcionaba para unas pocas galaxias cercanas, incluso con el telescopio de 100 pulgadas. Con el fin de calcular distancias a galaxias más remotas, tuvo que usar otros indicadores, tales como medir el brillo de las estrellas que estaban explotando en dichas galaxias. Es razonable suponer que todas las novas y supernovas alcanzan más o menos el mismo brillo, por lo que éste puede constituir una guía aproximada para el cálculo de las distancias. Sin embargo, este tipo de técnica sólo puede ser aproximada, razón por la cual todas las estimaciones sobre distancias a galaxias situadas más allá de la cercanía inmediata de la Vía Láctea comportaron distintas inexactitudes durante muchos años; en realidad, hasta la llegada del telescopio espacial Hubble al final de la década de los ochenta. No obstante, incluso con estas dificultades, hacia el final de los años veinte Hubble y su colega Milton Humason habían hallado las pruebas para el más profundo descubrimiento astronómico, quizás el más profundo descubrimiento realizado en el marco de la ciencia en general. El desplazamiento hacia el rojo de una galaxia es proporcional a la distancia a la que se encuentra de nosotros. Esto significa que su «velocidad de recesión» es proporcional a su distancia. El universo entero se está expandiendo, por lo que en el pasado ha tenido que encontrarse en un estado más compacto. Desde luego, ha debido tener un comienzo.

El descubrimiento de Hubble fue tan sorprendente (y, si somos sinceros, basado inicialmente en tan pocas mediciones de distancias a las galaxias) que podría haber tardado mucho tiempo en ser aceptado por sus colegas. Después de todo, ¿por qué tendría que estar expandiéndose el universo? Sin embargo, la respuesta a esta pregunta existe ya. En efecto, existe ya una teoría completa que describe el comportamiento del universo en expansión. Dicha teoría es la teoría general de la relatividad, que fue terminada por Albert Einstein en 1916 y se aplicó a la cosmología (lo hizo el propio Einstein) en 1917.

La teoría general de la relatividad es una teoría que habla de espacio y tiempo, que trata el tiempo como la cuarta dimensión de un continuo espacio-tiempo y describe las relaciones entre espacio, tiempo y materia en una serie de fórmulas. El modo de hacerse una idea de esta descripción del universo, sin esforzarse en pasar por las matemáticas, es imaginarnos que suprimimos dos de las cuatro dimensiones y visualizar el espacio-tiempo como si fuera una lámina de goma bidimensional estirada. El efecto de la presencia de un objeto en el espacio se puede representar imaginando una bola de jugar a los bolos que se coloca en la lámina estirada. La bola ocasionará una deformación haciendo que la lámina presente una abolladura. Si hay unos pequeños objetos que se mueven pasando junto al objeto pesado (canicas que ruedan por la lámina), éstos seguirán trayectorias curvas, debido a la abolladura causada por el objeto pesado. Esta curvatura de sus trayectorias es el proceso que

corresponde a la fuerza que denominamos gravedad. En casos extremos, un objeto muy compacto y de gran masa ocasionará un agujero que atravesará la lámina estirada; éste sería un agujero negro, con un campo gravitatorio tan potente que nada, ni siquiera la luz, puede escapar de él.

La teoría de Einstein es uno de los mayores triunfos de la ciencia y ha sido comprobada de muchas maneras diferentes. Explica cómo la luz de estrellas distantes parece estar doblada cuando pasa por el borde del Sol (un efecto que es visible durante un eclipse solar, cuando la deslumbrante luz del propio Sol queda tapada por la Luna) y cómo los agujeros negros, con unas masas que son millones de veces mayores que la masa del Sol, tragan la materia mientras liberan energía gravitatoria, convirtiéndola en una radiación electromagnética que lanzan al universo desde los centros de algunas galaxias, formando los llamados cuántares. En principio, la teoría general describe todo el espacio y todo el tiempo: el universo entero. Naturalmente, Einstein deseaba utilizarla para describir el universo matemáticamente, pero cuando intentó hacerlo, en 1917, se encontró con que las fórmulas no permitían la posibilidad de un universo estático e invariable. Según su magnífica nueva teoría, el espacio-tiempo podía estar expandiéndose, o podía estar contrayéndose, pero no podía permanecer sin variaciones.

Recordemos que todo esto sucedía en una época en que aún se creía que la Vía Láctea era todo el universo y que éste era invariable, del mismo modo que lo es un bosque. Así pues, Einstein, siendo muy consciente de que «si no concuerda con el experimento, es erróneo», añadió un término más a sus fórmulas, simplemente para mantener su modelo del universo. Este término añadido era la constante cosmológica, de la cual dijo más tarde que había sido el mayor patinazo de su carrera; sin embargo, esto parece un poco duro, ya que todo lo que intentaba era hacer que su modelo encajara con la mejor prueba experimental (es decir, procedente de la observación) de que se disponía en aquella época. Poco más de diez años después, con el descubrimiento de que el universo se está expandiendo, la constante cosmológica se pudo descartar. Las fórmulas cosmológicas iniciales de Einstein habían predicho el descubrimiento de la relación entre las distancias y el desplazamiento hacia el rojo de las rayas espectrales, a pesar de que incluso el propio Einstein no había creído que podía ser cierta. La teoría sí que coincidía por fin con la observación.

Lo que es aún mejor, la teoría contribuye a eliminar lo que es obviamente la característica más enigmática de la relación entre las distancias y el desplazamiento hacia el rojo. ¿Por qué tendría que alejarse uniformemente de nosotros todo lo que existe en el universo? Seguramente no vivimos en el centro del mismo.

Desde luego que no. El tipo de ley relativa al desplazamiento hacia el rojo que prevé la teoría general de la relatividad y que Hubble descubrió, según la cual el desplazamiento hacia el rojo es proporcional a la distancia, es el único tipo de ley

(aparte de la relativa a un universo estático) que se cumple del mismo modo independientemente de cuál sea la galaxia en la que nos encontremos. La mejor manera de ver esto consiste en volver a la imagen del espacio-tiempo como una lámina de goma estirada o, según una analogía que se emplea a menudo, como la superficie en expansión de un globo que se está inflando constantemente.

Imaginemos que la superficie del globo está salpicada de puntos pintados sobre ella, de tal modo que cada punto corresponde a una galaxia. A medida que el globo se hace más grande, todos los puntos se van alejando los unos de los otros, no porque la pintura se mueva sobre la superficie del globo, sino porque la superficie del globo (espacio-tiempo) se está estirando y se lleva sobre sí los puntos. Si el globo se expande de tal forma que la distancia entre dos puntos vecinos se multiplica por dos, entonces también se multiplica por dos la distancia entre cualquier par de puntos. Eligiendo cualquier punto del globo como referencia, otro punto que estuviera a 1 cm de distancia, estará ahora a 2 cm (con lo que parecerá que se ha movido 1 cm), otro punto que estuviera a 2 cm de distancia, estará ahora a 4 cm (y parecerá que se ha movido el doble de rápido que el anterior), y así sucesivamente. Cuanto más lejos esté un punto, parecerá que se está moviendo a una velocidad mayor (será mayor el desplazamiento hacia el rojo), sea cual sea el punto desde el que se esté midiendo.

Sin embargo, el desplazamiento hacia el rojo no es un efecto Doppler, esto es, no está causado por el movimiento de las galaxias a través del espacio. Es el propio espacio el que lo causa al estirarse, ya que esto hace que la luz de las galaxias lejanas, en su viaje a través del espacio, se estire a longitudes de onda más largas. Y, de la misma forma que no hay ningún centro en la superficie en expansión del globo, tampoco hay ningún centro en el universo que se expande. Todos los puntos del espacio están en igualdad de condiciones (salvo en el caso de perturbaciones locales del espacio-tiempo causadas por la presencia de estrellas, galaxias, etc.).

La teoría general de la relatividad también nos dice que, si nos imaginamos esta expansión yendo hacia atrás en el tiempo, la conclusión es que todo el universo visible ha tenido que emerger de un punto matemático —un objeto singular— en un momento determinado en el pasado. Ningún físico cree que la teoría general de la relatividad se pueda llevar tan lejos, porque es seguro que los efectos cuánticos tienen su importancia, junto a la de ese objeto singular. Sin embargo, está claro que el universo, por lo que sabemos, emergió de un estado muy denso, algo muy parecido a un objeto singular, en un momento determinado en el pasado. Las grandes incógnitas son: ¿cuándo sucedió esto? y ¿cómo sucedió?

La incógnita relativa al «cuándo» ha quedado aclarada gracias a los últimos estudios sobre la velocidad a la que se está expandiendo el universo, es decir, los últimos cálculos para calibrar la relación entre la distancia y el desplazamiento hacia el rojo. Como en la época de Hubble, la cuestión clave es medir las distancias a las

galaxias, ya que lo relativo al desplazamiento hacia el rojo es fácil. Una vez que se tienen las distancias y los desplazamientos hacia el rojo, se puede calibrar la ley que relaciona a ambos y combinar esto con las fórmulas de Einstein para calcular cuándo empezó la expansión. Cuanto mayor sea la velocidad a la que se expande el universo, menos tiempo habrá transcurrido desde el Big Bang para que las galaxias se encuentren tan alejadas como se hallan actualmente. La velocidad de la expansión se mide en función de un número llamado constante de Hubble, o H . Este número se determina a partir de esas mediciones, al día de hoy, de desplazamientos hacia el rojo y distancias a las galaxias, y cuanto menor sea el valor de H , mayor será la edad del universo.

Existen en la actualidad muchas técnicas diferentes para medir distancias en el universo y, felizmente, todas parecen converger hacia la misma respuesta. Una de las técnicas más sencillas es la que yo utilicé, junto con Simon Goodwin y Martin Hendry, derivada de la técnica que habíamos utilizado para medir el tamaño relativo de nuestra propia galaxia. Dado que esta técnica nos dice el tamaño medio de una galaxia espiral como la nuestra, es relativamente fácil tomar un par de miles de galaxias espirales que tengan desplazamientos hacia el rojo conocidos (pero que estén demasiado lejos para que sus distancias se puedan medir directamente mediante la técnica de las cefeidas) y deducir qué valor de la relación entre distancia y desplazamiento hacia el rojo sería el apropiado para garantizar que el tamaño medio de todas estas galaxias se corresponda con el tamaño medio de las espirales cercanas.

Para muchos valores de esta relación, las galaxias muy distantes parecerán estar muy lejos y, por consiguiente, parecerán muy grandes puesto que se ven tan grandes desde la Tierra; para otra serie de valores, parecerán estar muy cerca de nosotros y, por consiguiente, se podrá pensar que son muy pequeñas; sin embargo, para un valor de la relación parecerán estar justo a las distancias correctas para que su tamaño medio se corresponda con el tamaño de la Vía Láctea y sus vecinas. Esta calibración de la «ley de Hubble» —que es así como se conoce— nos dice que el universo tiene unos trece mil millones de años. Por la comodidad que supone trabajar con números redondos, muchos cosmólogos utilizan una edad de quince mil millones de años, ya que nadie es capaz de discutir en profundidad este asunto. Lo que importa es la clara evidencia de que el universo tuvo un comienzo hace más o menos esa cantidad de años; no hace mil millones de años, ni tampoco hace cien mil millones de años, sino hace un número de años que se sitúa en torno a los quince mil millones.

Es importante recalcar que esto es ciencia de buena calidad. Tenemos modelos basados en la teoría general de la relatividad de Einstein, y también contamos con observaciones: no sólo el modo en que la luz procedente de las galaxias experimenta un desplazamiento hacia el rojo, sino también un suave silbido de ruido radiofónico que llena todo el universo y se interpreta como el resto de la radiación

electromagnética que dejó la bola de fuego en la que nació el universo. La teoría y la observación concuerdan perfectamente. Pero ¿dónde acaba la ciencia de buena calidad? Ya hemos dicho que ningún científico cree que el universo actual comenzara en un objeto singular infinitamente denso —los infinitos en las fórmulas matemáticas son un signo claro de que las fórmulas se están llevando demasiado lejos—. Entonces, ¿hasta qué punto podemos acercarnos al objeto singular en las investigaciones realizadas con el tipo de ciencia que podemos hacer en los laboratorios aquí en la Tierra? Puede que la respuesta les sorprenda.

Las condiciones más extremas de densidad que los físicos, según dicen, pueden entender del todo aquí en la Tierra son las condiciones que existen en el núcleo de un átomo. Los protones, los neutrones y los núcleos se comprenden a fondo, y el comportamiento de la materia a tales densidades se ha comprobado muchas veces en experimentos en los que intervienen los aceleradores de partículas. Ciertamente tenemos muy buenos modelos del comportamiento de la materia con densidades nucleares. La mayoría de los físicos irían más allá: añadirían que comprendemos bastante bien la estructura que hay dentro de los protones y los neutrones, a nivel de los quarks. Sin embargo, seamos cautos y limitémonos a aceptar que entendemos completamente todas las fuerzas que actúan a escala nuclear. ¿Cuánto tiempo ha transcurrido desde que todo el universo que podemos ver actualmente se podría haber encontrado en un estado de densidad nuclear, al salir del Big Bang?

Tomando el hipotético objeto singular como la «hora cero» —el momento en que el universo empezó a existir— todo lo que hoy podemos ver se habría encontrado en este estado nuclear justo durante una diezmilésima de segundo después del comienzo. Y todo lo que ha sucedido desde entonces, durante los siguientes 13.000 o 15.000 millones de años de historia del universo puede explicarse, en principio, mediante las mismas leyes de la física que se han ensayado y comprobado muchas veces en experimentos realizados aquí en la Tierra.

Lo que habitualmente se denomina Big Bang es la primera fase de la existencia del universo, a partir de una fecha que se puede datar como una diezmilésima de segundo. Y esto es ciencia de buena calidad, según los criterios que hemos establecido: los modelos concuerdan con las observaciones. Lo que está menos claro es cómo el universo llegó al estado en que se encontraba a la «edad» de una diezmilésima de segundo, pero está siendo objeto de una respetable investigación científica, como veremos en breve. Lo que es cierto es que todo lo que sucedió a partir de esa diezmilésima de segundo está tan bien fundamentado como las leyes de Newton o las ecuaciones de Maxwell. Desde luego está tan bien fundamentado como la teoría general de la relatividad, ya que se le aplica esta teoría. En números redondos, podemos decir que el Big Bang duró cerca de medio millón de años, desde una diezmilésima de segundo hasta el momento en que la radiación electromagnética

y la materia fueron por caminos separados.

La radiación electromagnética desempeñó un papel dominante en el primitivo universo de los primeros tiempos, dado que éste se encontraba muy caliente. Si se comprime el aire en una bomba de bicicleta, el aire se calienta; si se deja que el gas se expanda saliendo por la boquilla de un vaporizador, dicho gas se enfría. Del mismo modo, el universo estaba muy caliente cuando se encontraba en un estado de gran compresión, pudiéndose calcular su temperatura utilizando las fórmulas de Einstein y nuestras propias observaciones del estado en que se halla el universo actualmente. En la época de la densidad nuclear, la temperatura del universo era de cerca de un billón de grados (Celsius o Kelvin, no importa la diferencia cuando se trata de números tan grandes) y su densidad era 100 billones de veces la densidad del agua. En esta bola de fuego, cada uno de los fotones llevaba tanta energía que los pares partícula-antipartícula (como, por ejemplo, un protón y un antiprotón) se creaban constantemente a partir de la energía pura y se aniquilaban constantemente la una a la otra, liberando de nuevo energía electromagnética (fotones). Al principio hubo cantidades iguales de protones y de neutrones (y también muchos pares electrón-positrón), pero cuando el universo se expandió y se enfrió, sucedieron varias cosas.

En primer lugar, la temperatura cayó hasta el punto en que los fotones eran incapaces de formar más pares protón-antiprotón o neutrón-antineutrón, con lo que las cantidades de estas partículas quedaron congeladas. Debido a un pequeñísimo desequilibrio en las leyes de la física, que rara vez se manifiesta a las temperaturas que podemos alcanzar en la Tierra actualmente, en la bola de fuego original las cantidades de partículas y antipartículas no se habían acoplado perfectamente en pares, es decir, por cada mil millones de antiprotones había mil millones y un protones, mientras que por cada mil millones de antineutrones había mil millones y un neutrones. Todos los miles de millones de pares de partículas y antipartículas se aniquilaban unas a otras, produciendo fotones de alta energía y quedando sobrante la partícula impar que existía por cada mil millones. Todo lo que vemos actualmente en el universo está hecho a partir de esta pequeñísima parte de materia que quedó sobrante, y por cada nucleón que existe en el universo hay mil millones de fotones en la radiación de fondo.

En aquel estadio, estos fotones estaban lejos de terminar de ejercer su influencia sobre la materia. Aunque ya no tenían la energía necesaria para fabricar nuevos protones y neutrones, participaban en interacciones violentas con las partículas que quedaban, provocando la desintegración progresiva de los neutrones para convertirse en protones, proceso en el que cada uno de los neutrones soltaba un electrón y un antineutrino, para convertirse en un protón.

A medida que transcurrió el tiempo, iba descendiendo la proporción de neutrones dentro de la sopa primitiva. Justo un segundo después del comienzo, la temperatura

del universo había descendido a diez mil millones de grados, la densidad era sólo 380.000 veces la densidad del agua y sólo quedaban 24 neutrones por cada 76 protones. A los 14 segundos después de la hora cero, sólo quedaban 17 neutrones por cada 83 protones, pero el ritmo de cambio se estaba frenando dramáticamente a medida que el universo se enfriaba. Unos tres minutos después del comienzo, la temperatura del universo había descendido hasta aproximadamente mil millones de grados, que es 70 veces la temperatura que tiene hoy el núcleo del Sol, y aún había 14 neutrones por cada 86 protones.

Si los neutrones que aún existían se hubieran quedado como partículas libres, en unos pocos minutos más se habrían desintegrado de forma natural dando lugar a protones (incluso sin ser bombardeados por fotones de alta energía). Sin embargo, el universo estaba ya en ese momento lo suficientemente frío como para que se unieran algunos de esos protones y neutrones, constituyendo núcleos estables de helio-4 (partículas alfa). Para entonces el universo tenía cuatro minutos de existencia, además todos los neutrones que quedaban ya se habían bloqueado del modo anteriormente citado, dando una mezcla de 74 por 100 de núcleos de hidrógeno (protones) y un 26 por 100 de núcleos de helio (partículas alfa) cuando el universo entró en la fase siguiente de su existencia.

Toda esta historia se basa en la comprensión de las leyes de la física que han sido determinadas a partir de experimentos de laboratorio realizados aquí en la Tierra, y además también se fundamenta en el hecho observado de que el universo está en expansión y en las fórmulas cosmológicas de la teoría general de la relatividad. Uno de los éxitos más impactantes de los modelos cosmológicos es que predicen esta mezcla (26 por 100 de helio y 74 por 100 de hidrógeno), que es justo la que se observa en las estrellas más viejas, que están hechas de la materia primitiva.

Durante cientos de miles de años posteriores a la formación del helio, el universo se expandió tranquilamente. Todavía se encontraba en gran medida demasiado caliente para que los electrones quedaran ligados a estos núcleos y constituyeran átomos, por lo que los electrones vagaban libres entre los núcleos, formando un plasma, y todas las partículas cargadas (lo mismo los electrones que los núcleos) participaban en interacciones con la radiación electromagnética que aún llenaba el universo. Los fotones iban de una partícula cargada a la siguiente en una loca danza zigzagueante, exactamente como lo hacen hoy en día en el interior del Sol. El siguiente cambio dramático —el final del Big Bang— sucedió entre unos 300.000 y 500.000 años después del comienzo, cuando la totalidad del universo se enfrió hasta llegar aproximadamente a la temperatura que tiene la superficie del Sol en la actualidad, sólo unos 6.000 grados.[50] A una temperatura como ésta, los electrones y los núcleos se pueden asociar para formar átomos eléctricamente neutros; y, desde luego, los átomos que son eléctricamente neutros difícilmente pueden participar en

interacciones con la radiación electromagnética. De repente, el universo se volvió transparente. Todos los miles de millones de fotones que había en la primitiva bola de fuego se vieron libres para discurrir sin interrupciones por el espacio, mientras que los átomos podían empezar a agruparse, sin que la radiación les estorbase, para formar nubes de gas y colapsarse bajo su propia gravedad para dar lugar a las primeras estrellas, y a las galaxias.

La radiación, que tenía una temperatura de cerca de mil millones de grados cuando se formó el helio y una temperatura de 6.000 grados cuando se disoció de la materia, medio millón de años después del comienzo del universo, tendría según los modelos que haberse enfriado, descendiendo hasta llegar actualmente a unos 3 K (es decir, menos 270 grados Celsius). El segundo gran triunfo del modelo del Big Bang es que exactamente este tipo de radiación se detecta proveniente de todas las direcciones del espacio: la radiación cósmica de fondo constituida por microondas.

Esta radiación, que se detectó por primera vez en la década de los sesenta, es casi perfectamente uniforme; sin embargo, en la década de los noventa, primero el satélite COBE y luego otros detectores descubrieron unas pequeñísimas variaciones en la temperatura de la radiación que llega de diferentes zonas del cielo. Dichas ondulaciones son como marcas fósiles que fueron impresas en la radiación la última vez que participó en interacciones con la materia, 500.000 años después del comienzo del universo. Así, nos hablan de cómo estaba distribuida la materia por el universo, la última vez que hubo interacciones entre ésta y la radiación, justo antes de que dicha materia empezara a colapsarse para formar estrellas, galaxias y clusters (enjambres) de galaxias. Estas irregularidades tienen exactamente el tamaño adecuado para haber sido las semillas a partir de las cuales, y debido al colapso gravitatorio, creció la estructura que vemos en el universo actualmente, desde las galaxias hacia arriba.

¿De dónde vinieron inicialmente estas ondulaciones? ¿Qué fue lo que sembró esta estructura en el propio Big Bang? Ahora tenemos que desplazarnos a un terreno ligeramente más especulativo y traer a la cosmología las teorías de la física de partículas. Como suele pasar en cualquier campo de la ciencia que se esté desarrollando, existen varios modelos para explicar lo que sucedió al comienzo. Para simplificar, voy a dar sólo uno, mi favorito, como guía para llegar a ver el modo en que la ciencia piensa que empezó todo esto. Las variaciones sobre el tema presentan el mismo tipo de panorama a grandes rasgos, aunque los detalles pueden diferir.

La gran pregunta, tanto en filosofía y religión como en la ciencia, es por qué tiene que existir el universo: ¿por qué hay algo allí, en vez de no haber nada? Mediante una combinación de la física cuántica y la teoría general de la relatividad, la respuesta es que el propio universo puede ser «nada», en cierto modo. Concretamente, la energía total del universo puede ser cero.

Esta idea es sorprendente, ya que sabemos que existen cientos de miles de

millones de galaxias, cada una de las cuales contiene cientos de miles de millones de estrellas, y Einstein nos enseñó que la masa es una forma de energía almacenada con una elevada concentración. Sin embargo, Einstein también nos enseñó que la gravedad (espacio-tiempo deformado) es una forma de energía, y lo que resulta raro es que un campo gravitatorio en realidad almacena energía negativa. Es bastante probable que toda la energía másica que hay en el universo sea precisamente contrarrestada por toda su energía gravitatoria. El universo puede ser nada más —o nada menos— que la última fluctuación cuántica.

Para aclarar esta idea de energía gravitatoria negativa, pensemos en algo mucho más pequeño que el universo, por ejemplo, una estrella. Imaginemos que desmontamos una estrella dividiéndola en las partes que la componen y dispersando estas partes lo más alejadas unas de otras que sea posible, hasta el infinito. Es igual que consideremos que sus componentes son átomos o núcleos o quarks; el argumento es en todo caso el mismo. Recordemos que la fuerza de la gravedad obedece una ley de proporcionalidad inversa con respecto al cuadrado de una magnitud. Por este motivo, la fuerza entre dos partículas que están separadas por una distancia infinita queda dividida entre el cuadrado del infinito, lo cual muy claramente da como resultado cero. La cantidad de energía gravitatoria asociada a un conjunto de partículas que están infinitamente alejadas es también cero.

Ahora imaginemos que dejamos que las partículas caigan hasta juntarse. Desde luego, si están a una distancia infinita, esto no sucederá nunca, pero podemos figurarnos que les damos un codazo para que la bola eche a rodar.[51] Uno de los aspectos más importantes que conocemos sobre la gravedad es que, cuando un conjunto de partículas cae juntándose éstas en el mismo lugar por el efecto de la gravedad, se libera energía; las partículas se calientan. Después de todo, así es en principio cómo consiguen suficiente calor en su interior el Sol y las estrellas para que las reacciones nucleares comiencen. Sin esta liberación de energía gravitatoria en un objeto que se está colapsando, nosotros no estaríamos aquí. A nivel de partículas, lo que está sucediendo es que la energía del campo gravitatorio se está convirtiendo en energía de movimiento (energía cinética), con lo que las partículas caen más rápido. Cuando chocan unas con otras, esta energía cinética de cada partícula se pone de manifiesto en forma de calor. Pero esto es lo extraño con respecto al campo gravitatorio. Se empieza con energía cero. Luego, cuando las partículas caen juntas, el campo gravitatorio les aporta energía, lo que significa que en este momento tiene menos que energía cero. Por lo tanto, para todos los objetos reales del universo real, la energía asociada al campo gravitatorio es negativa.

Va a resultar todavía mejor. Si nos imaginamos que el colapso gravitatorio continúa todo el tiempo hasta llegar a un punto (un objeto singular como el que, según la teoría general de la relatividad, estuvo vinculado con el nacimiento del

universo), la cantidad total de energía liberada por el campo gravitatorio es exactamente la misma que la energía másica de todas las partículas que han caído juntas. Si la masa de una estrella como el Sol estuviera concentrada en un punto, la energía negativa de su campo gravitatorio compensaría exactamente la energía másica positiva de toda su materia. Habría energía cero en todas partes. El mismo argumento es aplicable a todo el universo. Si nació en un punto, en vez de colapsarse hasta un punto, entonces el conjunto de la enorme cantidad de energía másica de toda la materia del universo podría quedar equilibrada de forma exacta por la igualmente enorme cantidad de energía negativa de todo su campo gravitatorio, dando como resultado en total energía cero.

Por supuesto, del mismo modo que los efectos cuánticos difuminan el objeto singular que supuestamente se encuentra en el nacimiento del universo, así también difuminarían la singularidad asociada a un colapso de este tipo. Sin embargo, el tipo de efectos cuánticos que actúan aquí lleva implícita la incertidumbre, dado que una auténtica singularidad no puede existir en la física cuántica, porque no existe nada que pueda ser una cantidad de energía exactamente determinada en un punto del espacio-tiempo determinado exactamente.

La incertidumbre cuántica, como ya comentamos en el capítulo 3, permite que el «espacio vacío» esté vivo, con paquetes de energía que surgen de la nada y desaparecen dentro del tiempo límite que establecen las reglas de la física cuántica. Cuanta menos energía contenga, más tiempo puede existir un paquete de energía «virtual», una fluctuación cuántica, antes de que el universo se dé cuenta y dicho paquete tenga que desaparecer. Así, una fluctuación cuántica que contuviera exactamente un total de energía igual a cero podría, por lo que respecta al mundo cuántico, durar para siempre.

En la década de 1970, unos pocos cosmólogos jugaban con la idea de que todo el universo podría ser producto de una fluctuación cuántica de este tipo, es decir, algo que salió de repente del vacío sin que hubiera ninguna razón para ello, simplemente porque esto no está prohibido por las leyes de la física. El modo en que esto podría haber sucedido, según los modelos cuánticos, no sería un caso singular. Las leyes de la física cuántica dicen que es posible la existencia de un intervalo mínimo de tiempo, llamado el tiempo de Planck, y también la de un valor máximo de la densidad (la densidad de Planck), así como existe la posibilidad de un intervalo mínimo para la distancia (la longitud de Planck). Poniendo todos estos límites juntos, el universo habría nacido en un estado determinado cuando su edad era de 10^{-43} segundos (el tiempo de Planck) y habría tenido una densidad inicial de 10^{94} gramos por centímetro cúbico. La totalidad del universo visible podría haberse originado a partir de una partícula de Planck, de unos 10^{-33} cm de diámetro (la longitud de Planck), que es una centésima de una millonésima de una billonésima del tamaño de un protón.

Sin embargo, el gran inconveniente de esta teoría era que parecía obvio, en la década de 1970, que esta partícula de Planck dispondría de un tiempo de vida muy corto, fuera lo que fuera lo permitido por las reglas de la física cuántica. Después de todo, tendría un campo gravitatorio enormemente fuerte y parece ser que su gravedad destrozaría la partícula en cuanto ésta apareciese.

Pero la cosmología experimentó una transformación durante la década de los ochenta, debido a las aportaciones provenientes del mundo de la física de partículas, y especialmente de las investigaciones realizadas sobre el modo en que las fuerzas de la naturaleza se combinan cuando existen altas energías. Como dijimos en el capítulo 3, los mejores modelos del modo en que se producen las interacciones entre partículas y campos nos explican que, con muy altas energías, las cuatro fuerzas de la naturaleza que conocemos actualmente habrían estado unidas en una superfuerza. Ésta habría sido la situación en la partícula de Planck (o partículas) cuando nació el universo. Sin embargo, inmediatamente, a la escala de la longitud de Planck, la gravedad se habría separado del resto de las fuerzas y se habría ido por su propio camino; las otras fuerzas habrían seguido su ejemplo muy rápidamente. Según estos modelos, la separación de las fuerzas de la naturaleza habría convertido parte de la energía existente en un enorme empuje hacia el exterior, obligando a la semilla del universo a expandirse dramáticamente en una pequeñísima fracción de segundo, venciendo así completamente (durante ese segundo de la separación) la tendencia del campo gravitatorio a hacer que se colapsara la semilla cuántica.

La energía liberada de esta manera se compara a menudo con la energía, conocida como calor latente, que se libera cuando el vapor de agua se condensa para producir agua líquida. Cuando el agua pasa de un estado de alta energía a un estado de baja energía, se libera calor latente. Ésta es, por ejemplo, la fuerza conductora de los huracanes, donde la condensación del agua libera tanto calor que se producen unas fortísimas y turbulentas corrientes de convección de aire. Los vientos de un huracán forman remolinos a causa de la rotación de la Tierra; pero el universo no rota. El nacimiento del universo fue como un huracán cósmico, pero aquí los vientos soplaban todos en la misma dirección: hacia afuera.

Esta teoría se conoce como inflado, y la característica fundamental del inflado es que se trata de un proceso exponencial. El modo en que el universo se expande actualmente es más o menos lineal: si dos galaxias estuvieran situadas de tal manera que la distancia entre ellas se duplicara en cinco mil millones de años, entonces en otros cinco mil millones de años esta nueva distancia volvería a duplicarse (en realidad, no llegaría a tanto como duplicarse en el segundo período de cinco mil millones de años, ya que esta expansión se vuelve más lenta por la influencia de la gravedad).

Sin embargo, durante la expansión exponencial, si la distancia entre dos partículas

se duplica en un segundo, en el siguiente segundo se cuadruplica, en el tercer segundo queda multiplicada por ocho, y así sucesivamente. La expansión exponencial crece a una velocidad rapidísima. En los primeros tiempos del primitivo universo, el tiempo de duplicación era de unos 10^{11} segundos, lo cual significa que en sólo 10^{32} segundos el tamaño del primitivo universo se duplicaba cien veces. Esto fue suficiente para tomar una semilla de una centésima de una millonésima de una billonésima del tamaño de un protón e inflarla hasta que adquiriera el tamaño de un grano de uva, todo ello antes de que el universo tuviese tan sólo 10^{32} segundos de edad. En ese momento las fuerzas de la naturaleza se habían separado todas unas de otras y el inflado se detuvo. A pesar de ello, aquel universo del tamaño de un grano de uva (que contenía toda la energía másica de lo que actualmente constituye todo el universo visible) siguió expandiéndose tan rápidamente, en el Big Bang, que la gravedad no pudo detener la expansión durante cientos de miles de millones de años.

[52]

Los astrónomos están emocionados con la teoría del inflado, porque proporciona un modelo que explica cómo comenzó el Big Bang. Los expertos en física de partículas también están emocionados con esta teoría, porque les proporciona una comprobación de las suyas propias, con energías mucho mayores que las que se pueden conseguir aquí en la Tierra. Ambos campos de la física están encantados porque el inflado, como todas las buenas teorías científicas, realiza predicciones que se pueden comprobar; en este caso mediante la observación del universo a gran escala. En general, la teoría del inflado supera las pruebas, aunque todavía hay muchas cosas que no entendemos en relación con el nacimiento del universo.

La comprobación más importante que todo este paquete de teorías ha superado hasta el momento se refiere a esas ondulaciones en la radiación de fondo que ya hemos mencionado anteriormente. Las ondulaciones que vemos en la actualidad fueron impresas cuando el universo tenía alrededor de un millón de años, y muestran que la materia (en forma de nubes de gas caliente) estaba distribuida en aquella época justo de la manera adecuada para colapsarse y formar estrellas y galaxias dentro de un universo en expansión. Según la teoría del inflado, cuando el universo que hoy día es visible para nosotros ocupaba un volumen de sólo 10^{-11} cm de diámetro (cien millones de veces mayor que la longitud de Planck), las fluctuaciones cuánticas «ordinarias», como las que ya hemos descrito en las páginas precedentes de este libro, hubieron de producir unas ondulaciones muy pequeñas en la estructura del universo. Al mismo tiempo que expandía todo lo demás durante el inflado, el proceso de crecimiento exponencial tuvo que tomar esas ondulaciones cuánticas y estirarlas, dejando por todo el universo su huella en forma de variaciones en la densidad de la materia de un lugar a otro.

El tipo de ondulaciones ocasionadas en la radiación de fondo y observadas por el

COBE y por otros detectores sería producido sin duda por una ampliación de las fluctuaciones cuánticas realizada de este modo. Si los modelos son correctos, sin la fluctuación cuántica primordial no habría existido en absoluto un universo; y sin esas fluctuaciones cuánticas secundarias no habría seres humanos que hicieran cábalas sobre el origen del universo, ya que todo habría sido liso y uniforme, y así no se habrían creado las estrellas. Ningún otro modelo que se haya desarrollado hasta ahora puede explicar por qué el universo es muy uniforme por todas partes (gracias al inflado), pero sin embargo contiene exactamente el tipo adecuado de ondulaciones que se necesitan para formar las galaxias y los clusters de galaxias (también gracias al inflado).

Éste es un ejemplo del modo en que la teoría del inflado explica algo que ya había sido observado: el hecho de que existimos. ¿Puede hacer una predicción real y estable? ¿Una predicción que se pueda comprobar en laboratorios situados aquí en la Tierra? Sí. Uno de los requisitos clave que es común a todas las variaciones sobre este tema es que la cantidad de materia existente en el universo debe, desde luego, equilibrar la energía negativa del campo gravitatorio. Este preciso equilibrio es una condición requerida para que el universo se expanda hasta el infinito y se quede así, detenido, sin volver a colapsarse (la imagen en espejo, inversa en el tiempo, de nuestra descripción inicial de una nube de partículas infinitamente dispersas que se colapsan confluyendo todas en un punto). En el lenguaje de la teoría general de la relatividad, esto significa que el espacio-tiempo debe ser plano en todas partes, sin curvaturas pronunciadas. [53]

En un lenguaje más cotidiano, esto quiere decir que la expansión que vemos en curso todavía actualmente no continuará siempre, pero tampoco será invertida por efecto de la gravedad para hacer que el universo vuelva a colapsarse. Esta suspensión, por un tiempo infinito, de la expansión sólo se producirá si hay una cantidad crítica de materia en el universo, que correspondería a unas cien veces más materia que toda la que vemos brillando en forma de estrellas y galaxias. En otras palabras, la teoría del inflado predice que la mayor parte de la masa del universo se encuentra en forma de materia oscura.

Esto resulta magnífico para los astrónomos, porque ya saben que hay más universo que el que ven nuestros ojos. A partir de estudios realizados sobre el modo en que las galaxias rotan y se mueven por el espacio, está claro que están siendo sometidas a una atracción causada por materia invisible; en efecto, las evidencias astronómicas muestran que la cantidad de materia oscura que existe es al menos diez veces la cantidad de materia brillante, y puede que cien veces esta cantidad.

Sin embargo, los cálculos realizados sobre la manera en que la materia fue procesada en el Big Bang, que ya hemos esbozamos anteriormente en este capítulo, establecen límites en relación con la cantidad de esta materia (brillante u oscura) que

puede haber en forma de protones y neutrones (que juntos constituyen núcleos de materia bariónica, la clase de materia de la que estamos hechos). Los procesos que tuvieron lugar en el Big Bang —interacciones en las que participaban protones, neutrones, electrones y neutrinos— son procesos que están decisivamente afectados por la densidad de esta materia durante unos pocos primeros segundos, en el momento en que se están produciendo estas interacciones. Si el objetivo era obtener la proporción correcta de 26 por 100 de helio y 74 por 100 de hidrógeno, hoy en día no podemos tener en forma de bariones más que un pequeño porcentaje de la cantidad crítica de materia. Puede haber otros tipos de materia, con tal de que ésta no se vea implicada en las reacciones de fusión nuclear que produjeron helio cuando el universo era joven, pero lo que no se puede es conseguir que el universo sea plano utilizando sólo bariones.

Esto es magnífico para los expertos en física de partículas, porque sus teorías estándar sobre partículas y campos predicen también la existencia de diferentes tipos de partículas, que no se parecen a nada de lo que hemos visto hasta ahora; se trata de materia no bariónica. Estas partículas nunca se han detectado porque, según los modelos, no son sensibles a ninguna gran fuerza, ni a fuerzas electromagnéticas (ni siquiera son sensibles a las fuerzas débiles). Sólo participan en interacciones con el resto del universo por efecto de la gravedad; de hecho, quedaron apartadas de cualquier otra materia desde el momento en que la gravedad se separó de las otras fuerzas, en los tiempos remotos del nacimiento del universo. Aunque estas partículas están aún por detectar, los modelos predicen qué tipo de masas han de tener, así como otras propiedades, y se les han dado nombres tales como, por ejemplo, fotinos.

De un modo más general, todos los objetos de este tipo se denominan Partículas Masivas de Interacción Débil, en inglés Weakley Interacting Massive Particles con las siglas WIMP; este nombre viene determinado por la sencilla razón de que poseen masa, pero no participan muy activamente en interacciones con la materia ordinaria. Aunque en principio pueden existir muchas clases diferentes de WIMP, y se supone que pudieron formarse profusamente en las condiciones de energía de los primeros tiempos del universo primitivo, es probable que sólo las más ligeras de las WIMP (posiblemente el fotino) sean estables y que todas las demás WIMP se hayan desintegrado de forma progresiva, en fases muy tempranas de la historia del universo, hasta llegar a su estado más ligero.

Desde luego, lo que les gustaría tanto a los astrónomos como a los expertos en física de partículas es tener un modo de detectar estas partículas directamente, lo cual puede que se consiga dentro de unos pocos años. Ahora que ya se ha llegado a sospechar su presencia, tendría que ser relativamente fácil capturar una WIMP, dado que las supuestas masas de estas partículas son aproximadamente la masa del protón, y también porque los modelos sugieren que estamos nadando en un mar de WIMP.

Si las masas de las WIMP están en la misma gama de valores que la masa de un protón o de un neutrón, y actualmente en el universo la cantidad de materia oscura es hasta cien veces la cantidad de materia brillante, entonces la cantidad de WIMP debe ser hasta cien veces la de bariones. A diferencia de los bariones, que se concentran en las estrellas y las galaxias, las WIMP estarían distribuidas más uniformemente por todo el universo (hay una prueba indirecta de esto en la forma de los patrones que dibujan los clusters de galaxias en el cielo; las simulaciones por ordenador muestran que los patrones de galaxias observados sólo pueden haberse formado si la materia bariónica brillante queda sumergida en un mar más uniforme de materia fría y oscura).

Unos sencillos cálculos sugieren que pueden existir unas 10.000 WIMP en cada metro cúbico del espacio que nos rodea; no sólo en el espacio vacío, sino también en el que está ocupado por la materia ordinaria, incluidos el aire que respiramos, la Tierra sólida y el núcleo del Sol (aunque la cifra parece menos impresionante, si recordamos que a una temperatura de cero grados Celsius cada centímetro cúbico de aire al nivel del mar contiene 45 trillones de moléculas). Estas WIMP correrían por un laboratorio a una velocidad de un par de cientos de kilómetros por segundo, como un enjambre de abejas cósmicas, y la mayoría de ellas pasarían a través de cualquier objeto del tamaño de un ser humano sin tan siquiera notar que está allí.

De todas estas WIMP sólo unas pocas chocarían contra el núcleo de un átomo y serían rechazadas por él (recordemos que, tal como Rutherford descubrió, un átomo es sobre todo un espacio vacío en el que está contenido un núcleo central pequeñísimo). En un kilogramo de materia (cualquier materia) existen cerca de 10^{27} bariones (mil cuatrillones de bariones). Cada día una cantidad de WIMP comprendida entre unos pocos y unos pocos cientos (dependiendo de sus propiedades exactas, que aún no se conocen) participará en una colisión nuclear dentro de un kilogramo de materia, siendo ésta la base sobre la que trabajan los detectores de WIMP que están empezando a funcionar en la actualidad.

Si sucede realmente que una WIMP choca con el núcleo de un átomo, habrá una repercusión en este núcleo. Sin embargo, en condiciones ordinarias los núcleos están recibiendo sacudidas constantemente: por parte de las radiaciones cósmicas del espacio, por el movimiento térmico, etc. Este ruido de fondo ha de ser eliminado si se desea discernir los efectos de las colisiones con las WIMP. Por ello, los detectores de WIMP toman bloques de materia muy pura que se mantiene aislada de influencias exteriores a muy bajas temperaturas, en laboratorios situados en el fondo de pozos mineros o en túneles excavados bajo montañas. En esos laboratorios, los bloques de materia se enfrían a pocos grados sobre el cero absoluto (unos pocos grados Kelvin) y se controlan mediante detectores muy sensibles para ver si sufren perturbaciones por efecto del impacto de una WIMP.

Están empezando a entrar en funcionamiento varios detectores con sensibilidad para registrar el impacto de aquellas WIMP que tienen el tipo de propiedades establecidas por las Grandes Teorías Unificadas de la física de partículas (las mismas propiedades que exigen los astrónomos para explicar la naturaleza de la materia oscura del universo) y ya se han dado un par de casos en que un equipo de científicos ha comunicado la detección de estas partículas, aunque estas detecciones han sido descartadas categóricamente, porque las han considerado prematuras otros científicos que no pertenecen a los equipos que alegan la detección.

Los experimentos decisivos se llevarán a cabo a principios del siglo XXI; entonces, si las WIMP existen, se verán realmente de forma individual. Por primera vez tendremos contacto experimental directo con unas partículas que representan el 99 por 100 de la masa del universo. Sin embargo, si estos experimentos dieran resultados negativos, esto significaría, de una manera decisiva, que las WIMP no existen, es decir, que los modelos están equivocados.

Eso sería una lástima, porque lo más emocionante en relación con toda la historia de estas partículas es que las propiedades de las WIMP que se han predicho mediante los modelos son exactamente las propiedades de las partículas que necesitan los cosmólogos para determinar la existencia de la materia oscura del universo. Se trata de la concordancia más extraordinaria entre ramas de la ciencia que inicialmente parecen estar yendo en direcciones divergentes, la una explorando hacia el interior en el mundo de lo muy pequeño, y la otra mirando hacia afuera en el mundo de las grandes dimensiones.

Si se detectan las WIMP dentro de unos pocos años, y tienen las propiedades predichas, este encaje entre lo grande y lo pequeño constituirá la mayor justificación del método científico entendido en su totalidad. Sin embargo, si no se detectan, tendremos que ponemos a reflexionar de nuevo, y no sobre el método científico, sino sobre este conjunto particular de modelos. En cualquier caso, ahora tiene usted una butaca de primera fila para presenciar la comprobación de una de las teorías más importantes y de mayor alcance que la ciencia ha propuesto jamás. No obstante, sea cual sea el resultado de esta comprobación, éste no alterará la verdad de la cosa más fundamental e importante que la ciencia nos debe enseñar. No importa lo hermoso que pueda ser un modelo, no importa que ahora todo parezca encajar de una forma natural; si el modelo no concuerda con el experimento, es falso.

Notas

[1] Por razones históricas, esta radiación se conoce como radiación del «cuerpo negro», porque este modelo también explica cómo es absorbida la radiación por un objeto negro. Sin embargo, un objeto caliente y radiante puede emitir cualquier color de la radiación del «cuerpo negro», siempre que tenga la temperatura adecuada. <<

[2] *The Character of Physical Law*, p. 129. <<

[3] Cuando un electrón está en un nivel de alta energía (llamado un estado «excitado») y puede elegir entre dos o más niveles de energía inferior para saltar, elige su destino totalmente al azar. Ésta es una de las características del llamado «salto cuántico». La otra característica es que el salto es muy pequeño para lo que es habitual en el mundo que percibimos cotidianamente. Los espacios situados entre los escalones de energía son sólo lo suficientemente grandes para permitir la emisión (o absorción) de un fotón de luz cada vez. Por lo tanto, un salto cuántico es un cambio muy pequeño que se realiza totalmente al azar. <<

[4] Una antipartícula es una especie de imagen de una partícula vista en el espejo, con propiedades opuestas, como sucede en el caso de la carga. Por ejemplo, un antielectrón (conocido también como positrón) tiene una unidad de carga positiva, pero posee la misma masa que un electrón. <<

[5] Es posible hacer que las reacciones químicas marchen en sentido opuesto, de tal forma que los electrones vayan a parar a niveles de energía superior, pero esto sólo se consigue si hay un aporte de energía procedente de algún otro lugar. Desde luego, esta forma de inversión es especialmente importante en la química de la vida en la Tierra, donde la fuente de energía es en última instancia el Sol. <<

[6] Éste es un buen ejemplo de cómo funcionan los efectos cuánticos. Para que un protón individual sea visible de esta manera, el velo de electrones ha de ser en realidad más fino que un simple electrón. Pero debe haber algún tipo de cobertura eléctrica o si no el protón formaría un enlace mucho más fuerte. Es exactamente como si un electrón se aplastara y se estirara hasta llegar a ser parcialmente transparente desde un punto de vista eléctrico. <<

[7] Dentro de ciertos límites. Actualmente sabemos que, si una sustancia se calienta lo suficiente, se deshará quedando separadas sus componentes; pero aquí estamos hablando de temperaturas moderadas, por ejemplo, algunos cientos de grados Celsius.

<<

[8] También explica cómo el fósforo, una excepción a la regla, puede formar a veces cinco enlaces simultáneamente, aunque no tiene una vida química tan interesante como la del carbono. El fósforo tiene número atómico 15 y cinco electrones en su capa ocupada más externa. En efecto, puede constituir cinco enlaces utilizando cuatro de estos electrones para formar enlaces covalentes y cediendo el quinto electrón para hacer un enlace iónico. Sin embargo, los cinco enlaces que forma son idénticos en cuanto a su carácter, a saber, una mezcla de una quinta parte de enlace iónico y cuatro quintas partes de enlace covalente, la cual conforma otro bello ejemplo de un proceso puramente cuántico. <<

[9] Por cierto, un ácido es sencillamente una sustancia cuyas moléculas liberan hidrógeno en las reacciones químicas con una relativa facilidad. Una base es una sustancia que cede el grupo OH. Cuando un ácido y una base reaccionan entre sí, producen agua (H_2O) y un tipo de compuesto conocido generalmente como sal (al menos en la química inorgánica). <<

[10] Lo que está haciendo el químico es, por consiguiente, suministrar energía al sistema y desplazar los átomos y moléculas desde el estado de energía más baja hasta hacerlos anidar en un hueco, o una serie de huecos, situados más arriba en la colina de la energía. <<

[11] En la década de 1980, un juguete conocido como serpiente de Rubik fue famoso durante un breve tiempo; cualquiera que tenga suficiente edad como para recordar dicha serpiente podrá imaginarse que la capacidad limitada para enrollarse que tiene una cadena de polipéptidos es similar a la capacidad limitada para enrollarse que tenía la serpiente de Rubik. <<

[12] Dicho sea de paso, en una molécula de ADN cada grupo fosfato tiene un exceso de carga negativa resultante del modo peculiar en que los enlaces están distribuidos (véase la nota 8 en el capítulo 4), por lo que todas las moléculas de ADN tienen realmente carga negativa y se deberían llamar, por decirlo con rigor, «iones moleculares». Pero no vamos a preocuparnos por estas sutilezas de la nomenclatura, ya que casi nadie lo hace. <<

[13] Mientras que Crick y Watson no podrían haber resuelto el problema sin sus fotografías. <<

[14] Cada uno de los tipos de base (A, C, G y T) se presenta, por supuesto, en ambas hebras. Cuando decimos «A opuesta a T», esto incluye también la situación inversa, en la que T es opuesta a A, y lo mismo para GC y CG. <<

[15] Existe una descripción ligeramente simplificada. De hecho, la construcción de la nueva pareja de hebras tiene lugar simultáneamente en varios emplazamientos a lo largo de cada hebra original, creciendo nuevos fragmentos de ADN a lo largo de dicha hebra original y ligándose unos con otros. <<

[16] Todo esto es la descripción del tipo de células que constituyen el cuerpo humano y los cuerpos de los animales y las plantas. Una célula de este tipo, con un núcleo claramente definido, recibe el nombre de célula eucariota. Las bacterias son diferentes; tienen una estructura celular que no está tan claramente definida, en la que no existe núcleo y el ADN no está dispuesto formando cromosomas, sino enrollado, constituyendo lo que se denomina un nucleoide. Las células de este tipo se llaman procariotas y representan claramente un estadio previo en la evolución. Sin embargo, no se puede decir que se trate de un fallo en el proceso evolutivo, ya que las bacterias siguen siendo en la actualidad una parte importante de la vida en la Tierra. <<

[17] La única diferencia real está en el número de los cromosomas que participan en el proceso. En los seres humanos hay 23 pares, pero el número de pares de las diferentes especies parece no guardar relación con la clase de «cuerpo» en el que habitan. Así, por ejemplo, hay 7 pares en el guisante de huerta, 24 pares en cada célula de una patata (curiosamente uno más de los que tenemos nosotros) y 100 pares de cromosomas en cada célula de un cangrejo de río. <<

[18] Poca gente supo de la obra de Mendel mientras éste vivió, ya que Mendel no tenía relación con las corrientes importantes del mundo científico y publicó sus principales hallazgos en 1866 en una revista bastante poco conocida, llamada *Proceedings (Métodos)* de la Sociedad para el Estudio de las Ciencias Naturales de Brno. <<

[19] O, desde luego, puede suceder que el nuevo alelo sea neutral, es decir, que no conlleve ni ventajas ni inconvenientes. En ese caso, se quedará indefinidamente en la reserva de genes de la especie. Sin embargo, puede que un día algún cambio en el medio ambiente exterior haga que ese alelo se convierta en algo ventajoso para la especie, en cuyo caso empezará a propagarse. <<

[20] Aunque resulte asombroso, doscientos años después de la primera edición del ensayo de Malthus, todavía hay gente ignorante que en ocasiones plantea un razonamiento similar como argumento para que no se aporte ayuda a los países pobres. Sin embargo, hoy en día es fácil hallar una salida para evitar la trampa de Malthus: el hecho de disponer de medios anticonceptivos eficaces significa que las poblaciones humanas ya no se incrementan automáticamente cuando el nivel de vida sube y una mayor proporción de los niños nacidos vive hasta alcanzar la madurez. Lo que sucede es sencillamente que la gente tiene menos hijos, con la seguridad de saber que los que tienen no morirán en la infancia. <<

[21] Felizmente, esta datación molecular concuerda bien con las pruebas fósiles. <<

[22] Si desea usted saber más, vea *Being Human* de Mary y John Gribbin (Phoenix, 1995). <<

[23] Algo que quizá tenga un significado equivalente es que, en una comunidad pequeña, si ayudáramos a alguien de este modo, había bastantes probabilidades de que, si más adelante en nuestra vida estuviéramos en peligro, podríamos tener cerca a alguien que nos ayudase. Esto reforzaría el argumento que se esboza aquí. <<

[24] En realidad, el sexo de los seres humanos y especies similares está determinado por un gen que se encuentra en un par de cromosomas concretos. Todas las hembras tienen un par idéntico de estos cromosomas, que se denomina XX, de manera que cada madre debe transmitir un X a cada uno de sus descendientes. Todos los machos tienen un par de cromosomas no idénticos llamado XY. Dado que pasa a cada uno de sus descendientes un cromosoma de este par, seleccionado al azar, hay una probabilidad exactamente del 50 por 100 de que el descendiente sea XX o XY. En términos evolutivos, sería una mutación muy sencilla la que garantizara que un macho pasase todos los X o todos los Y a su descendencia. El hecho de que una mutación así no se haya propagado en la reserva genética humana confirma que el equilibrio de los sexos al 50 por 100 es una estrategia evolutivamente estable. <<

[25] Includo, dicho sea de paso, el astrónomo George Darwin, uno de los hijos de Charles Darwin. Esto sirve para recordarnos que tuvimos una buena teoría de la evolución mucho antes de tener una buena teoría sobre la deriva continental. <<

[26] «Rift es morfológicamente un surco continental alargado y estrecho que está limitado por fallas normales de distensión. Marca una zona a lo largo de la cual la litosfera se ha fracturado bajo condiciones tectónicas de extensión o relajación.» Bates, R. L. y Jackson, J. A. (eds.) (1987), *Glossary of Geology* (3a edición), American Geological Institute, Alexandria, Virginia (USA), p. 569. (N. de la T.) <<

[27] Aunque el manto es sólido, debido a que su materia está caliente, ésta puede aún desplazarse siguiendo unas pautas de convección muy lentas, fluyendo más bien como el material que se denomina masilla ligera. Es fluido en este sentido, pero no es líquido. <<

[28] Las islas volcánicas del Pacífico, por ejemplo las Hawai, son una excepción a esta regla. Aunque se encuentran lejos de cualquier límite entre placas, se trata de lugares donde un punto caliente situado en el manto ha perforado un orificio en la delgada corteza oceánica. Cuando la placa del Pacífico cruza en su deriva el punto caliente, éste perfora una serie de orificios en la corteza, produciendo una cadena de islas volcánicas. Curiosamente, todas las cadenas de islas del Pacífico de este tipo describen una curva que muestra cómo la deriva de la totalidad de la placa cambió de dirección hace unos 40 millones de años, desde una deriva más dirigida hacia el norte hasta la deriva hacia el noroeste que vemos actualmente. Este cambio se produjo cuando América del Norte avanzó sobre la parte norte de la vieja dorsal del Pacífico, dándole a la placa un empujón hacia el oeste. <<

[29] Hay que matizar que el cámbrico no es otra era, sino meramente un período geológico, que duró unos cien millones de años. Es el primer período de la era paleozoica. <<

[30] O ambos, pero solamente el polo Sur estaba afectado al final del ordovícico. <<

[31] Esto sólo resulta bien si los nuevos pastos no están ya ocupados. La razón por la que la vida marina no se traslada a tierra actualmente es que ya hay mucha vida asentada en este medio y cualquier hipotético protoanfibio que se dejara caer saliendo del agua, sería pronto devorado. <<

[32] La prueba irrefutable de que el suceso que cerró el cretácico fue efectivamente causado por un impacto procedente del espacio está explicada en nuestro libro *Fire on Earth* (Simon and Schuster). <<

[33] En el original, en inglés, compara las expresiones «ferrous ions» y «iron ions». (N. de la T.) <<

[34] Sin embargo, recordemos que actualmente están inmovilizadas en las rocas 60 atmósferas de dióxido de carbono, por lo que, teniendo en cuenta lo que ha sido emitido por los volcanes desde que se formó el planeta, la totalidad de la atmósfera de hoy en día es sólo un 1'6 por 100 de lo que podría haber sido si la Tierra hubiera terminado siendo como Venus. <<

[35] Esto es posible debido a una resonancia cuántica, como la que hace que el ion de carbono sea estable, comentada ya en el capítulo 4. Si un átomo de oxígeno cede un electrón a otro átomo de oxígeno, el primero se queda con una carga positiva y una capa exterior de cinco electrones, por lo que puede hacer tres enlaces covalentes, uno de ellos con el átomo de oxígeno que ha captado el electrón y tiene ahora siete en su capa exterior, y otros dos enlaces con un átomo normal de oxígeno. La resonancia garantiza que, de hecho, cada uno de los enlaces de la molécula de oxígeno sea como un enlace y medio de los normales. Sin la resonancia cuántica no existiría la capa de ozono (entre otras cosas) y nosotros no estaríamos aquí. <<

[36] Por supuesto que, dado que lo que estamos viviendo es algo que ha evolucionado en la Tierra, en cierto sentido lo que hacemos es también «natural», pero confiamos en que esté claro lo que queremos decir. <<

[37] Es decir, la penetración en la estratosfera de grandes cantidades de aire que se desplacen en bloque no es posible; pero las moléculas aisladas que están en el aire pueden cruzar la barrera por difusión, que es el modo en que los CFC llegan allí y causan el daño consiguiente. <<

[38] Esta teoría fue propuesta por primera vez por John Gribbin y Jeremy Chertas al principio de los años ochenta y desarrollada en su libro *The Monkey Puzzle*. En aquel momento fue considerada como una extravagante especulación; sin embargo, tras el desarrollo de estudios moleculares relativos al ADN de distintas especies, los resultados hicieron que la idea se viera como respetable en la década de los noventa, cuando fue aceptada por Simon Easteal y Genevieve Herhert, de la Universidad Nacional de Australia. Actualmente está totalmente aceptado que, en sus propias palabras, «el antepasado común de los seres humanos y de los chimpancés fue bípedo; por consiguiente, lo que sucede es que esta característica se ha perdido en los chimpancés, y no que sea algo adquirido por los seres humanos» (*Journal Molecular Evolution*, febrero de 1997). <<

[39] Un año luz es, por definición, la distancia que la luz puede recorrer en un año. No es una medida de tiempo, sino de distancia, equivalente a 9'46 billones de kilómetros. A modo de ejemplo, digamos que la luz cubre los 150 millones de kilómetros que separan al Sol de la Tierra en 499 segundos. En consecuencia, la distancia de la Tierra al Sol es de 499 segundos luz, o 8'3 minutos luz. <<

[40] La forma de espiral de una galaxia como nuestra propia Vía Láctea es ligeramente parecida a la forma en espiral que hace la crema cuando se remueve en el café. <<

[41] Es decir, la mayor parte de la masa que se colapsó; algo de la masa original se disipó en el espacio por el calor de la nube que se enfriaba, o fue arrastrada lejos por campos magnéticos, contribuyendo así a sustraer parte del momento angular. <<

[42] Se da la coincidencia de que se necesitarían 107 Soles, colocados uno junto a otro, para cubrir la distancia que hay desde el Sol a la Tierra, siguiendo el radio de su órbita, no el diámetro. <<

[43] Venus se encuentra aún más cerca, a unos 42 millones de kilómetros, pero esto sucede sólo cuando se sitúa directamente entre nosotros y el Sol, siendo imposible su estudio mediante telescopios, ya que lo obstaculiza la deslumbrante luz solar y además se da el hecho de que estaríamos mirando el lado nocturno del planeta, dejando a un lado otra circunstancia: que Venus está totalmente cubierto de nubes. <<

[44] Tradicionalmente, los astrónomos utilizan la palabra «evolución» para referirse al ciclo vital de un objeto determinado, tal como una estrella o una galaxia, y concretamente al modo en que cambia a medida que envejece. No se refieren a la evolución darwiniana, es decir, no hablan de una variedad determinada de estrellas o galaxias que sean reemplazadas por otras especies cósmicas. <<

[45] Y el diámetro de la órbita de la Tierra lo conocemos aplicando técnicas similares de triangulación dentro del sistema solar. <<

[46] El número de átomos de hidrógeno que se necesitarían para formar con todos ellos una estrella que tuviera la cantidad mínima de masa es aproximadamente 10^{80} , un número que los astrónomos recuerdan como «el parámetro de la sopa Heinz», aunque lamentablemente, según nos han dicho, nunca ha habido en realidad «57 variedades» de productos Heinz. <<

[47] Para ver esto con una cierta perspectiva, recordemos que actualmente el Sol tiene cerca de un millón de veces el volumen de la Tierra. <<

[48] Para ver esto con una cierta perspectiva, recordemos que actualmente el Sol tiene cerca de un millón de veces el volumen de la Tierra. <<

[49] Al mismo tiempo que existen las nebulosas formadas por estrellas, hay también otras manchas borrosas, llamadas también nebulosas, que no son más que simples nubes de gas situadas dentro de nuestra galaxia. No queremos que la confusión de nombres ocasione aquí ningún problema. <<

[50] Vale la pena hacer una pausa aquí para reflexionar, para tratar de situar el Big Bang en perspectiva. En la historia del Big Bang, viene a ser ésta la primera vez que se encuentra una temperatura que podemos entender en términos humanos. Al fin y al cabo, todos hemos visto el Sol y hemos sentido su calor, a una distancia de 150 millones de kilómetros. En la época en que la materia y la radiación se desemparejaron, el universo entero era justamente como lo que es hoy la superficie del Sol. Después, para conseguir esta fantástica temperatura (según apreciaciones humanas) que tiene en la actualidad, se ha estado enfriando durante medio millón de años. <<

[51] Todo esto constituye un ejemplo de lo que los astrónomos llaman un argumento «para salir del paso», pero sirve para lo que nos proponemos hacer. Si los cálculos equivalentes se realizan como es debido, es decir, utilizando la teoría general de la relatividad, proporcionan la misma respuesta sorprendente. <<

[52] Hay una curiosa e intrigante variación sobre esta teoría y no nos resistimos a mencionarla. Hemos sugerido que, en el equivalente cuántico al objeto singular, podía haber aparecido la semilla del universo completo partiendo de la nada. Pero estas singularidades cuánticas son precisamente lo que se cree que se forma en los núcleos de los agujeros negros, donde la materia se colapsa hasta llegar a ser sólo un punto. Algunos científicos proponen seriamente que este colapso hacia la singularidad se puede invertir mediante procesos cuánticos, incluyendo el inflado, el salto (o, en cierto sentido, la desviación hacia un lado) para crear nuevos universos que se expandan en sus propias versiones del espacio y el tiempo. Todos los agujeros negros de nuestro universo pueden ser la puerta de entrada a otros universos, configurando un mar infinito de universos burbuja en la vasta extensión del espacio y del tiempo, y haciendo que se abandone la idea de un único origen para todo lo existente. Además, dado que la energía total que participa en el colapso y en la expansión es igual a cero no importa si es mucha (o poca), ni cuánta es la masa que entra en el agujero negro que se está colapsando: de todas formas puede configurar un universo completo al otro lado. <<

[53] Existen sutiles variaciones sobre el tema que permiten que el universo sea casi plano, pero no del todo. Esto no afecta a la discusión que se presenta a continuación.

<<