

BABELPDB: Extracción de información en la base de datos PDB para la enseñanza

BABELPDB: Information from the PDB Database for teaching

FRANCISCO TORRENS.

En el banco de datos Brookhaven Protein Data Bank (PDB) cada estructura se direcciona en un fichero cuyo nombre se codifica como iABC.BRK (?) donde i es el código de Brookhaven.⁽¹⁾ Un fichero PDB tiene dos partes. La primera parte contiene los autores, grupo, estructura espacial secundaria y secuencia. La segunda parte contiene las coordenadas (X, Y, Z), átomos (i), iones, conectividades entre átomos, factores de agitación térmica anisótropos (B); X, Y, Z, B, i. PDB contiene ca. 1000 estructuras cristalinas de proteínas, ácidos nucleicos y virus.

El programa BABEL implementa un marco general para convertir formatos de ficheros utilizados en modelización molecular.⁽²⁾ BABEL lee 37 y escribe 29 tipos de ficheros incluyendo PDB. BABEL puede llamarse bien utilizando las opciones de la línea de comandos, bien menús. BABEL tiene la capacidad de añadir y borrar hidrógenos de cualquier formato de fichero. Cuando se leen ficheros PDB, BABEL asigna enlaces examinando las distancias interatómicas y asignando un enlace allí donde la distancia interatómica es menor que la suma de los radios covalentes atómicos.

Hemos escrito el programa BABELPDB para la búsqueda, recuperación, análisis y representación de información basada en el ordenador de la base de datos PDB. Se permiten varias opciones: (1) convertir de PDB a otros formatos, (2) añadir o eliminar átomos de hidrógeno, (3) eliminar moléculas de agua y (4) extraer sólo los carbonos α .

Con BABELPDB, se han convertido las coordenadas de varias proteínas de PDB a coordenadas cartesianas, eliminado las moléculas de agua de hidratación y

añadido los átomos de hidrógeno. Con estas coordenadas se ha analizado la presencia de *enlaces de hidrógeno* en las macromoléculas. Los enlaces de hidrógeno se caracterizan por la presencia de interacciones X-H...Y donde los átomos X e Y son N, O, F o Cl, distancia(??)Å y ángulo de enlace (??)°. Este algoritmo se ha implementado en el programa TOPO para la simulación teórica de la forma molecular.^(3,4) TOPO permite calcular los descriptores geométricos e índices topológicos de las macromoléculas incluyendo la dimensión fractal de la superficie accesible al disolvente.

Se he estudiado la lisozima utilizando las coordenadas de rayos X de los átomos (2LYM) extraídas con BABELPDB. La enzima contiene 129 aminoácidos (1906 átomos) y tiene un peso molecular de 14 307 Daltons. Hay 151 moléculas de agua alrededor de la proteína enzimática. Para la lisozima con átomos de hidrógeno los resultados muestran una *dimensión fractal D* de 1.908. Para la lisozima sin átomos de hidrógeno se observa un valor para *D* de 1.920.

Se ha extraído el esqueleto de carbonos α de la lisozima, el cual ha permitido dibujar la imagen de cintas. Esta imagen determina los elementos de la estructura secundaria. En la imagen se distinguen cuatro regiones helicoidales (A—D). Tres de ellas son hélices α distorsionadas (A, B, D) y la otra es una hélice 3.0_{10} (C). La lisozima contiene una hoja β antiparalela (E).

De los resultados se pueden extraer las siguientes conclusiones:

1. Se ha escrito el programa BABELPDB para la búsqueda, recupera-

ción, análisis y representación de información de la base de datos PDB. Las coordenadas han permitido localizar los enlaces de hidrógeno. El algoritmo se ha implementado en el programa TOPO para una mejor adecuación didáctica al estudio de las propiedades moleculares en la enseñanza.⁽⁵⁻⁷⁾

2. El esqueleto de carbonos a extraído con BABELPDB permite dibujar la imagen de cintas que determina la estructura secundaria de las proteínas. Los profesores encontrarán las opciones de BABELPDB muy útiles para demostraciones. El programa puede también emplearse por alumnos para explorar las técnicas de extracción de información de PDB.

3. Queda por explorar la acomodación metodológica de estos programas enriqueciendo las técnicas actuales de enseñanza.⁽⁸⁾ Implementar nuevos algoritmos en situaciones de aprendizaje es informáticamente sencillo, pero comprobar su utilidad en la enseñanza es extraordinariamente complejo y exige una investigación que apenas ha comenzado.



BIBLIOGRAFÍA

- [1] F. C. BERNSTEIN, T. F. KOETZLE, G. J. B. WILLIAMS, E. F. MEYER JR., M. D. BRICE, J. R. RODGERS, O. KENNARD, T. SHIMANOUCI Y M. TASUMI, *J. Mol. Biol.*, 112 (1977) 535.
- [2] P. WALTERS Y M. STAHL, Program BABEL, University of Arizona, Tucson, AZ, 1994.
- [3] F. TORRENS, E. ORTÍ Y J. SÁNCHEZ-MARÍN, *J. Chim. Phys. Phys.-Chim. Biol.*, 88 (1991) 2435.
- [4] F. TORRENS, J. SÁNCHEZ-MARÍN E I. NEBOT-GIL, *J. Mol. Graphics Mod.*, 16 (1998) 57.
- [5] F. TORRENS, J. SÁNCHEZ-MARÍN Y E. SÁNCHEZ-PÉREZ, Didáctica empírica de la congelació de l'aigua, in: *Actes del II Symposium sobre l'Ensenyament de les Ciències Naturals*, S. Riera, Ed., Documents No. 11, Eumo, Vic, 1989, p. 595.
- [6] F. TORRENS, J. SÁNCHEZ-MARÍN Y E. SÁNCHEZ-PÉREZ, Estudi interdisciplinari de la congelació de l'aigua, in: *Actes del II Symposium sobre l'Ensenyament de les Ciències Naturals*, S. Riera, Ed., Documents No. 11, Eumo, Vic, 1989, p. 669.
- [7] F. TORRENS, E. SÁNCHEZ-PÉREZ Y J. SÁNCHEZ-MARÍN, Didáctica empírica de la forma molecular, *Ensenyament de las Ciencias*, Número Extra (III Congreso)-I (1989) 267.
- [8] F. TORRENS, E. ORTÍ Y J. SÁNCHEZ-MARÍN, Representación de propiedades moleculares en la didáctica de la química, in: *Colloquy University Pedagogy*, Horsori, Barcelona, 1991, p. 375.