

CAPÍTULO 5. MATERIALES Y MÉTODOS

5.1 Modelamiento por homología

Las proteínas homólogas son aquellas que están evolutivamente relacionadas. Generalmente realizan una misma función en distintas especies, y pueden tener longitud similar. Ciertos residuos ocupan la misma posición en todas las especies, y se llaman residuos invariables. Por el contrario, las posiciones con variación considerable se llaman residuos variables, y proporcionan datos acerca del parentesco filogenético entre dos proteínas (Nelson y Cox 2000).

5.1.1 Búsqueda del patrón

La construcción de un modelo comparativo o por homología, comienza con la búsqueda de proteínas de estructura tridimensional conocida a partir de la secuencia blanco. Esto se logra al comparar la secuencia a estudiar con cada una de las secuencias correspondientes a las estructuras incluidas en una base de datos (Sánchez y Sali 2000). El Brookhaven Protein Data Bank es la única base de datos a nivel mundial que contiene estructuras tridimensionales de biomoléculas determinadas experimentalmente. A cada estructura almacenada se le asigna una clave (ID de PDB) que permiten su fácil identificación (Berman *et al* 2000). BLAST (Altschul *et al* 1990) comprende una serie de programas diseñados para explorar todas las bases de datos de secuencias disponibles e identificar aquellas con secuencias similares al blanco indicado. Con dicho programa se realizó este paso.

5.1.2 Selección del patrón

Una vez que se cuenta con una lista de los IDs en PDB de posibles patrones, debe seleccionarse el más apropiado. Generalmente se elige aquel cuya secuencia es la más similar a la del blanco, es decir, tiene un mayor porcentaje de identidad (mayor número de residuos iguales) y un menor número y longitud de omisiones (*gaps*). Es también importante que pertenezcan a una misma familia, que su ambiente y/o ligandos sean lo más parecidos posible, y que el modelo experimental cuente con la más alta resolución posible

(Sánchez y Sali 2000). Para seleccionar el patrón más adecuado se realizaron alineamientos múltiples mediante el programa CLUSTAL W (Thompson *et al* 1994).

5.1.3 Alineamiento

Para la construcción de un modelo por homología es necesario establecer la equivalencia entre los residuos del blanco y del patrón. Dichas equivalencias se definen mediante el alineamiento de las secuencias (Sánchez y Sali 2000). El alineamiento se basa en la presunción de que las secuencias con que se trabaja tienen un ancestro común. Por lo tanto, de acuerdo a los eventos de evolución ocurridos, se encontrará un cierto número de sustituciones de amino ácidos. Tras realizar numerosos estudios sobre la frecuencia en que ciertos amino ácidos son sustituidos por otros, se les ha asignado un valor dado por una matriz. La sustitución es más frecuente entre amino ácidos con propiedades bioquímicas similares. Por ejemplo, el cambio entre dos residuos hidrofóbicos reporta un valor positivo, mientras que el de un residuo hidrofóbico por uno hidrofílico reporta un valor negativo. Las inserciones y deleciones son menos frecuentes, por lo que se les da un peso mayor y se hace referencia a ellas como omisiones (Thompson *et al* 1994). En el caso del programa CLUSTAL W se trabaja con la matriz BLOSSUM que se muestra en el Apéndice 10.1.

Encontrar un alineamiento adecuado es relativamente simple cuando su porcentaje de identidad es mayor al 40%, pudiéndose obtener mediante métodos de alineamiento automáticos (Sánchez y Sali 2000). Por dicha razón se trabajó con el alineamiento obtenido anteriormente mediante CLUSTAL W.

5.1.4 Construcción del modelo

En la construcción del modelo se utilizó el programa SYBYL (SYBYL[®] 6.7.1 Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA) en una computadora SGI 02.

Tras su construcción, el modelo puede refinarse al aplicar un programa de dinámica molecular. Las simulaciones de dinámica molecular estudian el movimiento y el espacio conformacional de un sistema molecular al integrar las ecuaciones de Newton dada una función de energía potencial y su campo de fuerza asociado. La posición de los átomos se calcula en intervalos discretos de tiempo, generalmente de fs. Se requiere definir el campo

de fuerza con que se trabaja, siendo en este caso el Amber 4.1 que posee las siguientes constantes: Constante dieléctrica= 1.0¹

Escala del radio con puentes de hidrógeno= 1.0²

Escala uno cuatro = 0.5³

La función de energía dependerá de la temperatura del sistema. Por lo tanto, para poder explorar las posibles conformaciones del sistema, puede iniciarse la simulación a una temperatura muy alta donde todas las conformaciones sean energéticamente accesibles. Posteriormente la temperatura disminuye gradualmente hasta un valor dado, donde se espera que el sistema adopte una conformación natural. Este ciclo puede repetirse varias veces y a distintas temperaturas para encontrar la conformación termodinámicamente más estable. En un rango de aproximadamente 100 kJ o 25 kcal no se observan conformaciones discretas, si no únicamente rotación libre, pudiendo estudiarse también los distintos rotámetros de la proteína (*Force Field manual*).

5.1.5 Evaluación

Tras la construcción del modelo, deben buscarse posibles errores en el mismo. La primera evaluación es interna, verificando que el modelo satisfaga las restricciones con que fue calculado. Posteriormente se hace una evaluación externa, que toma en cuenta distintos aspectos (Sánchez y Sali 2000). Para la evaluación externa se utilizó la versión en línea del programa WHAT IF (Rodríguez *et al* 1998).

El programa evalúa la estereoquímica (longitud y ángulos de enlace, ángulos dihedros, planaridad, etc) de los residuos en una proteína al compararlos con parámetros derivados de estructuras obtenidas experimentalmente y con los “ideales” teóricos. Pueden así identificarse segmentos con estructura inadecuada (Sánchez y Sali 2000). Para cada parámetro, el programa reporta un resultado numérico y un valor *Z* que indica el número de desviaciones estándar entre el resultado y el valor esperado. Además se calcula la raíz-media-cuadrada de un grupo de valores *Z*
$$\text{RMS } Z = \sqrt{\frac{\sum (Z^2)}{n}}$$
 que suele ser 1.0 (Vriend G, 1990).

¹ La constante dieléctrica=1 indica un ambiente abundante en amino ácidos y con pocas moléculas de agua.

² Se aplica como un factor de escala a la suma de los radios de van der Waals de los átomos involucrados en puentes de hidrógeno.

³ Factor de escala para el término de energía de la interacción de átomos separados entre si por tres enlaces.