Codificación de la geometría o ¿cómo se lo digo a la computadora?

Con la información que hemos desarrollado a lo largo de los capítulos anteriores tenemos todos los datos necesarios para crear una gran variedad de modelos de moléculas importantes, o incluso crear moléculas que nunca antes han sido sintetizadas o descriptas.

La receta es sencilla.

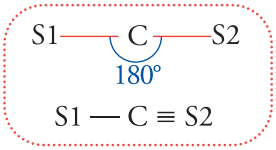
1. **¿Ingredientes?** Los que indique la fórmula molecular. Podemos elegir la cantidad de átomos de carbono, de oxígeno, hidrógeno, etc.
2. **¿De qué forma los conecto?** Elegimos una de todas las posibles combinaciones que permite el sistema de Lewis.
3. **¿A qué distancia se encontrará un elemento del vecino inmediato?** Aquellos átomos que se encuentran ligados estarán a la distancia que indiquen sus respectivos radios covalentes.
4. **¿Cuál será la distancia mínima entre los átomos que no están unidos directamente?** No podrán estar a una distancia menor que la suma de sus radios de Van der Waals.
5. **¿Cómo se dispondrán en el espacio los enlaces formados?** Lo harán según lo marque la TRePEV.

Ésta es una receta muy completa, pero podemos tener algunas dudas a la hora de terminar la molécula. Se trata de lo que vimos en la parte final del capítulo anterior. En una molécula de más de cinco átomos las geometrías deben combinarse. Esto quiere decir que cada átomo de carbono, nitrógeno, etc. presentará su hibridación que tendrá su orientación particular con respecto a su átomo vecino según sea el tipo de hibridación que presenten. También la forma de conectar los átomos a cada carbono presenta alternativas que hay que considerar.

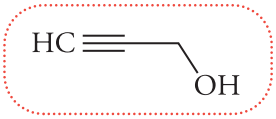
Variaciones en la conexión entre átomos

Hibridación sp^1

Esta hibridación de orbitales (un orbital *s* con un orbital *p*), permite que el átomo de carbono forme un enlace simple y un enlace triple que la TRePEV nos indica que forman un ángulo de 180°.



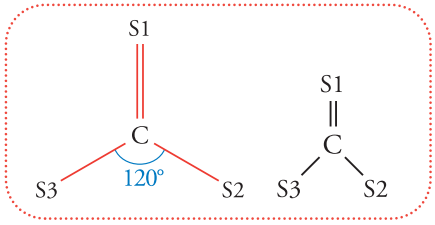
Por razones de mecánica cuántica, que describe en detalle cómo es la hibridación de tipo sp^1, el enlace triple impide que el sustituyente S2 rote. Por su parte, el sustituyente S1 puede rotar, libremente, teniendo como eje el enlace al carbono. A cada posición que podamos distinguir del sustituyente S1 decimos de la molécula que se trata de un confórmero. Para el uso práctico de los modelos moleculares es necesario definir la cantidad de confórmeros. Por ejemplo, analicemos el caso del modelo de propinol o también conocido como alcohol propergílico.



El carbono unido al oxidrilo (-OH) se encuentra unido por un enlace simple al carbono ***sp¹***, por lo tanto, el grupo oxidrilo puede rotar libremente, pero como el resto de la molécula es lineal y, por ello, no presenta asimetría, cualquier rotación a lo largo del eje carbono ***sp³***-carbono ***sp¹*** será indistinguible una de la otra, por lo tanto, tendrá una única conformación (si consideramos al oxidrilo como un único grupo). Si por el contrario lo consideramos como un grupo compuesto por un oxígeno[1] unido por un enlace simple a un hidrógeno, entonces éste puede rotar libremente y tenemos infinitas posibles conformaciones.

Hibridación *sp²*

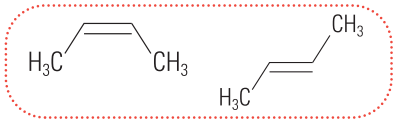
Este caso es similar al anterior ya que de la hibridación de un orbital ***s*** y dos orbitales ***p*** resultan dos enlaces simples y un enlace doble que impide que los átomos roten uno respecto del otro. La TRePEV establece que estos tres enlaces formarán un ángulo de 120° cada uno con su vecino.



Nuevamente, los sustituyentes que se encuentran unidos por un enlace simple sí pueden rotar.

Ahora nos interesa conocer las formas en que se puede presentar la unión de dos carbonos con hibridación ***sp²***.

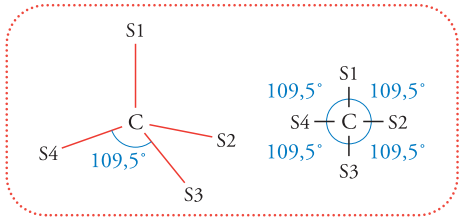
Veamos el n-2-buteno.



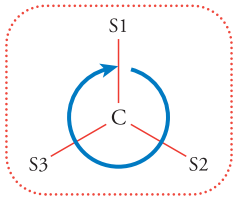
Éstas son las dos formas en que podemos dibujar a esta molécula. Si los sustituyentes se encuentran para el mismo lado del doble enlaces se habla del isómero[2] *cis*, cuando se trata del otro caso se habla del isómero *trans*.

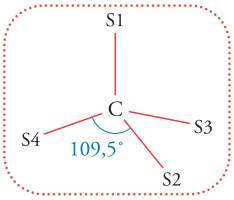
Hibridación *sp³*

En este caso el carbono presenta una geometría tetraédrica producto de la hibridación de un orbital ***s*** con tres orbitales ***p***. Cada sustituyente se encuentra unido al carbono central por un simple enlace.



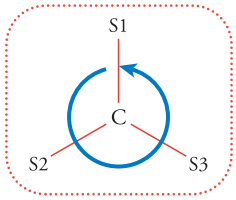
Cada sustituyente puede girar, libremente, dando lugar a infinitos confórmeros. Si observamos cualquier carbono con esta hibridación y con cuatro sustituyentes distintos a lo largo de un enlace, por ejemplo, poniendo el sustituyente S4 por detrás del carbono, podemos reconocer un sentido de giro marcado por el orden de los sustituyentes (de S1 a S3).





Podríamos haber dibujado nuestro tetraedro de la siguiente forma.

Lo que observamos si volvemos a poner el sustituyente S4 detrás del carbono es lo siguiente (yendo se S1 a S3).



[1] El átomo de oxígeno tiene dos pares de electrones no compartidos por lo tanto presenta una configuración lineal angular, con un ángulo definido por los enlaces < 109.5°.

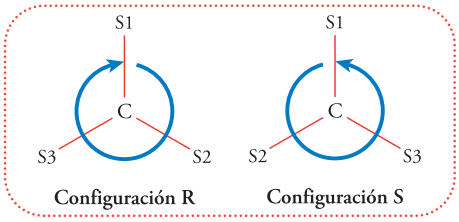
[2] Isómero: la diferencia entre este concepto y el de confórmero es que para pasar de un isómero al otro, a diferencia de los confórmeros, hay que romper y volver a formar enlaces.

El primer giro fue en el sentido de las agujas del reloj (se dice sentido dextrógiro) y, el segundo, en el sentido opuesto (se dice sentido levógiro).

Si ponemos estas dos representaciones una al lado de la otra, observaremos que una es la imagen especular de la otra, es decir la imagen que veríamos si pusiéramos el modelo molecular delante de un espejo.

Para pasar de un modelo al otro, necesariamente, hay que romper enlaces y volver a formarlos; por lo tanto, son isómeros, en particular se los llama enantiómeros. Sólo los carbonos que tienen cuatros sustituyentes distintos pueden presentar esta característica: se los llama carbonos quirales. Los carbonos con hibridación sp^3 que no tienen esta característica, es decir que tienen hasta dos sustituyentes distintos, se los llama carbonos aquirales. En estos casos, la imagen especular puede hacerse coincidir con la molécula que le dio origen por una simple rotación.

Si el carbono presenta una secuencia dextrógira se lo llama R, si por el contrario es levógira es S.

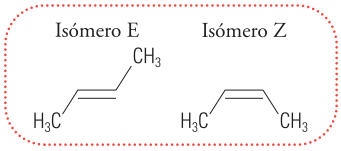


Configuración R Configuración S

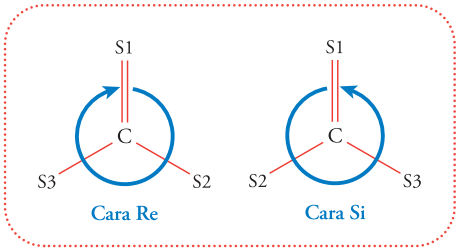
En 1966 tres investigadores, Robert Sidney Cahn, Christopher Kelk Ingold y Vladimir Prelog, publicaron un trabajo[3] en el que establecían la reglas para determinar, sin ambigüedades, la configuración de un carbono quiral. Las reglas CIP (por las iniciales de los autores) establecen la asignación de prioridades a los sustituyentes con un simple criterio. Se le asignará prioridad 1 al sustituyente de mayor número atómico, o sea de mayor Z, se le asignará prioridad 2 al sustituyente que tenga un Z, inmediatamente menor que el Z del sustituyente con prioridad 1. Así, siguiendo hasta asignar la prioridad 4. Si hay empate, se consideran los elementos a los que éstos están unidos. Luego observamos el carbono de la molécula poniendo el sustituyente con prioridad 4 detrás del carbono quiral y observamos en qué sentido gira el recorrido de mayor prioridad a menor prioridad (prioridad 1 → prioridad 2 → prioridad 3).

Por supuesto que hay casos especiales que no discutiremos aquí pero estas sencillas reglas nos permiten determinar la configuración de la gran mayoría de los carbonos quirales o asimétricos.

Las reglas CIP también establecen criterios para determinar si un doble enlace es un isómero *cis* (Z) o *trans* (E) (al considerar los dos sustituyentes de mayor prioridad en cada carbono).



Por último, con estas reglas, es posible distinguir los lados o caras de un carbono sp^2 si los sustituyentes unidos por los enlaces simples son distintos. Se dice que la cara es "cara Re" si, luego de asignar las prioridades, la secuencia gira en el sentido de las agujas del reloj (dextrógira), si la secuencia gira en el sentido contrario al de las agujas del reloj (levógira) se dice que la cara es la "cara Si". Es inmediato que, una vez determinado el tipo de cara que estamos observando, la otra cara tiene la configuración opuesta.



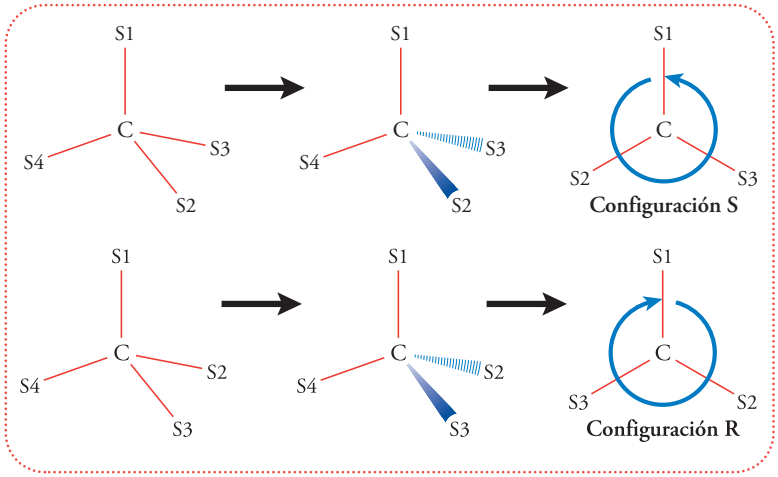
Cara Re Cara Si

[3] "Specification of Molecular Chirality". Angewandte Chemie International Edition 5 (4): 385–415.

Representación en el plano de un carbono asimétrico

Para poder representar correctamente un carbono asimétrico debemos adicionar dos símbolos más para representar enlaces simples. Se trata de la representación por medio de cuñas, llena y guionada. Ambas cuñas indican con su lado ancho que se apartan del plano que determinan el carbono central y los otros dos enlaces (representados por líneas de trazo uniforme). Se conviene que la cuña llena se acerca hacia el observador, es decir, sobresale del plano del papel, y que la cuña guionada se aleja del observador, es decir penetra el plano del papel.

Con todo esto es posible armar correctamente una cantidad sumamente grande de moléculas. Pero **¿cómo lo codificamos?**



Codificación o cómo escribimos todo esto

Como hemos visto, la forma de conectar cada sustituyente a cada carbono varía según el caso de hibridación que presenten y la orientación que les demos a unos con respecto a los otros. Pero una vez determinado todo lo anterior, es necesario escribirlo de forma tal que los programas dedicados a representar modelos moleculares lo entiendan. A continuación presentaremos en forma general en qué consiste codificar un modelo molecular tridimensional ya que cada programa tiene sus propias variantes.

Marco de referencia

Antes de comenzar a representar una molécula en el espacio, es necesario describir este espacio. **¿Cómo se hace esto?** Es bastante sencillo pero debemos comenzar por establecer un marco de referencia, éste consiste en un trío de ejes perpendiculares (**Figura 26**) que nos permitirá describir la posición de un punto, muy fácilmente.

Veamos.

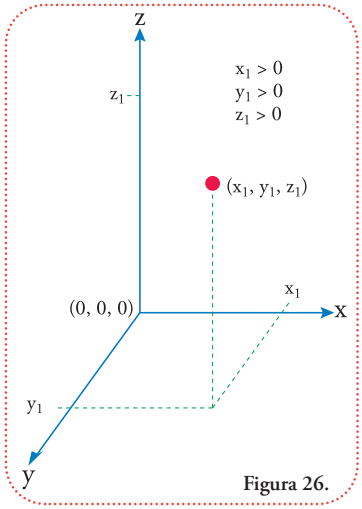


Figura 26.

Cada eje es una recta que se interseca con las otras dos. Tomadas de a pares siempre forman un ángulo de 90°.

Los valores que figuran entre paréntesis -x_1, y_1, z_1- se denominan coordenadas espaciales. Son números reales que repre-

senta la posición del punto con respecto a cada eje. En general no tienen unidades los ejes. Es muy habitual que los valores de las coordenadas se interpreten en Ångstroms (Å) cuya equivalencia con el metro y dos de sus submúltiplos (el nanómetro y el picómetro) es la siguiente:

$$
\begin{aligned}
1\ \text{Å} &= 1 * 10^{-10}\ \text{m} \\
&= 0{,}1 * 10^{-9}\ \text{m} \\
&= 0{,}1\ \text{nm (una décima de nanómetro)} \\
&= 100 * 10^{-12}\ \text{m} \\
&= 100\ \text{pm (cien picómetros)}
\end{aligned}
$$

En el caso que observamos en la **figura 26**, las tres coordenadas son positivas.

Observemos que el punto, a partir de donde comenzamos a medir la posición, es el 0 (cero) para los tres ejes, por lo tanto, lo anotamos así "(0, 0, 0)", a este punto se lo llama origen de coordenadas.

Analicemos el efecto de modificar el valor de las coordenadas. Si tuviéramos un caso en el cual la coordenada en el eje x fuera menor que cero (**Figura 27**) y las otras dos no se modificaran, tendríamos que el punto ahora se encuentra a izquierda del origen de coordenadas del eje x.

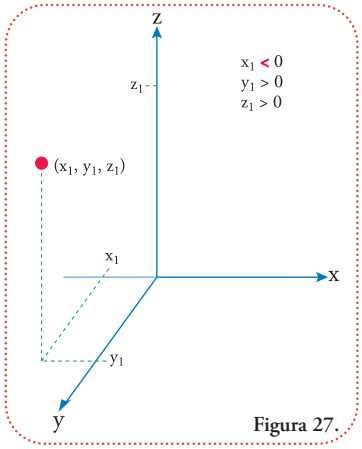


Figura 27.

Si volviéramos a la situación inicial y pusiéramos la coordenada en el eje y negativa (**Figura 28**) obtenemos:

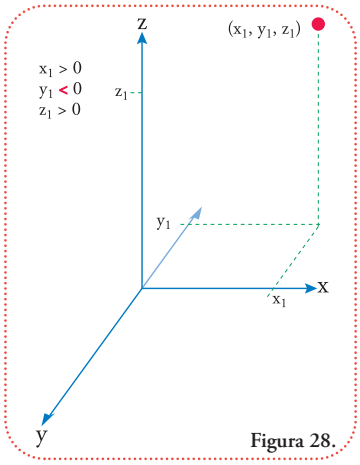


Figura 28.

Y si lo hacemos para el eje z tenemos lo siguiente (**Figura 29**).

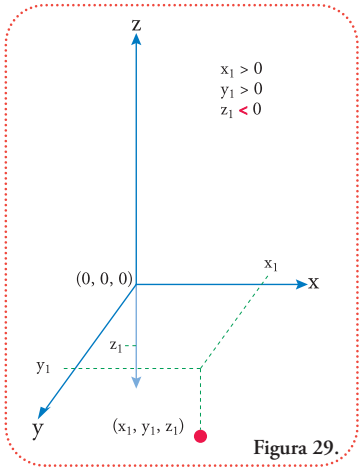


Figura 29.

Cada punto que agreguemos a este marco de referencia corresponderá a un núcleo atómico. Todos los puntos incorporados estarán a la distancia que corresponda a la suma de sus radios covalentes (si están ligados) o a una distancia mayor o igual a la suma de los radios de Van der Waals (si no estuvieran ligados directamente).

El resultado de posicionar todos los núcleos es una nube de puntos, cada uno de los cuales tiene su terna de coordenadas (x_i, y_i, z_i, donde i representa el número de identificación del átomo).
Para completar el objeto tridimensional hace falta explicitar la conectividad de la nube de puntos, es decir proveer la información que establece cuál átomo se conecta con cuál.

Una vez provista esta información tenemos codificado todo lo necesario para transmitir nuestro modelo molecular en forma de texto y, por lo tanto, en forma de ceros y unos (código binario) que es el código que utilizan los programas para comunicarse y ejecutar una tarea. Todo esto se agrupa en un objeto que se llama archivo.

En el próximo capítulo podremos trabajar con archivos reales de modelos reales.

Actividad

Para fijar estos conceptos respondamos las siguientes preguntas sobre un modelo molecular arbitrario de cuatro puntos (A, B, C y D). Obsérvelo detenidamente y responda las consignas, las repuestas correctas las encontrará al final del libro.

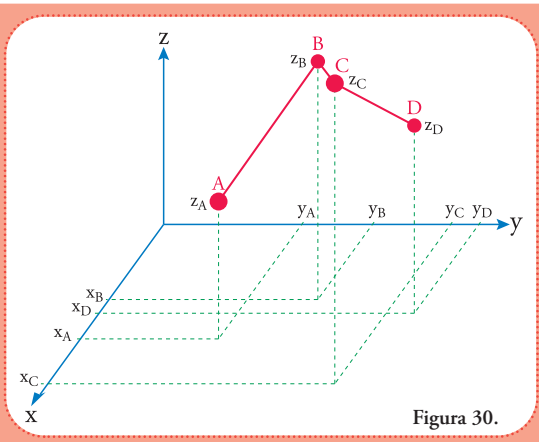


Figura 30.

Consigna I

Observe la **figura 30** e indique si cada una de las siguientes proposiciones son verdaderas (V) o falsas (F).

	V	F
1. $x_A > x_B$	☐	☐
2. $y_A > y_D$	☐	☐
3. $z_A < z_C$	☐	☐
4. $z_B < z_D$	☐	☐
5. $z_B > z_C$	☐	☐
6. A se conecta directamente con B	☐	☐
7. El tercer vecino de A es D	☐	☐
8. El tercer vecino de D es B	☐	☐
9. Hay dos líneas (enlaces) entre A y C	☐	☐
10. B, C y D definen un plano que contiene a A	☐	☐
11. La distancia entre A y B es menor que entre C y D	☐	☐
12. Existiría alguna forma de rotar esta representación tal que se vieran tres puntos eclipsados y uno no.	☐	☐

Consigna II

Dibuje una molécula que concuerde con el esquema de la **figura 30**, bajo la siguiente convención: los puntos A, B, C, D pueden ser cualquier elemento y los enlaces pueden ser simples (dobles o triples). No importa la geometría que presenta el esquema.

Consigna III

Suponga que debe inventar un código para describir este objeto en el espacio y comunicarlo por correo electrónico a un amigo (que tiene el mismo sistema de referencia), de tal forma que pueda reproducir el esquema de la figura por sí solo. Usted decide establecer las siguientes pautas de trabajo y comunicárselas.

1) Se definirá una línea (oración) para cada punto, la cual comenzará nombrándolo y luego

le seguirán su coordenada x (valor con respecto al origen del eje x), luego coordenada en y y, finalmente, su coordenada en z.

2) Luego se escribirán líneas que comiencen con la palabra "CONECTAR" y luego el nombre de los dos puntos que deben ser conectados.

Éste sería el resultado:

A x_A y_A z_A
B x_B y_B z_B
C x_C y_C z_C
D x_D y_D z_D
CONECTAR A B
CONECTAR B C
CONECTAR C D

a) Dibujar a escala y en hoja cuadriculada (para ayudarse mejor en la perspectiva) el siguiente objeto definido con la misma convención y utilizar el marco de referencia del esquema:

A 2 3 3
B 2 3 5
C 2 4 4
CONECTAR A B
CONECTAR A C

b) ¿Cómo cambiaría el dibujo anterior si el código fuera el siguiente?

A 3 3 3
B 2 3 5
C 2 4 5
CONECTAR A B
CONECTAR C B

Consigna IV

¿Podría completar el código del objeto que se muestra en la **figura 31**?

A _ _ _
B _ 6 _
C _ _ _
CONECTAR _ _
CONECTAR _ C

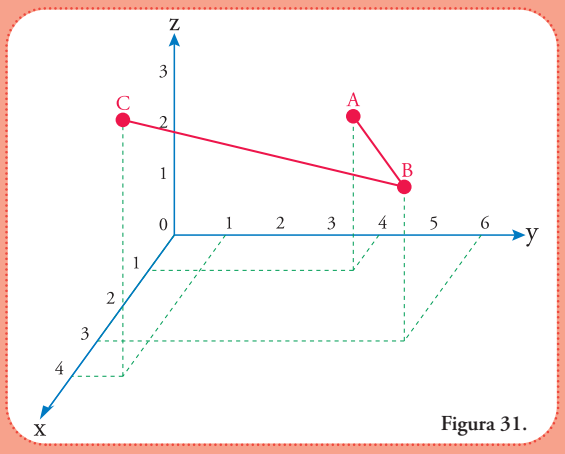


Figura 31.

Si ha logrado responder todas las consignas puede ver las respuestas para chequearlas.

Generación de modelos moleculares y su observación

Tenemos a disposición herramientas que nos permiten dibujar las moléculas en dos dimensiones y nos generan un modelo en tres dimensiones, el cuál es posible observarlo en un programa de visualización (**Figura 32**).

La primera etapa la llevamos a cabo con todos los conceptos que hemos aprendido en los capítulos anteriores.

La etapa de generación de coordenadas x, y, z y su conectividad es sumamente compleja para escribir el archivo nosotros mismos, aunque es posible. Aquí entran en juego los programas dedicados a tal fin. Estos programas no hacen otra cosa que seguir la receta que presentamos al comienzo del capítulo.

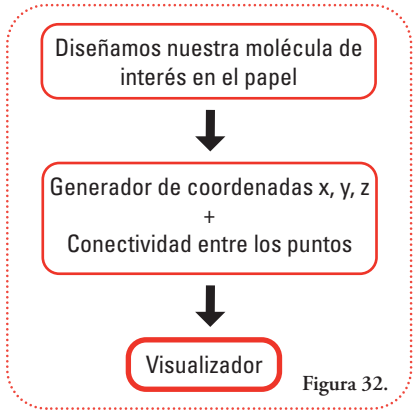


Figura 32.

1. Poseen una base de datos que le permite asignar a cada átomo su radio covalente y determinar la longitud de los enlaces.
2. Una base de datos de configuraciones de carbonos, para determinar los ángulos de los enlaces.
3. Codificación de las reglas del sistema de Lewis que nos impedirá equivocarnos en la conectividad de los átomos.
4. Están incorporadas las reglas CIP para poder determinar la configuración de los carbonos quirales a partir de un dibujo en dos dimensiones.

Finalmente, los programas visualizadores permiten interpretar los archivos generados por los programas anteriores para ofrecernos una visualización completa de la molécula. En general, los programas dedicados a la visualización suelen tener funciones que permiten hacer mediciones entre los átomos del modelo.

Los programas de generación de coordenadas suelen tener funciones de visualización del modelo terminado, pero pueden no tener las herramientas necesarias para profundizar en los detalles de la estructura.

Por último, suele haber estructuras demasiado complejas, con muchos átomos que suelen ser representadas con un detalle menor en cuanto el número de átomos, pero que permite entender la molécula a un nivel estructural superior. Este suele ser el caso de las proteínas, los ácidos nucleicos y los azúcares. De este último grupo de moléculas no generaremos coordenadas espaciales, sólo observaremos su estructura.

Generación

Comenzaremos por generar un modelo de una molécula sencilla en ACD/ChemSketch (Freeware), disponible en http://www.acdlabs.com/download/, y observaremos el modelo tridimensional que nos ofrece.

Trabajaremos con una molécula que tenga todos los tipos de carbonos y varios elementos, aunque ignoremos su existencia real, nos interesa observar el proceso de creación de sus coordenadas y su modelo tridimensional.

Ésta es nuestra molécula (**Figura 33**),

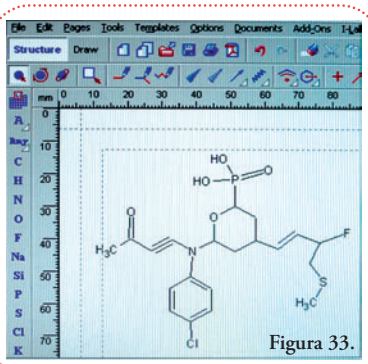


Figura 33.

Podemos observar, en la parte inferior de la pantalla, la fórmula molecular y, al lado, la masa molar (FW quiere decir en inglés: "Formula Weigth" que significa "peso de la fórmula") (**Figura 34**).

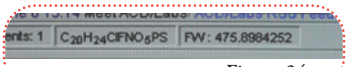
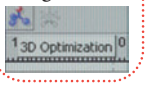


Figura 34.

Para observar el modelo molecular tridimensional apretamos el botón que lo indica (**Figura 35**):

Figura 35.



Este botón ("3D Optimization", que quiere decir Optimización en tres dimensiones) asignará, a cada carbono, sus propiedades estructurales. En el caso de la quiralidad aplicará por defecto un criterio que podremos cambiar si lo deseamos. Este es el resultado que veremos (**Figura 36**).

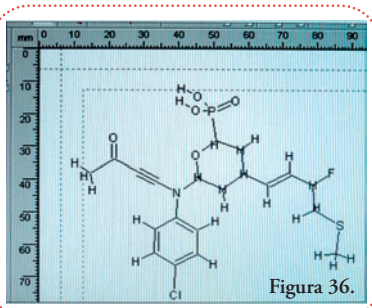
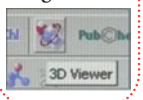


Figura 36.

Los hidrógenos ahora son explícitos y, en esta orientación, los vemos superpuestos a algunos enlaces. Para ver en detalle la estructura, apretamos el botón ("3D Viewer" que quiere decir Visualizador en 3D) (**Figura 37**).

Figura 37.



Ejecutará la función de visualización en tres dimensiones y nos dará la oportunidad de girar la molécula como deseemos a fin de observar sus distintos detalles, desde cualquier ángulo.

Esta es la primera imagen que obtenemos, básicamente la misma, pero en colores y sin letras. Cada color representa un elemento que se puede comprender al observar la estructura en dos dimensiones (**Figura 38**).

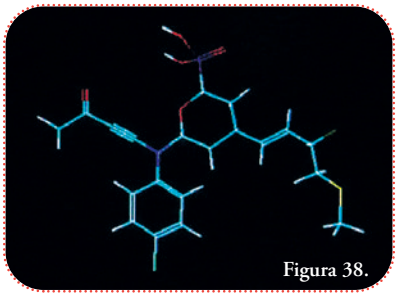


Figura 38.

Conviene ver en detalle algunos botones en la barra superior (**Figura 39**).

Figura 39.

El botón que se encuentra seleccionado indica que la representación que estamos viendo es la denominada representación en alambre ("Wireframe") que produce (**Figura 39**).

Luego está la representación de barras ("Sticks") (**Figura 40**)

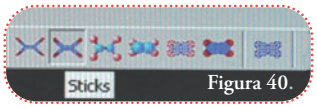


Figura 40.

Esta es la representación que produce seleccionar este botón (**Figura 41**).

Se pueden apreciar mejor los distintos elementos y las formas de los enlaces.

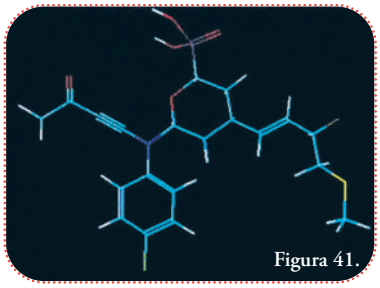


Figura 41.

A continuación activaremos el botón que se encuentra a la derecha, el que corresponde a la representación de bolas y barras ("Balls and Sticks") (**Figura 42**).

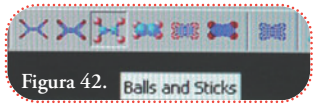


Figura 42.

Produce una representación similar a la anterior y le agrega esferas a los elementos de radio menor, pero proporcional, al radio covalente de cada elemento (**Figura 43**).

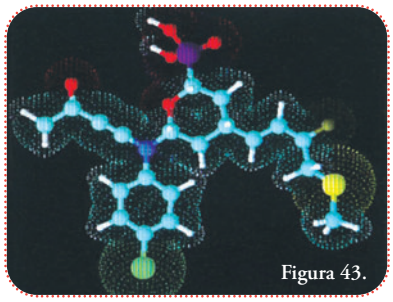


Figura 43.

Figura 44.

Continuamos con el botón de esferas sólidas de radio de Van der Waal ("Spacefill") (**Figura 44**).

Claramente se observa cómo las esferas se superponen unas con otras, sólo en el caso de que se trate de átomos unidos directamente (**Figura 45**).

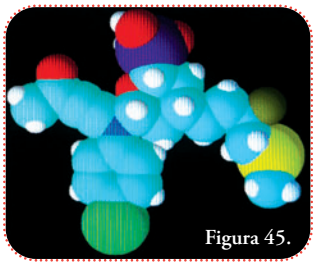


Figura 45.

Una variante de la representación que tiene en cuenta el radio de Van der Waals, es el botón que representa las esferas, no lisas, sino como una superficie de puntos ("Dots Only") y como discos ("Disks") (**Figuras 46 a 49**).

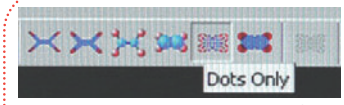


Figura 46.

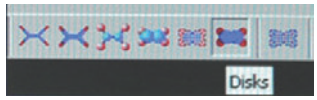


Figura 47.

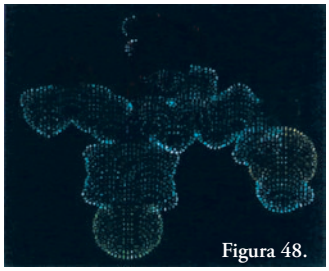


Figura 48.

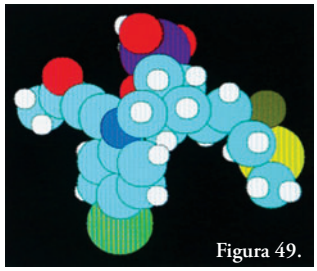


Figura 49.

Es posible combinar la representación de alambre, barras y bolas y barras con la representación de puntos para mejorar la visualización de la forma molecular y la superficie determinada por las esferas de radios de Van der Waals de cada elemento. Debemos seleccionar el botón "con puntos" ("with Dots") (**Figura 50**).

Mostramos, sólo, el resultado de combinar la represtación de bolas y barras con puntos (**Figura 51**).

Figura 50.

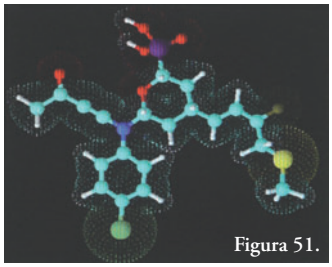


Figura 51.

Esta representación es bastante completa y la usaremos para observar este modelo desde distintos ángulos (**Figuras 52 a 54**):

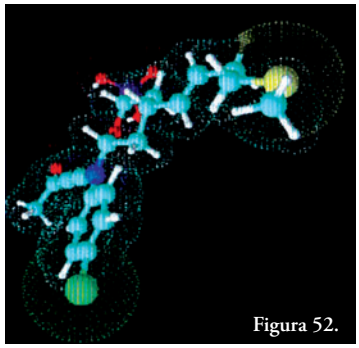


Figura 52.

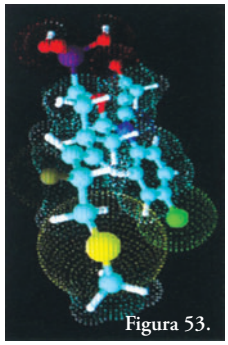


Figura 53.

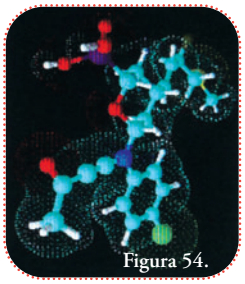


Figura 54.

Más adelante observaremos en forma detallada el formato MDL molfile. Por ahora continuaremos con otro programa generador de coordenadas y conectividad.

PRODRG

Este programa es un servicio en línea de acceso gratuito, se ejecuta desde el sitio http://davapc1.bioch.dundee.ac.uk/prodrg/[4] (está en idioma inglés) y fue desarrollado, especialmente, para generar modelos de coordenadas tridimensionales y archivos asociados (que describen con mayor detalle la molécula en términos de cargas parciales, tensión de los enlaces, etc.). Esta información, contenida en estos últimos archivos, es necesaria para realizar estudios sumamente interesantes de movimiento molecular; esta rama de la química, fusionada a la tecnología de la computación, se llama dinámica molecular. Justamente, se encar-ga de describir el movimien- to molecular en relación a las interacciones intermoleculares. Dependiendo de la complejidad del sistema es posible lograr resultados que predicen interacciones y ayudar en el diseño molecular.

Por ejemplo, en el diseño de fármacos inhibidores de enzimas.

La página de inicio es la que se visualiza en la figura 55.

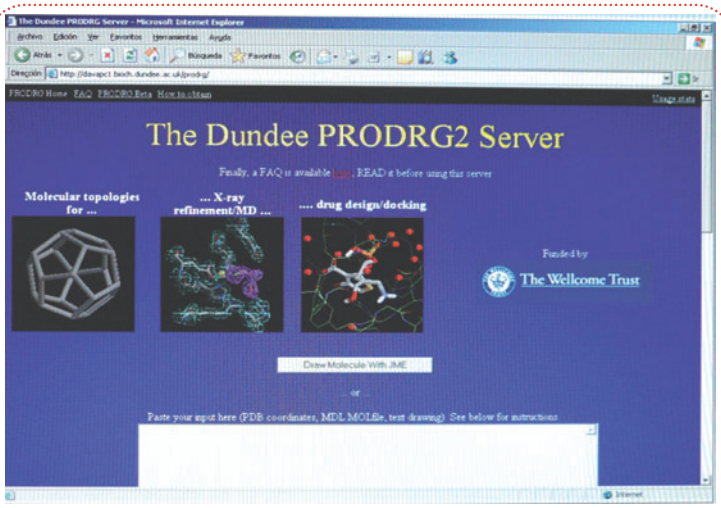


Figura 55.

La información para la generación del modelo molecular es un texto que dibuja una suerte de modelo en dos dimensiones.

Tiene **reglas muy sencillas de construcción**:

1. los átomos se nombran con letras mayúsculas;
2. los signos que se utilizan para indicar los enlaces son:
   a. simples: - (si es horizontal), | (si es vertical),
   b. dobles: = (si es horizontal), “ (si es vertical),
   c. triples: # (en cualquier caso);
3. no se permiten conexiones en diagonal;
4. la longitud de los enlaces no está restringida, se extienden por agregación del signo correspondiente;
5. la quiralidad se puede invertir usando la tipografía minúscula;

[4] A. W. Schuettelkopf and D. M. F. van Aalten (2004). PRODRG - a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallographica D60, 1355--1363.

6. sólo son interpretados los átomos de nitrógeno (N), carbono (C), oxígeno (O), azufre (S), fósforo (P), cloro (Cl), yodo (I), bromo (Br), flúor (F), a este conjunto de átomos se lo suele llamar conjunto de átomos orgánicos;
7. no es necesario explicitar los hidrógenos ya que el programa los completará automáticamente.

Por ejemplo, podemos dibujar, directamente, en la ventana el ciclo hexano, así:

C-C-C-C-C-C
|- - - - - - - -|

También, se puede hacer en un procesador de texto y luego "cortar y pegar" el texto en la ventana en blanco que se presenta en la primera página del sitio.

Podemos dibujar nuestra molécula del apartado anterior, pero como hacerlo manualmente puede ser un poco complicado, es posible generarlo a partir de una ventana de dibujo, similar al ACD/ChemSketch (Freeware), seleccionando el botón ("Draw Molecule With JME", que quiere decir Dibujar la Molécula Con JME) (**Figura 56**):

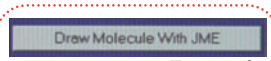


Figura 56.

Abre la siguiente ventana que tiene todas las herramientas que utilizamos en el ACD/ChemSketch (Freeware) para dibujar nuestra molécula (**Figura 57**).

Primero dibujamos toda la molécula con carbonos, usamos la plantilla para un anillo fenólico, la del ciclohexano y la de enlaces simples para construir el esqueleto (**Figuras 58 a 60**).

Figura 58.

Figura 59.

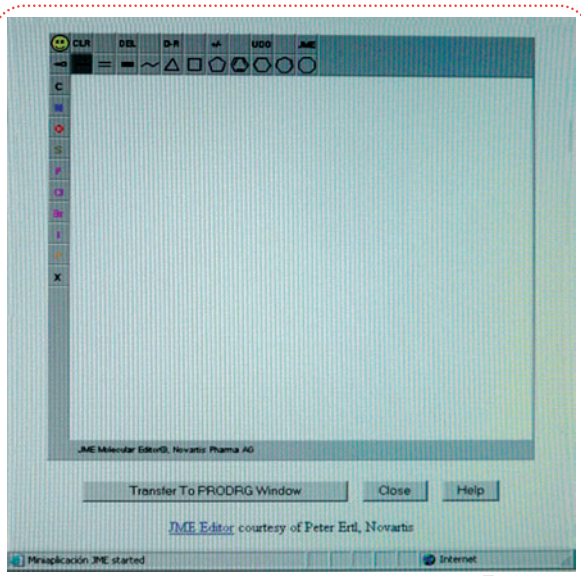


Figura 57.

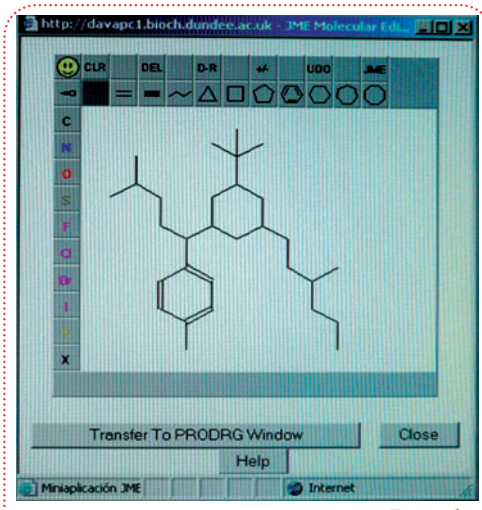


Figura 60.

A este modelo le agregaremos los dobles y triples enlaces que sean necesarios con los botones (**Figura 61**).

Figura 61.

Si lo quisiéramos, podríamos observar la quiralidad que le asignó el

ACD/ChemSketch ("Freeware") para repetirla agregando la representación en cuña con el botón correspondiente (**Figura 62**). Si la seleccionamos una vez la cuña será llena, si lo hacemos dos veces la cuña será guionada (**Figuras 63 y 64**):

Figura 62.

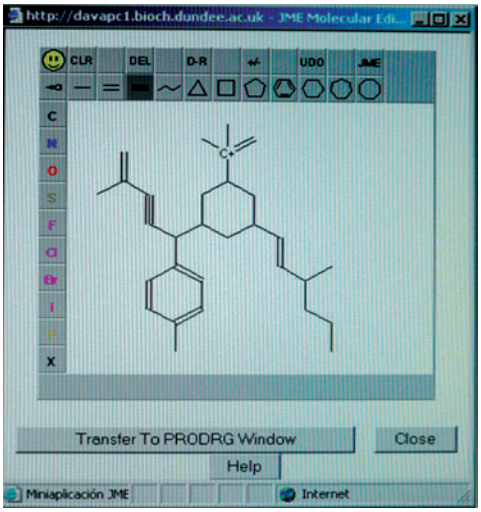


Figura 63.

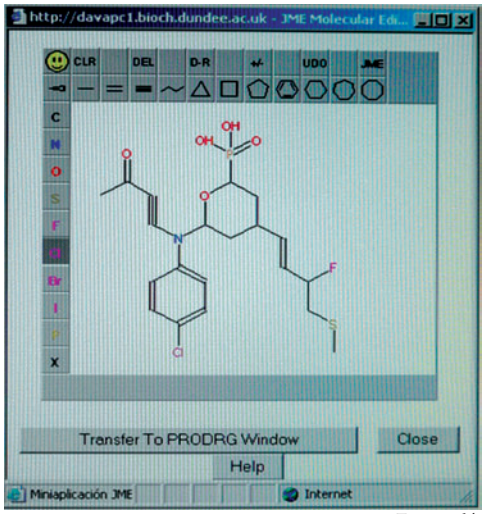


Figura 64.

Con la barra de elementos que se ubican a la izquierda, podemos reemplazar los carbonos por los elementos correspondientes en:

Al seleccionar el botón (**Figura 65**).

Figura 65.

Que indica "Transfer To PRODRG Window", que quiere decir "transferir a la ventana de PRODRG", obtenemos (**Figura 66**):

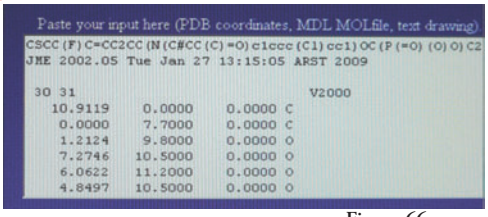


Figura 66.

En el primer renglón observamos la molécula codificada en SMILES, que veremos en el siguiente apartado. El segundo renglón indica la fecha y la hora en que fue realizado el modelo, luego un renglón en blanco.

A continuación se indican la cantidad de átomos de la molécula, sin contar los hidrógenos (30) y la cantidad de enlaces (31). V2.000 indica el modelo de código MDL molfile. Recordemos que este código de representación de modelos moleculares en tres dimensiones provee primero las posiciones y luego las conexiones.

A continuación se provee la coordenada del primer átomo, un carbono; en la siguiente línea están las coordenadas del siguiente átomo, que es otro carbono y, así, se completan las coordenadas de todos los átomos que forman la molécula en 30 renglones. Notemos que la tercera coordenada es 0,0000 para todos los elementos, es decir que todos están en el mismo plano, lo cual es correcto ya que proviene de una representación en dos dimensiones o sea plana.

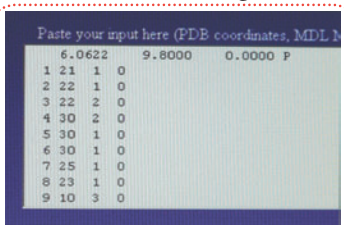


Figura 67.

Luego viene la descripción de los enlaces (**Figura 67**).

Luego de la última línea de coordenadas espaciales, que corresponde al fósforo, se describen la conectividad de la molécula.

El primer enlace descripto es el que une al átomo 1 (o sea el elemento cuyas coordenadas están en el primer renglón del conjunto de líneas de coordenadas) con el átomo 21 por medio de un enlace simple, indicado por el número 1 en la tercera columna. El cero de la cuarta columna indica que es un lugar disponible, pero no utilizado para proporcionar información adicional sobre el enlace.

Si el enlace es doble, el número de la tercera columna es 2, como en el enlace entre los átomos 3 y 22; si el enlace es triple, como entre los átomos 9 y 10, el número de la tercera columna es 3.

El modelo intermedio, el que aparece en la ventana del PRODRG está listo para generar el modelo final apretando el botón que dice "Run PRODRG" (**Figura 68**). Indica ejecutar el programa PRODRG. En segundos, este programa nos genera el modelo molecular en los formatos más importantes y necesarios para distintos usos.

Figura 68.

La primera parte de la pantalla nos presenta las acciones que realizó y nos avisa que finalizó correctamente; a nuestra izquierda nos muestra una representación tridimensional del modelo que produjo con el nombre de cada átomo de acuerdo a un código interno con el cual nombra los átomos en los archivos que generó (**Figuras 69 y 70**).

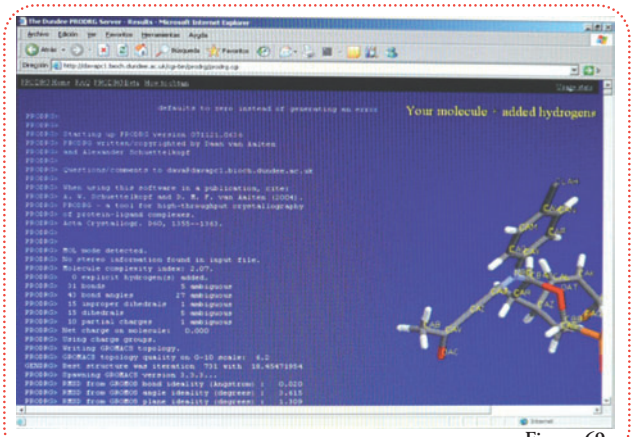


Figura 69.

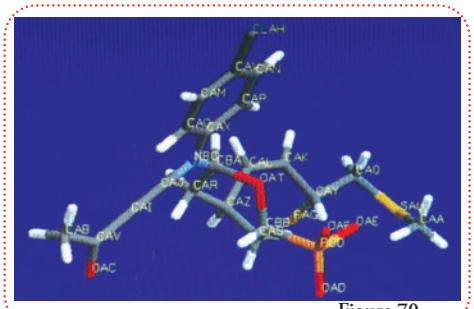


Figura 70.

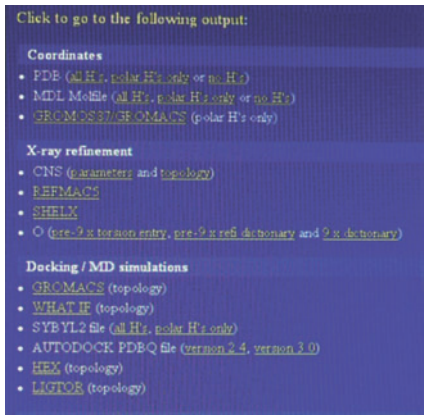


Figura 71.

A continuación vemos el menú de los archivos que produjo (**Figura 71**).

Las tres categorías de archivos son:

1. coordenadas (en inglés: “coordinates”): agrupa todos los archivos útiles para los visualizadores que permitirán describir mejor la conformación molecular.

2. refinamiento de datos de difracción de rayos – X (en inglés: "X-ray refinement"): agrupa todos los archivos que son útiles para mejorar los modelos moleculares generados con datos experimentales de difracción de rayos – X.

3. simulación de acople molecular y dinámica molecular (en inglés: "docking y MD simulations"): agrupa todos los archivos que son útiles para estudiar los movimientos moleculares y las interacciones entre moléculas de acuerdo a su forma pero, también, teniendo en cuenta las cargas de los átomos, la tensión de los enlaces, etc.

Guardaremos las coordenadas de todos los átomos (incluyendo los hidrógenos) en formato MDL molfile; para hacerlo, primero apretando el botón izquierdo del ratón, dentro de la primera categoría, la opción "all H´s" (que indica que contiene todos los hidrógenos) (**Figura 72**).

Figura 72.

Nos aparece (**Figura 73**).

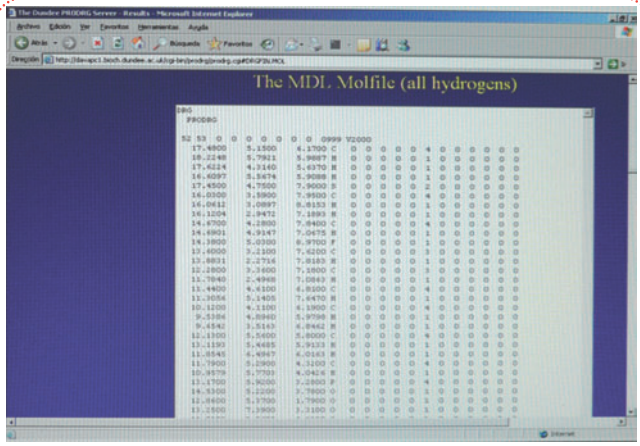


Figura 73.

Notemos que ahora la cantidad de átomos y enlaces aumentó de 30 a 52 y de 31 a 53, respectivamente (en la cuarta línea).

A continuación transferiremos esta información a un archivo de la siguiente manera:

1. seleccionamos todo el texto, pulsando el botón izquierdo del ratón dentro del cuatro de texto y, luego, CRTL + E (tecla "CONTROL" más tecla de la letra "E"). Esto produce que todo el texto esté listo para ser copiado (**Figura 74**);

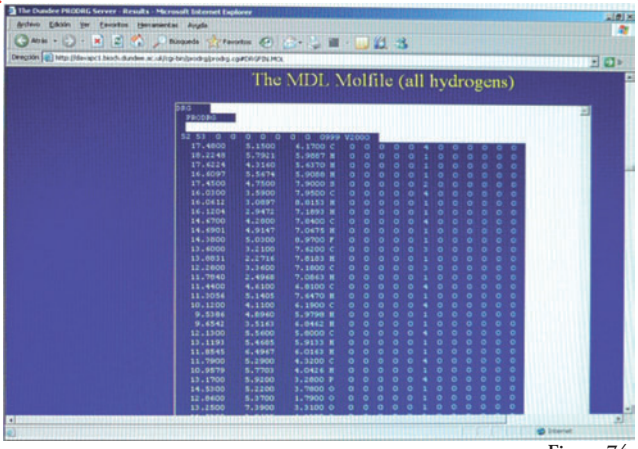


Figura 74.

2. apretamos CTRL + C. Esto produce que el texto quede en la memoria de la computadora listo para ser reescrito (apretando CTRL + V) en cualquier programa que maneje textos y luego guardarlo con un nombre que deseemos seguido de la extensión ".mol";

3. usamos el block de notas de Windows para reescribirlo y lo guardamos, entonces, con el nombre de "molécula.mol" (**Figuras 75 y 76**).

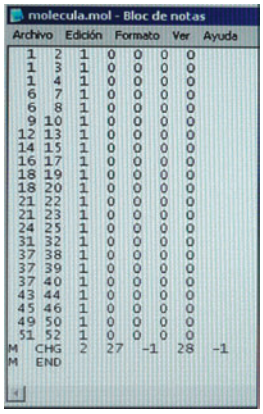


Figura 75.

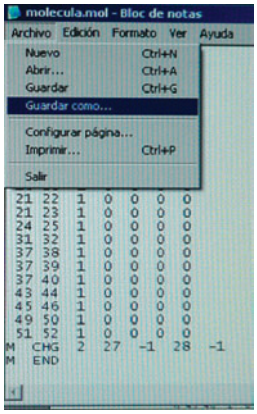


Figura 76.

las funciones para obtener y modificar un archivo .mol.

Pasemos a otro código de representación

Código SMILES

Como comentamos en el apartado anterior, muchos programas tienen extensas bases de datos que permiten obtener el modelo tridimensional de la molécula a partir de información en dos dimensiones.

Algunos necesitan aún menos información y pueden interpretar el código SMILES ("Simplified Molecular Input Line Entry Specification", que quiere decir Especificación Simplificada de Entrada Molecular en Línea de texto) el cual permite representar un gran número de moléculas, como una cadena de caracteres. Como se puede observar en este ejemplo (**Figura** 77) en el panel A tenemos la representación bidimensional de la molécula que deseamos pasar al código SMILES.

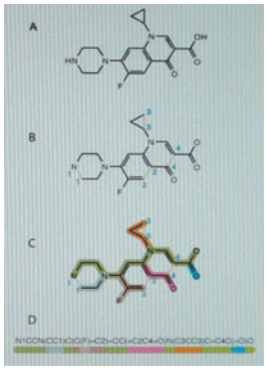


Figura 77.

El siguiente paso es cortar todos los anillos, de forma tal, que queden cadenas lineales ramificadas en otras cadenas lineales, a cada corte le corresponde un número.

En el panel C se encuentran coloreadas todas las cadenas lineales distinguibles. Observamos que en cada punto de bifurcación aparece un color nuevo.

En el panel D está el código SMILES de esta molécula.

La cadena de átomos más larga es la verde, por lo tanto, comenzamos por ésta a construir la codificación mostrada en el panel D.

Los programas que interpretan este código entienden que cada átomo se representa con el símbolo que le corresponde en la tabla periódica. Asumen que se unen por simples enlaces por defecto y se completa el octeto de cada átomo con la cantidad necesaria de hidrógenos. Si deseamos indicar un doble enlace, lo hacemos con el signo de igual (=), y si es triple lo hacemos con el signo de numeral (#).

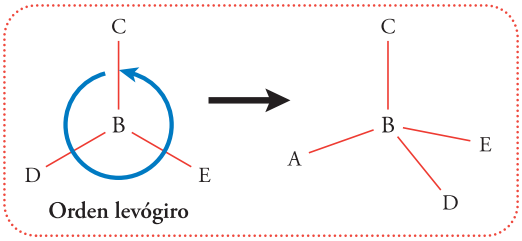
Como cada corte está numerado, a los átomos que deben unirse para reconstituir el ciclo se les agrega el número de corte correspondiente.

Cada rama, que aparece unida a la cadena principal, se representa dentro de un par de paréntesis, es decir, las cadenas van anidadas. Si bien este ejemplo no contiene centros asimétricos, ni dobles enlaces, el código SMILES permite describirlos.

La quiralidad de un elemento se señala con el signo arroba (@), simple o doble. Un signo indica giro en el sentido contrario al de las agujas del reloj (levógiro), dos signos indican giro en el sentido de las agujas del reloj (dextrógiro). Si hay hidrógenos que intervienen en la configuración se deben escribir.

Los centros asimétricos se identifican de la siguiente manera:

C[B@A](D)(E): aparecerán en sentido levógiro el primer sustituyente de B, es decir C, en orden de escritura, luego el segundo y el tercero (D y E figuran entre paréntesis ya que son dos bifurcaciones distintas a partir del punto B):



Como en esta molécula ejemplo NH_3CClI (la configuración del carbono es R) (**Figura 78**).

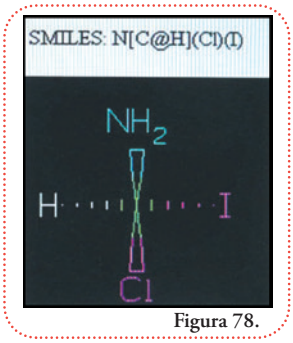
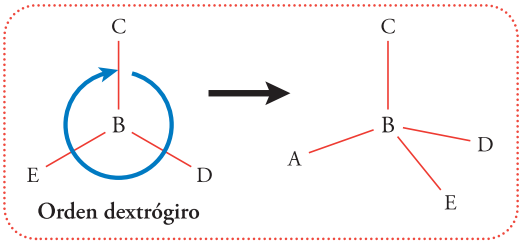


Figura 78.

C[B@@A](D)(E): los sustituyentes en orden de escritura, C, D y E aparecerán en sentido dextrógiro



Para nuestra molécula ejemplo NH_3CClI (la configuración del carbono resulta S) (**Figura 79**).

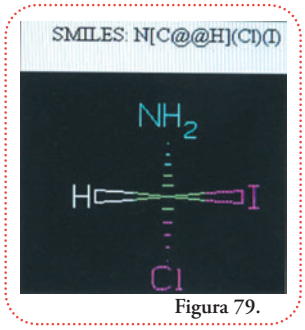
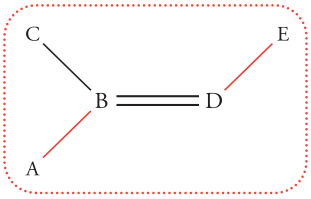


Figura 79.

NO HAY QUE INFERIR que un signo @ indica configuración R y @@ indican configuración S. La configuración (R o S) se determina con las reglas de CIP de priorización de sustituyentes de acuerdo al número atómico.

Los dobles enlaces son más sencillos de codificar. Se utilizan las barras / y \ para indicar si el doble enlace es *cis* o *trans*, considerando, sólo, los elementos unidos mediante las barras:



A/B(C)=D/E produce una representación en la que el sustituyente A y el E están enfrentados con respecto al doble enlace (notar que C figura entre paréntesis ya que es una ramificación en el átomo B).
Como en esta molécula ejemplo $C_2H_3BCl_2$ (configuración del doble enlace E *-trans-*, para los Cl –Z=17–) (**Figura 80**).

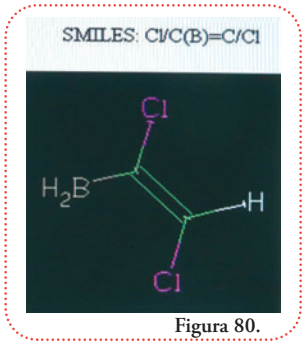
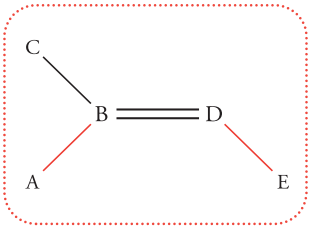


Figura 80.



A/B(C)=D\E produce una representación que el sustituyente A y el E están enfrentados con respecto al doble enlace.

En la molécula ejemplo $C_2H_3BCl_2$ (configuración del doble enlace Z *-cis-*, para los Cl –Z=17–) (**Figura 81**). Recordemos que los hidrógenos (H) se completan automáticamente y por lo tanto no se escriben. Nuevamente, NO HAY QUE CONFUNDIR el uso de / / o / \ con el hecho de que describan el isómero Z (*cis*, del mismo lado

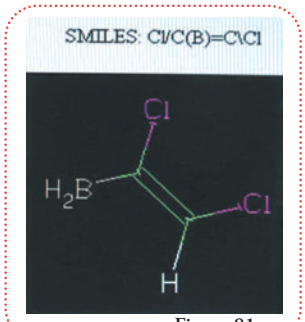


Figura 81.

los sustituyentes de mayor prioridad de cada carbono) o E (*trans*, de lados opuestos, bajo el mismo criterio). Que un doble enlace sea Z o E se determina aplicando las reglas CIP.

Pasemos ahora a describir la molécula que generamos para usar el PRODRG, siguiendo los pasos que vimos en los paneles A a D y agregando los criterios de descripción de carbonos asimétricos y dobles enlaces (**Figura 82**).

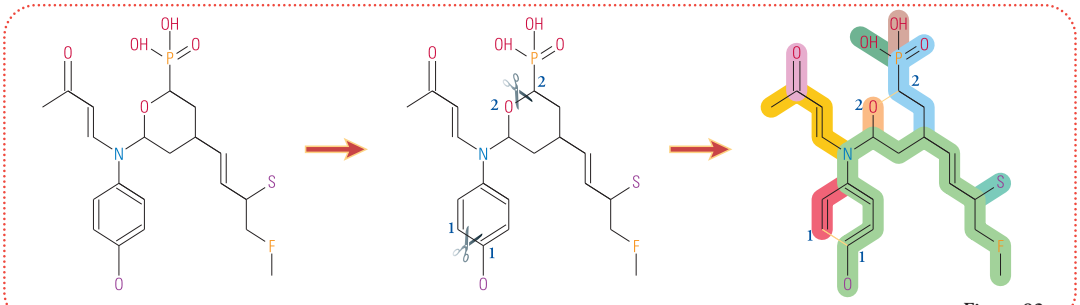


Figura 82.

Vamos a construir la codificación por partes y señalaremos la cadena con colores.

Iremos chequeando cada estructura parcial en el sitio: http://www.daylight.com/daycgi/depict

Es una única página que tiene una barra para ingresar el código y al seleccionar el botón "SUBMIT" (enviar) genera una imagen de la molécula (**Figura 83**).

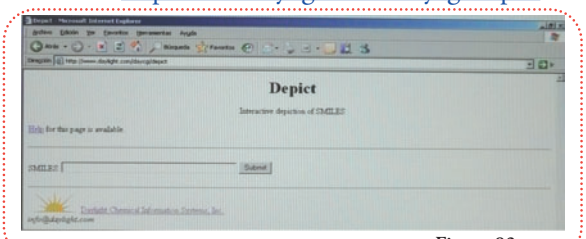


Figura 83.

Si la única molécula fuera la cadena verde (**Figura 84**).

SMILES: ClC=CC=CNCCC/C=C/CCSC

ClC = CC = CNCCC/C = C/CCSC

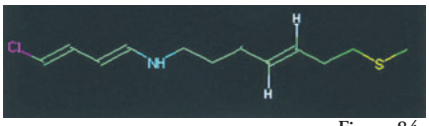


Figura 84.

Ahora agregamos las cadenas amarillas y fucsia (**Figura 86**).

SMILES: ClC1=CC=C(C=C1)N(C#CC(=O)C)CCC/C=C/CCSC

ClC1 = CC = C(C = C1)NCCC/C = C/CCSC

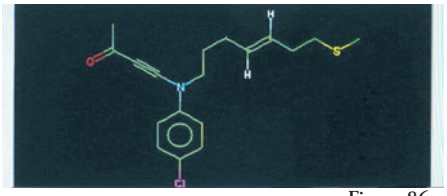


Figura 86.

Ahora agregamos la cadena roja, tenemos que indicar cuáles carbonos cierran el ciclo al que le hicimos el corte 1. Así obtenemos una molécula cerrada que podemos chequear (**Figura 85**).

ClC1 = CC = C(C = C1)N(C#CC(= O)C)CCC/C = C/CCSC

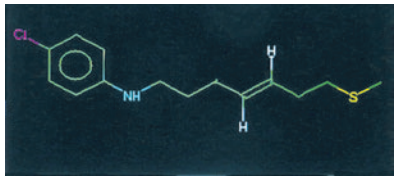


Figura 85.

A continuación agregamos las cadenas naranja (resaltado en gris), celeste, verde oscuro y marrón ya que forman el segundo ciclo. Se generan dos nuevos centros asimétricos que podemos explicitar si lo deseáramos. Supongamos que conocemos la configuración del segundo centro asimétrico tal

que se describiera con dos signos arroba (@@), (**Figura 87**).

Finalmente, agregamos la cadena celeste oscuro, que corresponde al flúor y que vamos a suponer que el centro asimétrico, que se genera, es perfectamente descriptible utilizando un solo signo arroba (@), (**Figura 88**).

Así logramos traducir nuestra modelo molecular al código SMILES. En rigor de verdad este código sólo puede almacenar la conectividad de una molécula, lo cual incluye sólo el tipo de átomo y los tipos de enlaces. El programa que interprete este código debe tener a disposición una base de datos de distancia de enlaces (radios covalentes) y ángulos predichos por la TRePEV.

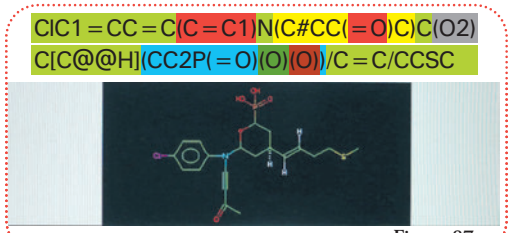


Figura 87.

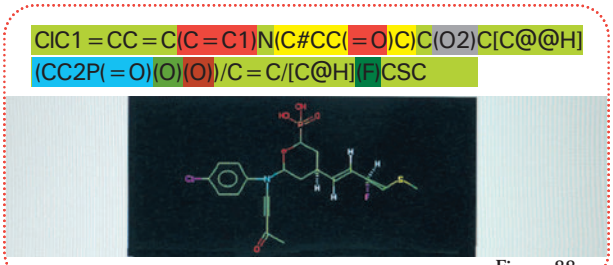


Figura 88.

Visualización

Hasta este momento sólo vimos como visualizador la función del ACD/ChemSketh (Freeware). Esta función es muy útil ya que nos ofrece la posibilidad de observar el modelo molecular desde cualquier ángulo y la superficie y el volumen molecular al representar a los átomos como esferas de radios de Van der Waals. Sin embargo, esta función no provee herramientas cuantitativas para medir distancias, ángulos o ángulos diedros, o seleccionar cierta parte de la molécula para estudiarla en detalle; éstas sí son funciones comunes a todos los programas dedicados a realizar visualizaciones de modelos moleculares. Además, otra característica muy importante, es que son capaces de presentar moléculas compuestas por muchos átomos (varios miles) y, a su vez, varias moléculas de este tamaño. También permiten cargar varios modelos y analizar si es posible superponerlos y cuantificar el grado de superposición, aunque éstas son funciones muy avanzadas para el enfoque de nuestro libro.

Deep View / Swiss PDB - Viewer

Este programa se puede descargar gratuitamente del sitio:

http://www.expasy.ch/spdbv/text/getpc.htm

En el capítulo siguiente veremos cómo usarlo en detalle para resolver un problema concreto en una molécula pequeña que consiste en verificar distancias y ángulos.

Aquí veremos las herramientas que ofrece para representar un tipo particular (aunque muy variado) de moléculas: las proteínas.

¿Qué son las proteínas?[5]

Esta pregunta tiene varias respuestas de acuerdo a la forma en que se la encare. Desde el punto de vista de la química orgánica, podemos afirmar que una molécula de proteína es un polímero lineal que resulta de la combinación de un grupo de veinte aminoácidos.

¿Qué es un polímero?

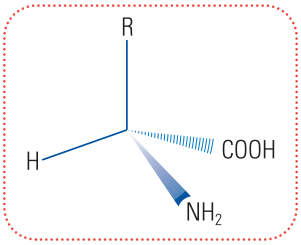
Un polímero es la unión sucesiva de partes menores (monómeros), más o menos parecidas,

[5] Puede encontrar en esta misma colección el libro Proteínas Estructuras Fascinantes del Dr. Javier Santos dedicado a estas moléculas.

en este caso los monómeros son los aminoácidos.

¿Qué son los aminoácidos?

Los aminoácidos son moléculas que tienen tanto el grupo funcional amino ($-NH_2$) como el grupo funcional ácido carboxílico (-COOH). No cualquier aminoácido puede formar parte de una proteína.

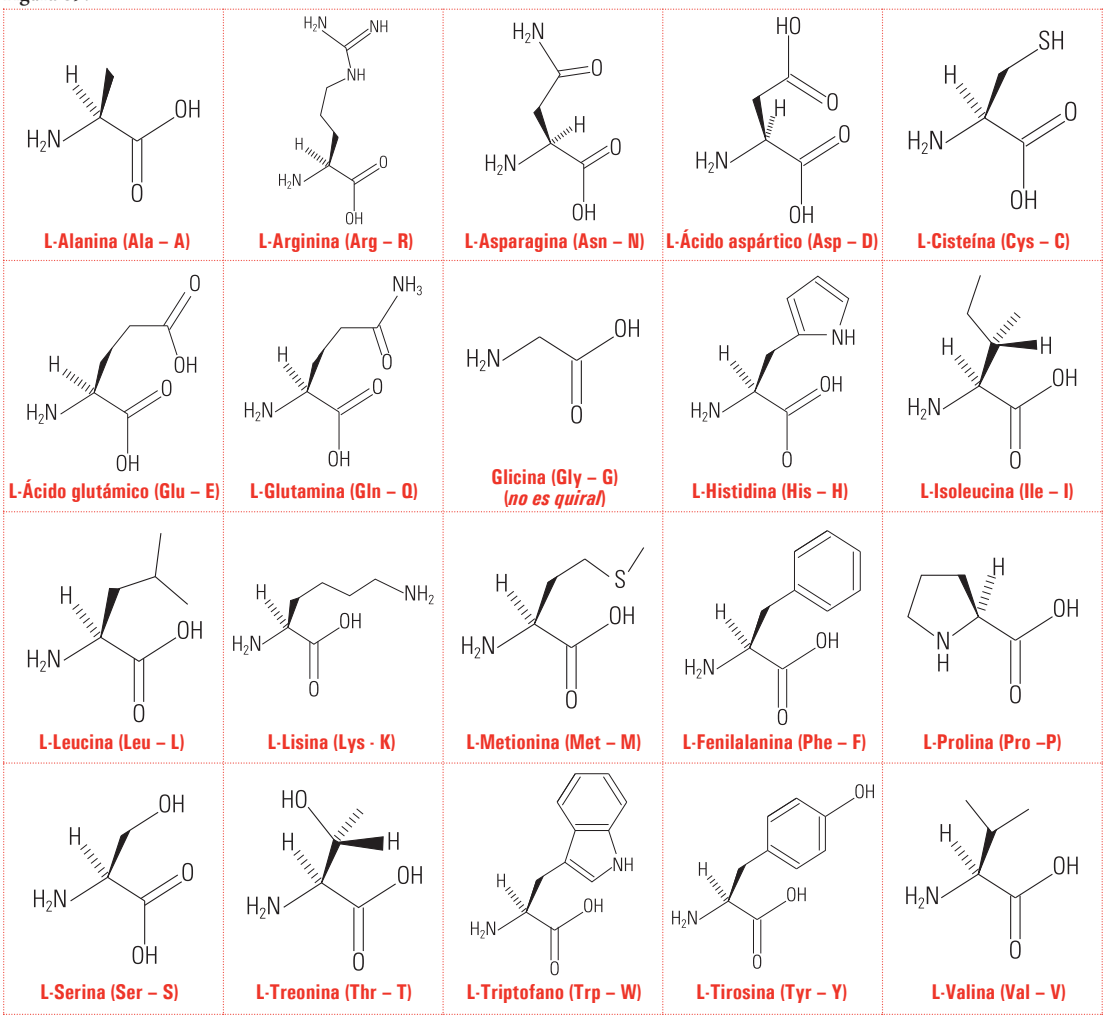


Los aminoácidos que están presentes en la gran mayoría de las proteínas pertenecen a un conjunto de (sólo) 20 moléculas[6] que comparten la siguiente estructura y quiralidad (se los llama L-aminoácidos) que se visualiza en el esquema.

R representa lo que se llama cadena lateral del aminoácido y lo distingue del resto de los otros 19.

A continuación listamos los veinte aminoácidos (entre paréntesis se mencionan el nombre en código de tres letras y de una letra) (**Figura 89**).

Figura 89.



[6] En ciertos organismos este conjunto es levemente alterado por la presencia en moléculas proteicas de alguno de los 20 aminoácidos citados modificado químicamente.

Los aminoácidos se conectan entre sí por una reacción que involucra el grupo funcional amino de un aminoácido y el grupo funcional ácido carboxílico del otro. El grupo funcional resultante es una amida (**Figura 90**), indicado como enlace peptídico.

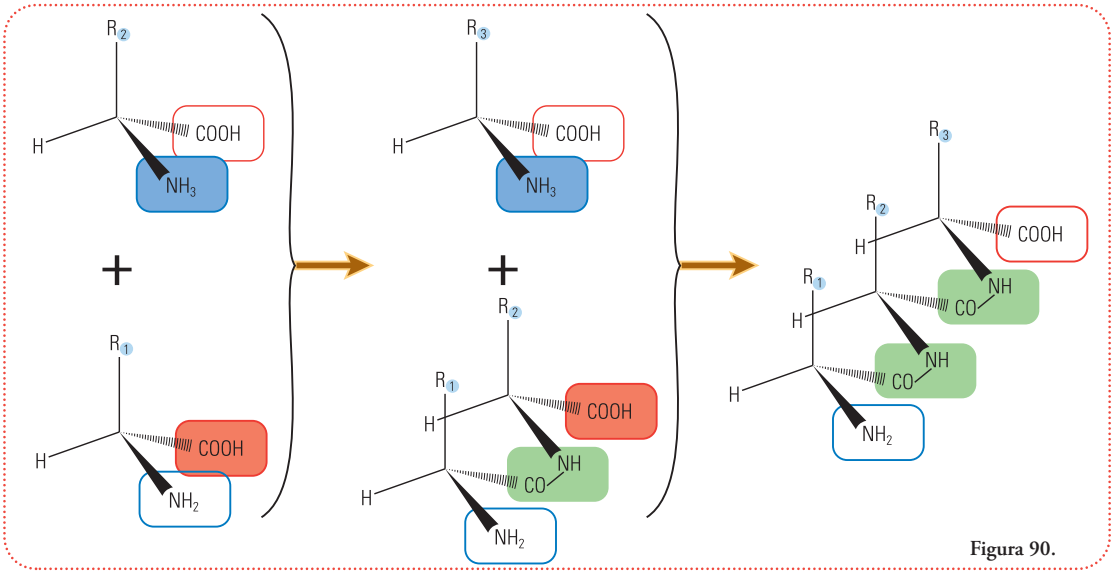


Figura 90.

Podemos ver que quedan sin reaccionar el grupo amino del primer aminoácido y el grupo carboxilo del último aminoácido. Estos dos grupos se nombran, respectivamente, amino terminal (N-t) y carboxilo terminal (C-t) y sirven para orientar el polímero de aminoácidos.

Se dice que una proteína comienza a sintetizarse desde el N-t al C-t dado que, efectivamente, es eso lo que ocurre en la célula (**Figura 91**).

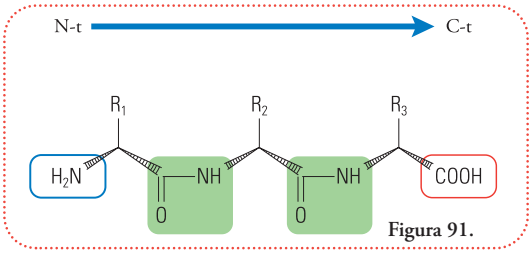


Figura 91.

Los códigos de una y tres letras permiten codificar las proteínas en lo que se conoce como estructura primaria. Hay que notar aquí que cada signo (letra, o grupo de tres letras) corresponde a un grupo de átomos, esto abrevia mucho la descripción de una proteína que cuenta con miles de ellos.

Hay un formato particular para codificar el modelo molecular de una proteína. Se llama formato .pdb. La característica diferente, más importante, con respecto al formato .mol es que, cada átomo, tiene sus coordenadas y, además, la información de a cuál aminoácido pertenece. Esta información es importante para generar una típica representación proteica que veremos a continuación.

Una característica importante para tener en cuenta en moléculas grandes y, sobre todo, en proteínas, es que éstas se pliegan sobre sí mismas y partes alejadas en la estructura primaria (secuencia) pueden estar cerca en la estructura plegada de la proteína (estructura terciaria).

Con estructura secundaria se señala el hecho de que determinadas regiones de la secuencia adoptan una conformación repetitiva particular. Se han caracterizado, principalmente, dos tipos de estructuras secundarias: las alfa hélices (“ -helix”, en inglés) y la hoja beta (“ -sheet”, en inglés).

Tienen las siguientes características:

Alfa hélices

Si orientamos el polímero de aminoácidos con el N-t hacia la izquierda y el C-t a la derecha y recorremos la molécula de izquierda a derecha por los enlaces peptídicos y los carbonos alfa (a continuación lo definiremos), observaremos un avance en tirabuzón que gira en el sentido de las agujas del reloj (dextrógiro).

Aclaramos que se denomina carbono alfa (C-α) al carbono que está unido, tanto al grupo amino como al grupo ácido carboxílico; y cuando el aminoácido se encuentra formando un polímero se encuentra entre dos grupos amida, salvo que sea el primer aminoácido, N-t (unido a un grupo amino y un grupo amida), o el último aminoácido, C-t (unido a un grupo amida y un grupo ácido carboxílico).

El conjunto de todos estos átomos (amino y carboxilo terminales, los carbonos alfa y los que forman el enlace peptídico, sin contar los hidrógenos) se llama esqueleto de carbonos (backbone, en inglés). En la **figura 92**, son los átomos con fondo verde.

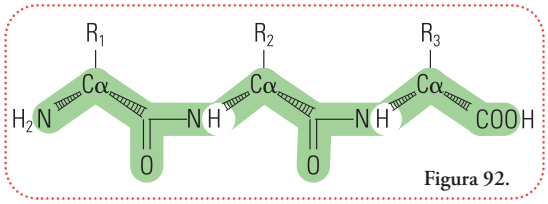


Figura 92.

El resultado es que las cadenas laterales se ubican hacia fuera. Típicamente, cada vuelta está formada por 3,6 aminoácidos (dado que cada aminoácido produce un giro de 100°) (**Figura 93**).

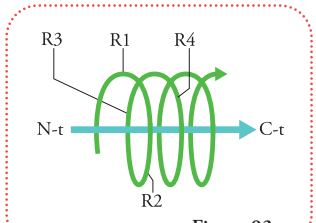


Figura 93.

Éste es un ejemplo real de hélice que podemos observar en el Deep View / Swiss PDB-Viewer. Este programa utiliza el código de colores CPK (debido a que fue propuesto por los investigadores Corey, Pauling y Koltun) el cual asigna el azul para el nitrógeno, el gris (o negro) para el carbono y el rojo para el oxígeno.[7] (**Figura 94, panel izquierdo**).

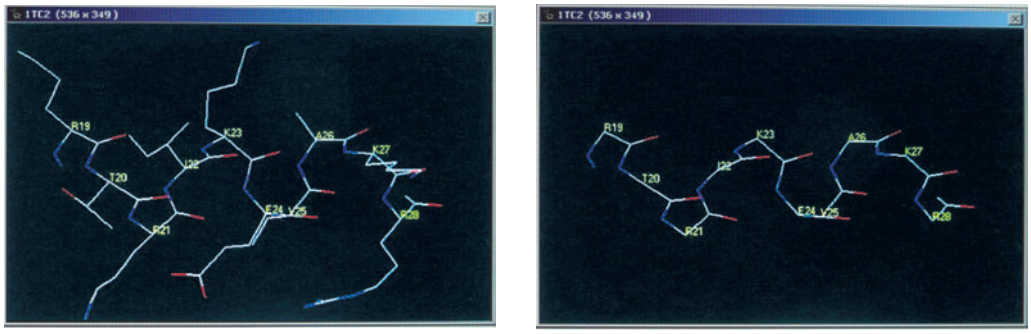


Figura 94.

Los aminoácidos se representan con el código de una sola letra y el número indica su posición en la estructura primaria (secuencia aminoacídica).

Si agregamos las cadenas laterales a la imagen se observará su ubicación, orientadas hacia el exterior (**Figura 94, panel derecho**).

[7] El código CPK además asigna el blanco para el hidrógeno, anaranjado para el fósforo y/o hierro, amarillo para el azufre y verde para el magnesio y/o cloro.

El Deep View / Swiss PDB Viewer, puede representar esta estructura alfa hélice como una cinta enroscada o un cilindro enroscado.

Hoja beta

En esta estructura secundaria sucede que cuando la secuencia es orientada como en el caso anterior (N-t hacia la izquierda y C-t hacia la derecha) las cadenas laterales quedan alternando una orientación y su opuesta. Nuevamente, al seguir el camino del esqueleto de carbonos (C-α más enlaces peptídicos) obtenemos (**Figura 95**).

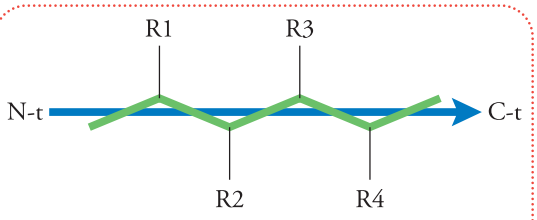


Figura 95.

Este segmento de secuencia con esta disposición espacial se llama hebra de hoja beta. Las hebras dejan al oxígeno y al nitrógeno del grupo funcional amida orientado de forma tal que se establece un tipo especial de unión débil (el puente de hidrógeno)[8] con una hebra contigua, dos hebras contiguas o más, forman lo que se denomina hoja beta.

Éste es un ejemplo real de una hebra de hoja beta que podemos observar su esqueleto de carbonos (**Figura 96**).

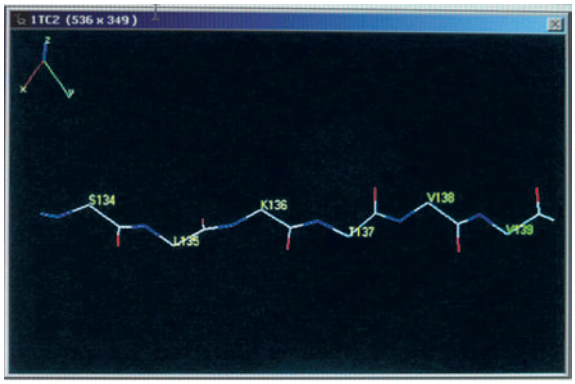


Figura 96.

Y la disposición de las cadenas laterales alternando sus orientaciones (**Figura 97**).

El Deep View / Swiss PDB-Viewer, puede representar esta hebra de hoja beta como una cinta plana con una punta de flecha en el C-t para identificar la orientación.

Por último hay que aclarar que no todos los aminoácidos se encuentran formando una estructura de alfa hélice u hoja beta.

El programa representa esta región, simplemente, con una curva gruesa.

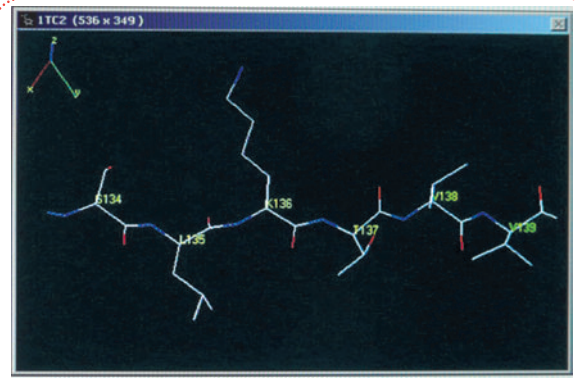
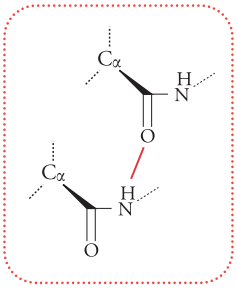


Figura 97.



A continuación observaremos las proteínas a las que pertenecen estos fragmentos que mostramos, se trata de la proteína Hipoxantina Guanina Fosforribosiltransferasa de Trypanosoma cruzi (el Tripanosoma cruzi es un parásito unicelular causante de la en-

[8] El puente de hidrógeno también está presente en la estructura α hélice, consiste en una unión débil (alrededor del 10% de la tensión de un enlace covalente) entre el hidrógeno unido al nitrógeno de la amida y los pares electrónicos no compartidos del oxígeno de otro enlace peptídico. Se suele representar con una línea de puntos (en rojo en la figura).

fermedad endémica de Sudamérica conocida como Mal de Chagas). El modelo molecular fue generado a partir de datos experimentales de difracción de Rayos X. Se encuentran a disposición en el PDB (por su sigla en inglés que quiere decir Protein Data Bank[9], en castellano Banco de Datos de Proteínas) bajo el código 1TC2. Los modelos se descargan vía INTERNET con el nombre 1TC2.pdb.

Se encontró que hay dos moléculas distinguibles en el modelo molecular generado a partir de los datos experimentales.

Para cargar el modelo, primero ejecutamos el Deep View / Swiss PDB-Viewer, aparecerá un barra con su menú y sus botones. Elegimos el menú desplegable "File" (archivo) y luego "Open PDB file" (abrir archivo PDB) (**Figura 98**).

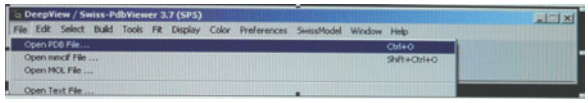


Figura 98.

Elegimos el archivo 1TC2.pdb y lo abrimos. Aparecerá el modelo en representación de alambre (**Figura 99**).

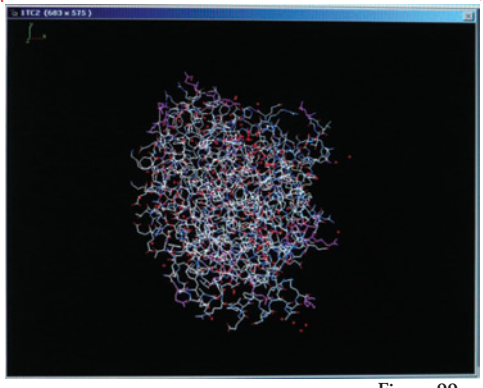


Figura 99.

Es difícil apreciar algo, ni siquiera es posible distinguir que se trata de dos moléculas distintas. Sí podemos ver pequeñas cruces rojas que representan el oxígeno del agua. El hidrógeno no es posible determinarlo ya que la metodología experimental no puede detectarlo (aunque el programa permite agregarlo por medio de un algoritmo que asigna la hibridación correspondiente a cada carbono).

A la derecha de la pantalla de trabajo aparece una ventana vertical con el título de "Control Panel" (panel de control)[10] (**Figura 100**).

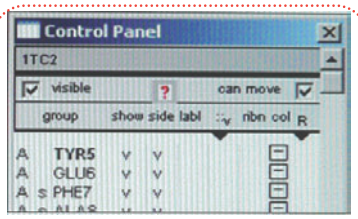


Figura 100.

Si seleccionamos el botón derecho sobre la palabra "side" (que quiere decir "lado" en inglés y se refiere a las cadenas laterales) obtendremos una representación que sólo es el esqueleto de carbonos y las aguas cristalográficas (las crucecitas rojas) (**Figura 101**).

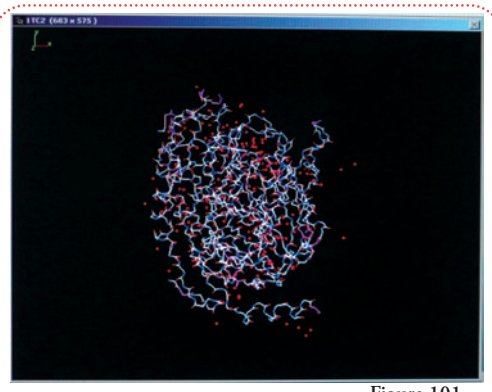


Figura 101.

Ahora es un poco más clara la representación, pero todavía no nos da una idea de la estructura terciaria.

En la ventana "Control Panel" seleccionamos el botón derecho del ratón debajo (en la columna en blanco) de la palabra "ribn" (que es la abreviatura de "ribbon" que en inglés significa "cinta") y sobre la palabra "show" (que en

[9] Accesible en la dirección electrónica: www.pdb.org.

[10] Si no apareciera es posible desocultarlo apretando las techas "ALT" + "+", o seleccionando del menú desplegable "Window" (ventana) la opción "Control Panel".

inglés significa “mostrar”) (**Figura 102**).

Éste es el resultado, el esqueleto de carbonos es representado en cintas (**Figura 103**).

Ahora podemos apreciar con mayor claridad la existencia de las zonas de estructura alfa hélice y hebras beta. Podemos girarla, y arrastrando el puntero sobre la ventana de trabajo observaremos claramente que estamos trabajando con dos moléculas separadas (**Figura 104**).

Para mejorar la representación seleccionamos el botón izquierdo del ratón sobre el triángulo que apunta hacia abajo en la ventana “Control Panel”, debajo de las letras “B S”, y seleccionamos la palabra “ribbon” (“cinta” en inglés) (**Figura 105**).

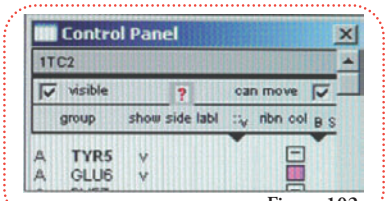


Figura 102.

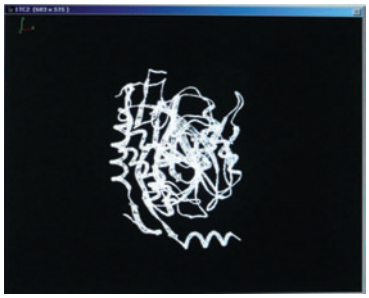


Figura 103.

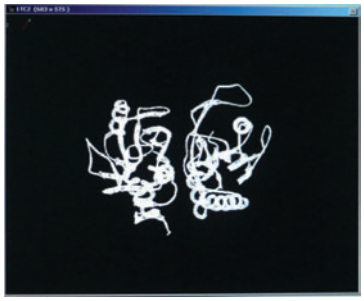


Figura 104.

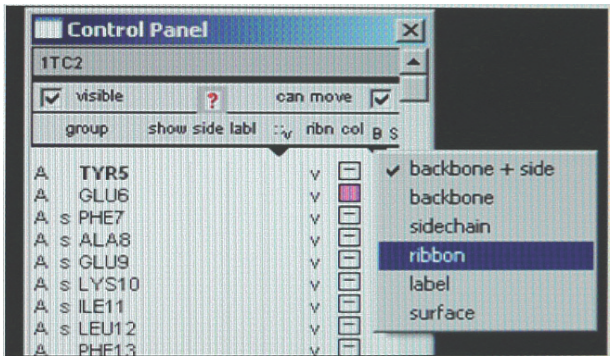


Figura 105.

Luego, vamos al menú desplegable “Display” (que significa en inglés “imagen presentada en una pantalla”) de la barra principal y seleccionamos las últimas dos opciones (se requiere entrar dos veces al menú “Display”) (**Figura 106**).

En la primera entrada seleccionamos la opción “Use OpenGL Rendering” indica usar el sistema OpenGL para presentar una imagen del modelo. En la segunda entrada al menú, seleccionamos la opción “Render in solid 3D” producirá una imagen de un objeto sólido iluminado (**Figura 107**).

Podemos mejorar la interpretación del modelo proteico si seleccionamos del menú desplegable “Color” la primera opción y luego sobre el segundo

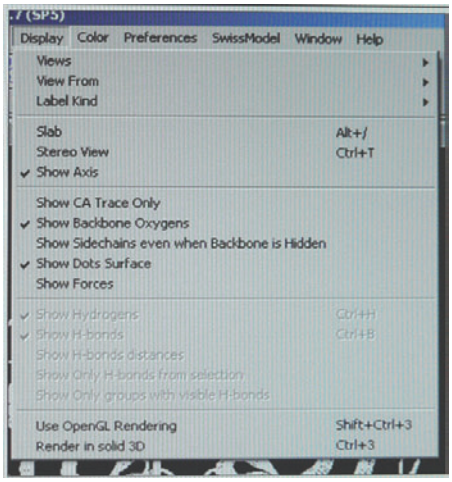
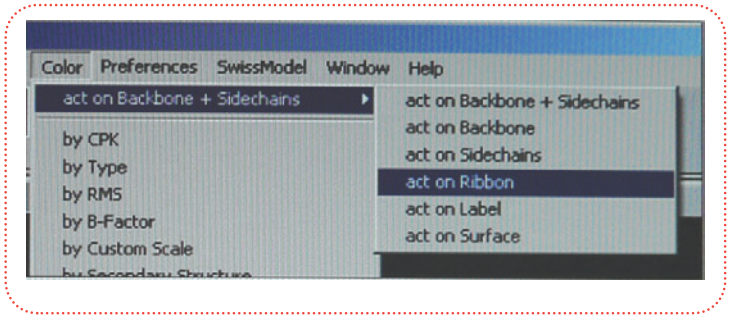


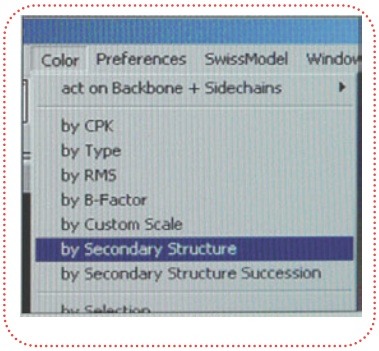
Figura 106.

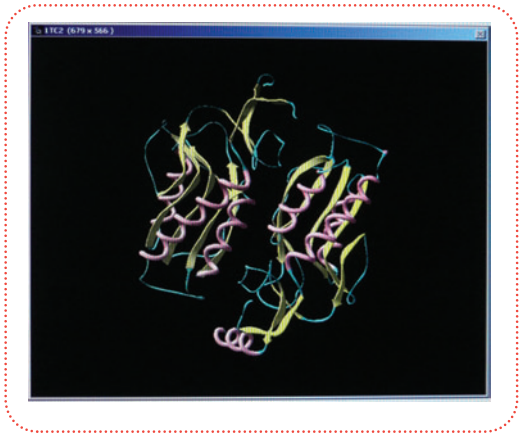
menú que se despliega, seleccionamos "act on Ribbon" (que quiere decir "actuar sobre la representación en cinta") (**Figura 108**).

A continuación seleccionamos, nuevamente de este menú, la opción "by Secondary Structure", esta opción indica que coloree el modelo sólido de forma tal que se distingan las distintas zonas de estructura secundaria (hélice alfa, hoja beta y ni una ni la otra) (**Figura 109**).

Este es el resultado luego de girar, arbitrariamente, el modelo final (**Figura 110**).







Hasta aquí llegaremos con la explicación de Deep View / Swiss PDB-Viewer. Les recomiendo que exploren las otras opciones por ustedes mismos.

Para quien lo desee (y pueda leer en idioma inglés) está disponible una guía completa de todas las funciones del programa online en la siguiente dirección: http://spdbv.vital-it.ch/TheMolecularLevel/SPVTut/