

ESTUDIO COMPARATIVO DE LOS POTENCIALES DE FUERZA MEDIA AL CONTACTO ENTRE RESIDUOS DE AMINOÁCIDOS.

D. G. Renzi^{1,2} y F. Vericat^{1,3}

¹ Instituto de Física de Líquidos y Sistemas Biológicos (IFLYSIB) UNLP-CONICET, La Plata, Argentina.

² Facultad de Ciencias Veterinarias, Universidad Nacional de Rosario.

³ Grupo de Aplicaciones Matemáticas y Estadísticas de la Facultad de Ingeniería (GAMEFI), Universidad Nacional de La Plata.

e-mail: drenzi@fveter.unr.edu.ar, vericat@iflysib.unlp.edu.ar

La hipótesis termodinámica establece al estado nativo de una proteína como el de más baja energía libre bajo condiciones fisiológicas normales. Luego, para juzgar la calidad de un modelo de proteína propuesto, o para identificar el plegamiento nativo de una secuencia de aminoácidos entre muchas alternativas posibles, necesitamos contar con funciones de energía efectiva apropiadas. Los términos "energía efectiva" o "potencial de fuerza media" se refieren a la energía libre del sistema para una conformación determinada de la proteína, es decir, la energía intramolecular de las proteínas más la energía libre de solvatación. Las funciones de energía efectivas utilizadas actualmente provienen generalmente de dos orígenes distintos. El primero está basado sobre la función de energía que se obtiene, en principio, de un análisis fundamental de las fuerzas entre partículas. El segundo está compuesto por funciones de energía "estadísticas" que derivan de frecuencias de pares de aminoácidos, observados en un conjunto de estructuras de proteínas conocidas. En este trabajo calculamos funciones de energía efectivas entre residuos de aminoácidos a partir de una teoría de líquidos clásicos, y construimos una tabla de valores de interacción al contacto a fin de comparar con las energías estadísticas obtenidas por otros autores.

The thermodynamical hypothesis says that the native state of a protein is that which has the lowest free energy consistent with the normal physiological conditions. Thus, in order to judge the quality of a given model of protein, or to identify the native folding of a sequence of amino acids among many possible sequences, we need appropriate effective energy functions. With the terms "effective energy" or "mean force potential" we refer to the system free energy for a given conformation of the protein, say the intra-molecular energy of the proteins plus the solvation free energy. The effective energies currently used belong to someone of two different groups. In one of them are the effective energies which are obtained from a basic analysis of the inter particles forces. In the other group are "statistical" energy functions which are calculated from the appearance frequency, in a set of structures corresponding to known protein sequences, of a given pair of amino acids at fixed distances. In this work we calculate effective energy functions between pairs of amino acids residues by using a classic liquid state theory. We construct a table of contact values for these interactions and compare our results with the statistical energies obtained by other authors.

Introducción

Predecir la estructura tridimensional de una proteína a partir de su secuencia de aminoácidos, es hoy una de las metas más ambiciosas de la Biología.

Actualmente decenas de miles de estructuras de proteínas conocidas se encuentran almacenadas en el *Protein Data Bank*⁽¹⁾.

Por otro lado, se conocen también cientos de miles de secuencias de proteínas, y aunque la información es algo redundante, para muchas de estas secuencias se desconoce aún su estructura y función.

Se estima que la velocidad en la determinación de secuencias es al menos unas 50 veces superior a la velocidad con que se determinan las estructuras.

Esta diferencia, llevó a que muchos investigadores se abocaran a la tarea de desarrollar técnicas aptas para la predicción o el reconocimiento de estructuras plegadas, o de ciertos *test* capaces de seleccionar una dada estructura como la más favorable de entre un conjunto de estructuras posibles⁽²⁾.

La técnica de reconocimiento de estructura plegada consiste en "ubicar" a una dada secuencia de aminoácidos, sobre un conjunto de estructuras plegadas conocidas.

El criterio que se utiliza para seleccionar a la estructura más probable está basado en el cálculo de energía, obtenida usando potenciales de fuerza media entre pares de residuos.

Hay tres elementos básicos en la implementación de esta técnica: (i) los potenciales de fuerza media o energías efectivas, (ii) la forma de "ubicar" a una dada secuencia sobre cada estructura posible, y (iii) el criterio usado para decidir cuál es la estructura "correcta" de todas las consideradas.

Si bien existen algunas variantes sobre cada uno de los pasos descriptos según los diferentes autores que han trabajado sobre el tema, aquí nos concentramos en el primero de estos elementos, es decir, en la determinación de los potenciales efectivos a utilizar.

Comúnmente se utilizan funciones de energía efectiva obtenidas con dos tipos de procedimientos distintos. Una, basada en el análisis físico de las fuerzas de interacción entre las partículas, y la otra es una función similar a la energía, obtenida de datos derivados de cálculos estadísticos sobre un conjunto de estructuras de proteínas conocidas.

En este trabajo presentamos funciones de energía efectiva de pares obtenidas a partir de un modelo de líquidos clásicos⁽³⁾, y comparamos los valores al contacto con algunos de los obtenidos por métodos estadísticos para estudiar posibles correlaciones entre los valores de cada uno.

Modelo teórico

Consideraremos a las moléculas de residuo como esferas disueltas en agua, que interactúan entre sí y con el solvente a través de un potencial $u_{12}(|R_1 - R_2|)$ que

depende solamente de la distancia entre los centros de los residuos, o entre el centro de un residuo y el oxígeno de una molécula de agua. Usamos "ρ" para designar a los residuos de aminoácido, y "ρ'" para la molécula de agua.

Expresamos a la función de distribución radial residuo-agua (ρ_α) y residuo-residuo (ρρ') por,

$$g_{\rho\alpha}(r) = y_{\rho\alpha}(r) \exp(-u_{\rho\alpha}(r)/k_B T)$$

$$g_{\rho\rho'}(r) = y_{\rho\rho'}(r) \exp(-u_{\rho\rho'}(r)/k_B T)$$

donde $y_{\rho\alpha}$ es la función de distribución radial de cavidad.

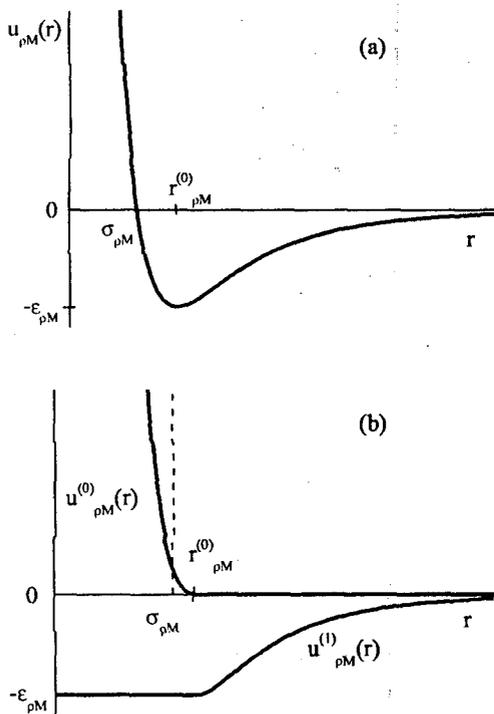


figura 1

Los potenciales de pares desnudos que consideramos, tienen todos la forma mostrada en la figura 1a. Indicamos con $r_{\rho M}^{(0)}$ a la posición del mínimo de $u_{\rho M}(r)$, y con $\epsilon_{\rho M}$ al valor que toma $u_{\rho M}(r)$ en $r_{\rho M}^{(0)}$.

Descomponemos al potencial desnudo $u_{\rho M}(r)$ en dos partes: una repulsiva de corto alcance $u_{\rho M}^{(0)}(r)$, y otra atractiva de largo alcance $u_{\rho M}^{(1)}(r)$ (figura 1b). El potencial $u_{\rho M}(r) = u_{\rho M}^{(0)}(r) + u_{\rho M}^{(1)}(r)$, donde $M = \alpha$ o ρ y

$$u_{\rho M}^{(0)} = \begin{cases} \epsilon_{\rho M} + u_{\rho M}(r) & r < r_{\rho M}^{(0)} \\ 0 & r > r_{\rho M}^{(0)} \end{cases}$$

$$u_{\rho M}^{(1)} = \begin{cases} -\epsilon_{\rho M} & r < r_{\rho M}^{(0)} \\ u_{\rho M}(r) & r > r_{\rho M}^{(0)} \end{cases}$$

Por otra parte, tratamos a las interacciones entre moléculas de agua en forma exacta, esto es, $g_{\alpha\alpha}(r)$ es la

función de distribución radial oxígeno-oxígeno obtenida en forma experimental⁽⁴⁾.

Introducimos funciones de correlación $c_{\rho M}(r)$ en las ecuaciones del tipo Ornstein-Zernike⁽³⁾,

$$h_{\rho\alpha}(r) = c_{\rho\alpha}(r) + \rho_{\alpha} \int c_{\rho\alpha}(|\mathbf{r} - \mathbf{r}'|) h_{\alpha\alpha}(r') d\mathbf{r}'$$

$$h_{\rho\rho'}(r) = c_{\rho\rho'}(r) + \rho_{\alpha} \int c_{\rho\alpha}(|\mathbf{r} - \mathbf{r}'|) h_{\rho'\alpha}(r') d\mathbf{r}'$$

donde $h_{\rho\alpha}(r) = g_{\rho\alpha}(r) - 1$ es la función de correlación de pares experimental para el agua líquida pura.

Consideramos las relaciones,

$$c_{\rho\alpha}(r) = 0 \quad r > \sigma_{\rho\alpha}$$

$$c_{\rho\rho'}(r) = 0 \quad r > \sigma_{\rho\rho'}$$

$$g_{\rho\alpha}(r) = h_{\rho\alpha}(r) + 1 = 0 \quad r < \sigma_{\rho\alpha}$$

$$g_{\rho\rho'}(r) = h_{\rho\rho'}(r) + 1 = 0 \quad r < \sigma_{\rho\rho'}$$

y resolvemos este sistema de ecuaciones utilizando un procedimiento variacional para determinar $c_{\rho\alpha}(r)$, con el que calculamos $g_{\rho\rho'}(r)$, y finalmente,

$$W_{\rho\rho'}(r) = -k_B T \ln(g_{\rho\rho'}(r))$$

Discusión y resultados

Existen en la naturaleza 20 aminoácidos distintos, con los cuales se forman las cadenas poliméricas conocidas como proteínas. Según su polaridad los podemos clasificar como *apolares*, *polares* y *cargados*.

Determinar las energías efectivas de interacción de pares entre estos 20 aminoácidos significa calcular un total de 210 potenciales. Por supuesto, resulta imposible mostrarlos a todos, así que sólo presentaremos un conjunto reducido de ellos.

Para calcular las funciones de energía efectiva $W_{\rho\rho'}(r)$ correspondientes a residuos *apolares*, utilizamos un potencial desnudo de la forma Lennard-Jones (12-6). El parámetro "σ" queda determinado experimentalmente⁽⁵⁾, y al "ε" lo ajustamos teniendo en cuenta el número de coordinación de cada residuo.

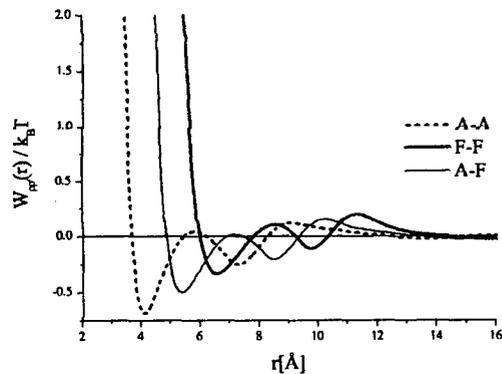


figura 2

Presentamos en la figura 2, las curvas correspondientes a dos de los residuos *apolares*, el más pequeño y el más grande: *alanina* (A) y *fenilalanina* (F). Estos potenciales

exhiben un comportamiento cualitativamente similar al mostrado por los demás residuos del grupo.

A distancias menores que el contacto son fuertemente repulsivos, presentan dos pozos y dos barreras bien definidas, y se anulan para distancias mayores a 14 Å. El primero de los pozos es el más profundo (0.3–0.75 $k_B T$) y se halla a una distancia cercana al valor al contacto. El segundo (0.1–0.3 $k_B T$) se encuentra separado del primero por una distancia del orden del tamaño de una molécula de agua.

Energéticamente podemos decir que resulta favorable el contacto entre dos residuos de alanina frente a dos residuos de fenilalanina. Además, si tenemos en cuenta la altura relativa de las barreras, vemos que los residuos de alanina estarán más fuertemente ligados que los de fenilalanina.

El grupo de los aminoácidos polares está formado por 8 moléculas con residuos distintos. Nuevamente elegimos sólo dos de ellos para mostrar. Uno de ellos, *serina* (S), es el aminoácido polar de más frecuente aparición en proteínas y el otro, *triptófano* (W), es el de menor aparición entre los 20 residuos de aminoácidos.

Para calcular la energía efectiva de interacción entre residuos polares, utilizamos un potencial desnudo que tiene un término repulsivo del tipo esferas blandas y un término atractivo dipolo-dipolo, que es función sólo de la distancia (angularmente promediado)⁽⁶⁾. En este caso tenemos tres parámetros de ajuste: " σ ", " ϵ " y " q_{ef} ".

En la figura 3 comparamos los potenciales efectivos de pares S-S, W-W y S-W. Las gráficas de estos potenciales son representativas de todos los residuos polares. A distancias menores que el contacto se comportan en forma repulsiva, presentan tres pozos y dos barreras y finalmente tienden a cero desde valores negativos.

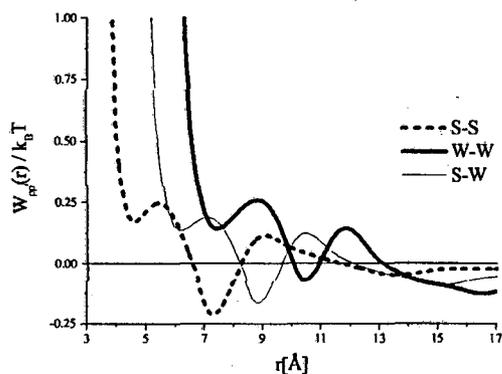


figura 3

El primer pozo o mínimo local toma en todos los casos valores positivos. El segundo mínimo, negativo para todos los pares de residuos, está separado del primero por una distancia algo menor que el tamaño de una molécula de agua. A este mínimo de valor negativo le sigue una barrera que oscila entre los 0.23 y 0.33 $k_B T$, y luego a una distancia aproximada de dos veces el diámetro de una molécula de agua aparece un tercer

mínimo de energía. A partir de allí, el potencial tiende a cero tomando valores mayores a $-0.1 k_B T$.

El último de los grupos de aminoácidos está formado por residuos cargados, dos con carga positiva: *lisina* (K) y *arginina* (R), y dos con carga negativa: *ácido aspártico* (D) y *ácido glutámico* (E).

Como en el caso de los residuos polares, usamos potenciales desnudos de interacción de pares que tienen un término repulsivo del tipo esferas blandas, más un término atractivo coulombiano. Nuevamente los parámetros que hay que ingresar son tres: " σ ", " ϵ " y " q_{ef} ".

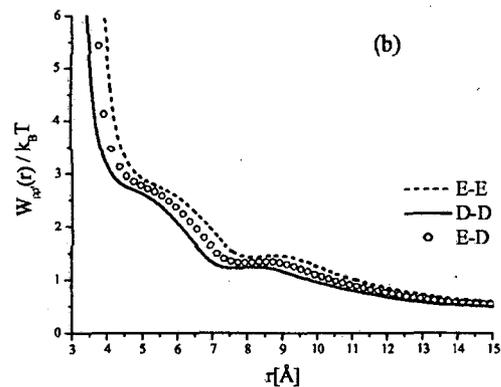
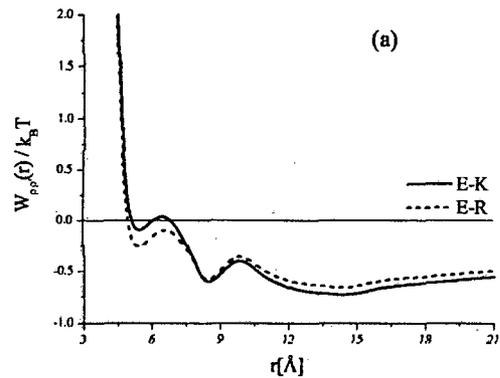


figura 4

En la figura 4a, mostramos las curvas correspondientes a las interacciones efectivas entre residuos con cargas de signos opuestos, y en la figura 4b las gráficas para residuos con cargas de igual signo.

Estas curvas muestran un razonable acuerdo cualitativo con lo obtenido por otros autores para sistemas formados por iones disueltos en agua⁽⁷⁾.

Finalmente presentamos en la figura 5, potenciales efectivos de interacción entre pares de residuos de aminoácidos que pertenecen a grupos diferentes. Seleccionamos tres residuos: uno *apolar* (A), uno *polar* (S) y uno *cargado* (E).

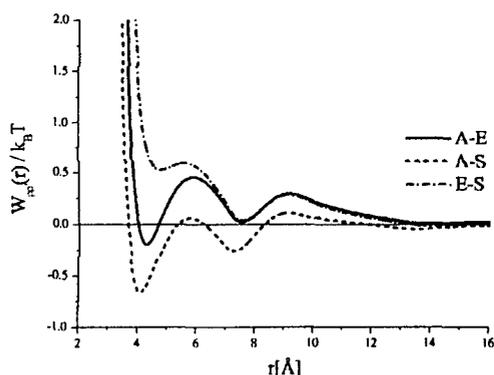


figura 5

Por supuesto no podemos mostrar a los 116 potenciales efectivos que corresponderían a este caso, ni siquiera podemos decir que éstos que hemos elegido sean los más representativos del conjunto.

Si comparamos las gráficas A-E con A-S, podemos observar la diferencia que introduce la presencia del solvente (agua) sobre las interacciones efectivas, teniendo en cuenta los potenciales desnudos de los diferentes residuos (cargado y polar) con el agua.

Como comentamos al principio, este no es el único camino para acceder al cálculo de energías efectivas. Hay otros métodos para derivar cantidades asociadas a potenciales efectivos entre pares de residuos, muchos de los cuales se basan en un cálculo de tipo estadístico sobre estructuras de proteínas conocidas⁽⁸⁾.

Aunque los caminos elegidos para llegar a obtener esos potenciales de energía efectiva de pares son muy diferentes, nos preguntamos si no existirá alguna correlación entre esos potenciales de tipo *estadísticos* y los que calculamos acá con elementos de la física de líquidos clásicos.

Para esto, generamos una tabla con los valores de energía al contacto entre pares de residuos de aminoácidos, y los comparamos con la tabla publicada por Thomas y Dill⁽⁸⁾.

Consideramos los 210 valores de interacción al contacto de cada una de las tablas y calculamos el *coeficiente de correlación dependiente de la posición de Spearman* (r_s)⁽⁹⁾.

Este parámetro, presenta varias ventajas con respecto al coeficiente de correlación lineal, el cual se calcula con los valores correspondientes de cada una de las tablas, que desde luego no coinciden.

Para calcular el coeficiente de correlación de Spearman, reemplazamos a cada valor en nuestra tabla, por su posición con respecto a los otros valores. Luego, hacemos lo mismo con la tabla de Thomas y Dill.

Si llamamos con V_i a la posición que ocupan los valores de energía c_i en nuestra tabla, y con D_i a las posiciones de los valores de energía ε_i en la otra tabla, el coeficiente de correlación se define como,

$$r_s = \frac{\sum_i (V_i - \bar{V})(D_i - \bar{D})}{\sqrt{\sum_i (V_i - \bar{V})^2} \sqrt{\sum_i (D_i - \bar{D})^2}}$$

Este coeficiente de correlación resulta ser más robusto que el coeficiente de correlación lineal, y menos dependiente de "anomalías" en los datos.

Utilizando la rutina "spear.f" de *Numerical Recipes*⁽⁹⁾, encontramos un valor 0.35 para el parámetro r_s y nivel de *significancia* de su desviación de cero igual a 2.1×10^{-2} . Éste último valor es muy pequeño (<0.01), lo que indica que existe una correlación significativa entre los valores de energía de las dos tablas comparadas.

Agradecimientos

Los autores agradecen a las Universidades Nacionales de Rosario y La Plata, al CONICET y a la ANPCyT por la ayuda económica que permite llevar adelante los proyectos de los cuales este trabajo forma parte. F.V. es miembro del CONICET.

Referencias

1. Protein Data Bank (<http://www.rcsb.org/pdb/>).
2. Bowie, J.U., Luthy, R. and Eisenberg, D., *Science*, **235**, 164 (1991); Maiorov, V.M. and Crippen, G.M., *J. Mol. Biol.*, **227**, 876 (1992); Moul, J., Pedersen, J.T., Judson, R. and Fidelis, K., *Proteins: Struct. Funct. Genet.*, **23**, 2 (1995).
3. Pratt, L.R. and Chandler, D. *J. Chem. Phys.*, **67**, 3683 (1977).
4. Soper, A.K., *J. Chem. Phys.*, **101**, 6888 (1994).
5. Creighton, T.E. *Proteins. Structure and Molecular Properties* (New York: W.H. Freeman and Company, 1993).
6. Laría, D. and Vericat, F. *Phys. Rev. A*, **43**, 1932 (1991).
7. Bagatella-Flores, N. and González-Mozuelos, P., *J. Chem. Phys.*, **117**, 6133 (2002).
8. Thomas P.D. and Dill, K.A., *Proc. Natl. Acad. Sci. USA*, **93**, 11628 (1996).
9. Press, W.H., Teukolsky, S.A. Vetterling, W.T. and Flannery, B.P., *Numerical Recipes in Fortran 77. The Art of Scientific Computing, Second Edition, Vol 1* (New York: Syndicate of the University of Cambridge, 1997)