

Evaluación de funciones de energía para el problema de empaquetamiento de la cadena lateral en proteínas

José D. Colbes y Carlos A. Brizuela

Departamento de Ciencias de la Computación

Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE)

Km. 107 Carretera Tijuana-Ensenada, Ensenada, B. C., 22860, Mexico.

Email: jcolbes@cicese.edu.mx, cbrizuel@cicese.mx

Resumen—El modelado de la conformación de la cadena lateral en proteínas es crucial para la predicción de estructuras de proteínas, el diseño de proteínas, y el acoplamiento entre proteínas. Debido a la atención que se dedica a este problema, en la literatura existen varios métodos que obtienen resultados razonablemente precisos y rápidos; pero aún queda un considerable espacio para mejoras. La mayoría de estos métodos utiliza un conjunto discreto de conformaciones de la cadena lateral, y explora el espacio de conformaciones de distintas maneras. En los últimos años sólo se tuvieron avances marginales en cuanto a precisión, lo cual puede deberse a limitaciones en la función de energía o en la exploración del espacio de conformaciones. En este trabajo se propone una metodología para la evaluación de funciones de energía, cuyos resultados preliminares indican que las mismas son las principales limitantes para la mejora de los resultados del problema considerado.

I. INTRODUCCIÓN

El problema de empaquetamiento de la cadena lateral en proteínas (PSCPP por sus siglas en inglés) consiste en predecir las coordenadas tridimensionales de todos los átomos que conforman la cadena lateral de cada uno de los residuos de una proteína, conociendo la cadena de residuos que conforman la proteína y las coordenadas de los átomos de la columna vertebral de la misma.

Si se discretiza el espacio de búsqueda con el uso de rotámeros y se usa una función de energía como función objetivo, el PSCPP se convierte en un problema de optimización combinatoria. El PSCPP es un problema NP-difícil [1].

Inicialmente se consideraba al PSCPP como parte del problema de predicción de estructuras [22]. Luego, varios trabajos se centraron en el análisis de la conformación de las cadenas laterales [8], [30], [3], [15], [14], los cuales dieron lugar a las bibliotecas de rotámeros [29], [2], [12], [10]. El PSCPP pasó entonces a ser un problema tratado separadamente, con aplicaciones en la predicción de estructuras (en particular, el modelado homólogo [4]), el diseño de proteínas y el refinamiento de estructuras [16].

I-A. Medidas de calidad de resultados

Los métodos propuestos para dar soluciones al PSCPP pueden evaluarse *in silico*, comparando las estructuras obtenidas por dichos métodos con las que se encuentran en el PDB. Se utilizan en general dos métricas: la desviación cuadrática media (RMSD) y la precisión absoluta.

Para la precisión absoluta se utilizan los valores $\chi_1(\%)$ y $\chi_{1+2}(\%)$, donde $\chi_1(\%)$ representa el porcentaje de residuos cuyos ángulos de torsión χ_1 son correctos, mientras que $\chi_{1+2}(\%)$ es el porcentaje de residuos cuyos ángulos de torsión χ_1 y χ_2 son ambos correctos. Se considera que un ángulo de torsión es correcto si se encuentra alejado a un máximo de 40° del ángulo de torsión de la estructura nativa de la proteína. Este valor arbitrario se ha venido utilizando desde los primeros trabajos que tratan sobre el PSCPP [35], [21], [37], [12], bajo la suposición de que los valores predichos corresponden al mismo mínimo de energía. De hecho, los algoritmos con mejores resultados para este problema siguen utilizando este valor como referencia.

El RMSD se utiliza como medida de diferencia entre estructuras, y su valor es la raíz cuadrada del promedio cuadrático de los valores de distancia entre los átomos correspondientes de las dos estructuras que se comparan. El PSCPP no requiere un alineamiento previo de las estructuras, pues se conoce de antemano los pares de átomos equivalentes.

Se ha comparado el desempeño de algunos de los algoritmos más populares del estado del arte para el PSCPP [27], considerando el conjunto de pruebas propuesto para el SCWRL4 (un total de 379 proteínas) [19]. Los resultados se muestran en la Tabla I.

Tabla I. COMPARACIÓN DE LA PRECISIÓN DE LOS MÉTODOS EN EL CONJUNTO DE PRUEBA DEL SCWRL4 [27]. SE INCLUYE AL SIDEPRO POR UTILIZAR EL MISMO CONJUNTO DE PRUEBA.

Método	$\chi_1(\%)$	$\chi_{1+2}(\%)$
OPUS-Rota [26]	85.03	71.74
SCWRL4 [19]	85.03	75.44
CIS-RR [7]	84.88	74.88
RASP [27]	85.10	74.71
SIDEpro [28]	86.14	74.15

Estos algoritmos obtienen una precisión de aproximadamente 85% para $\chi_1(\%)$ y 75% para $\chi_{1+2}(\%)$. Por otro lado, se ha demostrado que usando una biblioteca sencilla de rotámeros independiente de la columna vertebral, es posible lograr una precisión mejor [9], [23]. Para analizar si esta diferencia se debe a los algoritmos de búsqueda o a las funciones de energía utilizadas, se aplicó las funciones de energía de varios métodos del estado del arte junto con un algoritmo de búsqueda local, teniendo a la estructura nativa como referencia inicial [31]. En un caso ideal dicha estructura debería permanecer inalterada, ya que al menos debería ser un mínimo local. Pero luego de las pruebas, además de encontrar

estructuras con menores valores de energía, se observó que la precisión de las soluciones obtenidas decrece notablemente. Esto indica que las funciones de energía aún no logran modelar de manera correcta el empaquetamiento de la cadena lateral, y lo mismo ocurre en el diseño de proteínas [24].

II. EVALUACIÓN DE FUNCIONES DE ENERGÍA

Considerando al PSCPP como un problema de optimización, la función de energía es la función objetivo que debe minimizarse. La función de energía es un modelo que aproxima la energía resultante de las interacciones entre los átomos de una proteína, es decir, su energía libre. Existen dos tipos principales de funciones de energía [20], [24]: (i) las *basadas en la física* (*physical effective energy function* - PEEF) y (ii) las *basadas en el conocimiento* (*statistical effective energy function* - SEEF).

Las funciones de energía deben ser lo suficientemente precisas para capturar las características más importantes de las interacciones atómicas que ocurren en la proteína; y al mismo tiempo, deben tener una implementación computacional rápida [13], [5], [36]. Una de las propiedades que se espera de una función de energía efectiva es su habilidad de reconocer a la estructura nativa, y poder usarse entonces en la predicción de estructuras y el diseño de proteínas.

Debido a las limitaciones de las funciones de energía existentes, el diseño o rediseño de funciones de energía es un área de constante atención en biocomputación. Las funciones de energía basadas en la física han cambiado muy poco en las dos últimas décadas [6], [24]; mientras que para las basadas en conocimiento constantemente hay nuevas propuestas. Esto puede deberse en cierta medida a la dificultad y costo de realizar las pruebas experimentales que requieren las funciones basadas en la física. Como ambos modelos tienen sus ventajas y desventajas, varias propuestas se centran en la fusión de los mismos. Éstas utilizan generalmente a CHARMM o a AMBER para los componentes físicos, mientras que la parte de conocimiento se optimiza para la aplicación que se está tratando.

Para la evaluación de una función de energía para el PSCPP, podrían usarse las siguientes metodologías:

- Considerar un algoritmo de búsqueda y resolver el PSCPP, comparando los resultados con los del estado del arte. Este método se usa, por ejemplo, en las propuestas de SCMF [18] y SIDEpro [28]. Se debe resaltar que este método es dependiente del algoritmo de búsqueda empleado, el cual puede ser exacto o aproximado.
- Para varias aplicaciones, no es necesario que se produzca correctamente la superficie de energía¹, sino que la estructura nativa tenga el menor valor de energía. Por ello se utiliza un conjunto de (*estructuras señuelo*), que consisten de perturbaciones ligeras de la estructura nativa; y la función de energía debe reconocer la estructura nativa entre dichos señuelos. Esta metodología es la más utilizada, encontrándose en las propuestas de OPUS-PSP [25], DOPE [33], I-TASSER [39], [38], y Rosetta [32].

- Aplicar un algoritmo de búsqueda local teniendo como entrada a la estructura nativa de la proteína [31]. Como dicha estructura debe ser al menos un mínimo local en la función de energía, la búsqueda local no debería hacer cambios en la estructura inicial.

En el presente trabajo se pretende usar el tercer método para la evaluación de las funciones propuestas, para luego aplicarlo al PSCPP. Una ventaja de usar al PSCPP como caso de prueba de las funciones de energía es que los métodos propuestos para resolverlo pueden evaluarse *in silico*.

III. DESCRIPCIÓN DEL MÉTODO

Se utilizó un conjunto de 358 proteínas extraídas del PDB, y la biblioteca de rotámeros independiente de la columna vertebral generada por Dunbrack y Cohen [11] en su versión de mayo del 2002. Las pruebas (codificadas en Java) fueron ejecutadas en una laptop Acer Aspire-V5 con un procesador Intel Core i7-4500U de 4 núcleos a 1.8GHz, 12GB de memoria, y Windows 8.1 de 64 bits.

La función de energía evaluada es la siguiente:

$$E(S) = \sum_i^{N-1} \sum_{j=i+1}^N E_{EV}(r_i, r_j) + \sum_i^N E_{RSCE}(r_i)$$

donde S es la estructura considerada, N es la cantidad de residuos de la proteína, y r_i es el rotámero del residuo i . Esta función de energía cuenta con dos términos: (a) el primero se denomina *Excluded Volume* (EV), el cual es un potencial que crea una gran repulsión entre los átomos cuando éstos se acercan a una cierta distancia definida por sus radios de van der Waals; (b) el segundo término es un potencial de *energía de Ramachandran de la cadena lateral* (RSCE), el cual asigna un valor de energía según la probabilidad de encontrar los ángulos de torsión de cada aminoácido. Los parámetros de ambos términos fueron obtenidos del algoritmo MESHI [17].

El término EV entre una pareja de átomos se obtiene de la siguiente forma:

$$E_{EV}(a_i, a_j) = \begin{cases} k\delta^4 & \text{si } a_i \text{ o } a_j \text{ pertenecen a la columna vertebral,} \\ \delta^2 & \text{si } a_i \text{ y } a_j \text{ pertenecen a una cadena lateral,} \\ 0 & \text{si } d > \delta. \end{cases}$$

Se tiene que $\delta = d - \sigma$, donde d es la distancia entre el átomo a_i y el átomo a_j , y σ es un parámetro que se asigna de acuerdo al tipo de átomo de a_i y a_j , así como al aminoácido al que pertenezcan. El término EV es distinto al potencial de Van der Waals, pues no considera ningún potencial de atracción entre átomos, teniendo en cuenta únicamente la repulsión entre los mismos.

El término RSCE se obtuvo utilizando interpolación de splines [34], de tal forma a suavizar los valores de una base de datos que contiene el análisis estadístico de los ángulos de torsión de la columna vertebral y cadena lateral, observados en muchas estructuras de proteínas. Para calcular el valor de energía de RSCE de un aminoácido se requieren los ángulos de torsión de la columna vertebral ϕ y ψ , así como todos los ángulos de torsión de la cadena lateral de dicho aminoácido

¹valores de energía libre de todas las estructuras

[17]. El valor de energía que se obtiene intenta aproximar las siguientes probabilidades de obtener dichos ángulos de torsión:

$$E_{RSC E}(r_i) = \begin{cases} p(\phi_i, \psi_i) & \text{si } i_\chi = 0 \\ p(\phi_i, \psi_i, \chi_1^i) & \text{si } i_\chi = 1 \\ p(\phi_i, \psi_i, \chi_1^i) * p(\chi_2^i | \chi_1^i) & \text{si } i_\chi = 2 \\ p(\phi_i, \psi_i, \chi_1^i) * p(\chi_2^i | \chi_1^i) * p(\chi_3^i | \chi_1^i) & \text{si } i_\chi = 3 \\ p(\phi_i, \psi_i, \chi_1^i) * p(\chi_2^i | \chi_1^i) * p(\chi_3^i | \chi_1^i) * p(\chi_4^i | \chi_1^i) & \text{si } i_\chi \geq 4 \end{cases}$$

En donde i_χ representa el número de ángulos de torsión de la cadena lateral del aminoácido i . $p(\phi_i, \psi_i, \chi_1^i)$ representa la probabilidad conjunta de que el aminoácido i tenga los ángulos de torsión de la columna vertebral ϕ_i y ψ_i , así como el ángulo de torsión de la cadena lateral χ_1^i . $p(\chi_n^i | \chi_1^i)$ representa la probabilidad condicional de que el aminoácido i tenga el valor χ_n^i como el n -ésimo ángulo de torsión de la cadena lateral, dado el valor χ_1^i para el primer ángulo de torsión.

III-A. Estructura inicial y algoritmo de búsqueda local

Debido al uso de una biblioteca de rotámeros para la selección de las conformaciones de la cadena lateral de los aminoácidos de una proteína, puede ser imposible crear una estructura idéntica a la presente en el PDB; por ello se utiliza la estructura *Best* de dicha proteína. La misma se construye seleccionando para cada aminoácido el rotámero que mejor se adapte a los ángulos de torsión de la proteína en el archivo *pdb*.

Se aplica el método de búsqueda local a estructuras *Best*, siendo éste guiado por la función de energía a evaluar. El objetivo principal de esta prueba es comprobar si la estructura *Best* es al menos un óptimo local para la función de energía empleada. El algoritmo de búsqueda local es el siguiente:

Algoritmo 1 Método de Búsqueda Local

Entrada: Estructura S_{Best} .

Salida: Estructura con un óptimo local de energía.

```

1:  $S = S_{Best}$ 
2:  $menorEnergia = energia(S)$ 
3: repetir
4:    $fin = Verdadero$ 
5:   para cada residuo  $i$  hacer
6:     para cada rotámero  $r$  de  $i$  hacer
7:       Asignar temporalmente el rotámero  $r$  al residuo  $i$ ,
         obteniendo  $S_r$ 
8:        $energia = energia(S_r)$ 
9:       si  $energia < menorEnergia$  entonces
10:         $menorEnergia = energia$ 
11:         $S = S_r$ 
12:         $fin = Falso$ 
13:     fin si
14:   fin para
15: fin para
16: hasta que  $fin == Verdadero$ 
17: devolver  $S$ 

```

IV. RESULTADOS Y CONCLUSIONES

Los resultados obtenidos se muestran en la Tabla II. Los valores mostrados son promedios simples y ponderados (entre

paréntesis), considerando la cantidad de residuos de cada proteína del conjunto de prueba. Para el tiempo se tiene en cuenta únicamente el proceso de búsqueda local. El porcentaje de reducción de energía expresa la relación entre la disminución de energía y la energía de la estructura inicial.

Tabla II. RESULTADOS DEL MÉTODO DE BÚSQUEDA LOCAL, CONSIDERANDO A LA ESTRUCTURA *Best* COMO ENTRADA.

Parámetros	<i>Best</i>	<i>Best + BusqLoc</i>
χ_1 (%)	99.65 (99.64)	81.12 (81.45)
χ_{1+2} (%)	93.25 (93.28)	61.98 (62.27)
Colisiones	0.38 (0.54)	0.74 (1.09)
Energía	8604.72 (11336.69)	2574.35 (3430.72)
Reducción de energía(%)		46.93 (50.19)
Tiempo(ms)		20.00 (36.41)
RMSD		1.804 (1.792)

En relación a la precisión, puede notarse que la misma se redujo considerablemente, teniéndose una variación alrededor de 18 % para χ_1 (%), y alrededor de 31 % para χ_{1+2} (%). En cuanto a la energía, se observa que ésta se reduce en casi un 47 % (en promedio) de su valor original. Considerando que la estructura *Best* es bastante cercana a la estructura nativa y que ésta debería ser al menos un mínimo local de la función energía evaluada, esta última no se acerca al modelo real, ya que existen importantes variaciones tanto en precisión como en valores de energía.

Se realizó una segunda prueba, considerando en este caso una estructura inicial con rotámeros elegidos aleatoriamente para cada residuo de la proteína. Los resultados se muestran en la Tabla III.

Tabla III. RESULTADOS DEL MÉTODO DE BÚSQUEDA LOCAL, CONSIDERANDO A LA ESTRUCTURA *Random* COMO ENTRADA.

Parámetros	<i>Random</i>	<i>Random + BusqLoc</i>
χ_1 (%)	34.89 (34.72)	80.08 (80.58)
χ_{1+2} (%)	14.87 (14.63)	60.62 (61.14)
Colisiones	38.79 (51.50)	0.76 (1.15)
Energía	513193.88 (685003.56)	2574.47 (3430.88)
Reducción de energía(%)		99.37 (99.40)
Tiempo(s)		25.23 (44.24)
RMSD		1.8316 (1.8167)

A pesar de que en la primera prueba se inició con una estructura muy cercana a la nativa, y que el método de búsqueda local empleado es simple; se observa que la precisión y la energía final promedio es bastante cercana para ambas pruebas. Teniendo en cuenta esta cercanía de los resultados, y un porcentaje mayor a 99 % de reducción promedio de la energía de la estructura aleatoria inicial, se puede concluir que la función de energía empleada es la mayor limitante para el PSCPP y no el método de búsqueda elegido.

Como trabajo futuro se propone realizar estas mismas pruebas con funciones de energía de los métodos del estado del arte para el PSCPP.

V. AGRADECIMIENTOS

Este trabajo fue apoyado por el Consejo Nacional de Ciencia y Tecnología bajo el proyecto SEP-CONACYT-CB-2010-154737.

REFERENCIAS

- [1] T. Akutsu, "NP-hardness results for protein side-chain packing," *Genome Informatics*, vol. 8, pp. 180–186, 1997.
- [2] E. Benedetti, G. Morelli, G. Nemethy, y H. Scheraga, "Statistical and energetic analysis of sidechain conformations in oligopeptides," *International Journal of Peptide and Protein Research*, vol. 22, pp. 1–15, 1983.
- [3] T. Bhat, V. Sasisekharan, y M. Vijayan, "An analysis of side-chain conformation in proteins," *International Journal of Peptide and Protein Research*, vol. 13, no. 2, pp. 170–184, 1978.
- [4] T. Blundell, B. Sibanda, M. Sternberg, y J. Thornton, "Knowledge-based prediction of protein structures," *Nature*, vol. 326, no. 26, pp. 347–352, 1987.
- [5] F. Boas y P. Harbury, "Potential energy functions for protein design," *Current Opinion in Structural Biology*, vol. 17, no. 2, pp. 199–204, 2007.
- [6] B. Brooks, C. Brooks, A. MacKerell, L. Nilsson, R. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoseck, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. Pastor, C. Post, J. Pu, M. Schaefer, B. Tidor, R. Venable, H. Woodcock, X. Wu, W. Yang, D. York, y M. Karplus, "CHARMM: The biomolecular simulation program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [7] Y. Cao, L. Song, Z. Miao, Y. Hu, L. Tian, y T. Jiang, "Improved side-chain modeling by coupling clash-detection guided iterative search with rotamer relaxation," *Bioinformatics*, vol. 27, no. 6, pp. 785–790, 2011.
- [8] R. Chandrasekaran y G. Ramachandran, "Studies on the conformation of amino acids. XI. Analysis of the observed side group conformations in proteins," *International Journal of Protein Research*, vol. 2, pp. 223–233, 1970.
- [9] I. Corona, "Análisis comparativo de dos heurísticas para el problema de empaquetamiento de la cadena lateral en proteínas," Tesis de Maestría, CICESE, 2010.
- [10] R. Dunbrack, "Rotamer libraries in the 21st century," *Current Opinion in Structural Biology*, vol. 12, no. 4, pp. 431–440, 2002.
- [11] R. Dunbrack y F. Cohen, "Bayesian statistical analysis of protein side-chain/rotamer preferences," *Protein Science*, vol. 6, no. 8, pp. 1661–1681, 1997.
- [12] R. Dunbrack y M. Karplus, "Backbone-dependent Rotamer Library for Proteins Application to Side-chain Prediction," *Journal of Molecular Biology*, vol. 230, no. 2, pp. 543–574, 1993.
- [13] D. Gordon, S. Marshall, y S. Mayo, "Energy functions for protein design," *Current Opinion in Structural Biology*, vol. 9, no. 4, pp. 509–513, 1999.
- [14] M. James y A. Sielecki, "Structure and refinement of penicillopepsin at 1.8 Å resolution," *Journal of Molecular Biology*, vol. 163, no. 2, pp. 299–361, 1983.
- [15] J. Janin y S. Wodak, "Conformation of amino acid side-chains in proteins," *Journal of Molecular Biology*, vol. 125, no. 3, pp. 357–386, 1978.
- [16] T. Jones y S. Thirup, "Using known substructures in protein model building and crystallography," *The EMBO Journal*, vol. 5, no. 4, pp. 819–822, 1986.
- [17] N. Kalisman, A. Levi, T. Maximova, D. Reshef, S. Zafirri-Lynn, Y. Gleyzer, y C. Keasar, "MESH: a new library of Java classes for molecular modeling," *Bioinformatics*, vol. 21, no. 20, pp. 3931–3932, 2005.
- [18] P. Koehl, H. Orland, y M. Delarue, "Adapting Poisson-Boltzmann to the self-consistent mean field theory: Application to protein side-chain modeling," *The Journal of Chemical Physics*, vol. 135, no. 5, pp. 055104–11, 2011.
- [19] G. Krivov, M. Shapovalov, y R. Dunbrack, "Improved prediction of protein side-chain conformations with SCWRL4," *Proteins*, vol. 77, no. 4, pp. 778–795, 2009.
- [20] T. Lazaridis y M. Karplus, "Effective energy functions for protein structure prediction," *Current Opinion in Structural Biology*, vol. 10, no. 2, pp. 139–145, 2000.
- [21] C. Lee y S. Subbiah, "Prediction of protein side-chain conformation by packing optimization," *Journal of Molecular Biology*, vol. 217, no. 2, pp. 373–388, 1991.
- [22] C. Levinthal, "Are there pathways for protein folding?" *Journal de Chimie Physique et de Physico-Chimie Biologique*, vol. 65, pp. 44–45, 1968.
- [23] C. Lezcano, "Problema de empaquetamiento de la cadena lateral de proteínas: Análisis de la calidad de soluciones de una biblioteca de rotámeros simple," Tesis de Maestría, Universidad Nacional de Asunción, 2012.
- [24] Z. Li, Y. Yang, J. Zhan, L. Dai, y Y. Zhou, "Energy Functions in De Novo Protein Design: Current Challenges and Future Prospects," *Annual Review of Biophysics*, vol. 42, pp. 315–335, 2013.
- [25] M. Lu, A. Dousis, y J. Ma, "OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing," *Journal of Molecular Biology*, vol. 376, no. 1, pp. 288–301, 2008.
- [26] —, "OPUS-Rota: A fast and accurate method for side-chain modeling," *Protein Science*, vol. 17, no. 9, pp. 1576–1585, 2008.
- [27] Z. Miao, Y. Cao, y T. Jiang, "RASP: rapid modeling of protein side chain conformations," *Bioinformatics*, vol. 27, no. 22, pp. 3117–3122, 2011.
- [28] K. Nagata, A. Randall, y P. Baldi, "SIDEpro: A novel machine learning approach for the fast and accurate prediction of side-chain conformations," *Proteins: Structure, Function, and Bioinformatics*, vol. 80, no. 1, pp. 142–153, 2012.
- [29] J. Ponder y F. Richards, "Tertiary templates for proteins: use of packing criteria in the enumeration of allowed sequences for different structural classes," *Journal of Molecular Biology*, vol. 193, pp. 775–792, 1987.
- [30] F. Richards, "Areas, Volumes, Packing, and Protein Structure," *Annual Review of Biophysics and Bioengineering*, vol. 6, pp. 151–176, 1977.
- [31] D. Rodríguez, "Diseño de heurísticas para el problema de empaquetamiento de la cadena lateral en proteínas," Tesis de Maestría, CICESE, 2014.
- [32] C. Rohl, C. Strauss, K. Misura, y D. Baker, "Protein structure prediction using rosetta," *Numerical Computer Methods, Part D, Methods in Enzymology*, vol. 383, pp. 66–93, 2004.
- [33] M. Shen y A. SaliProtein, "Statistical potential for assessment and prediction of protein structures," *Protein Science*, vol. 15, no. 11, pp. 2507–2524, 2006.
- [34] Y. Subbotin, "Piecewise-polynomial (spline) interpolation," *Mathematical Notes of the Academy of Sciences of the USSR*, vol. 1, no. 1, pp. 41–45, 1967.
- [35] N. Summers y M. Karplus, "Construction of side-chains in homology modelling: Application to the C-terminal lobe of rhizopuspepsin," *Journal of Molecular Biology*, vol. 210, no. 4, pp. 785–811, 1989.
- [36] M. Suárez y A. Jaramillo, "Challenges in the computational design of proteins," *Journal of the Royal Society, Interface*, vol. 6, no. 4, pp. S477–S491, 2009.
- [37] J. Wendoloski y F. Salemme, "PROBIT: A statistical approach to modeling proteins from partial coordinate data using substructure libraries," *Journal of Molecular Graphics*, vol. 10, no. 2, pp. 124–126, 1992.
- [38] P. Widera, J. Garibaldi, y N. Krasnogor, "GP challenge: evolving energy function for protein structure prediction," *Genetic Programming and Evolvable Machines*, vol. 1, no. 11, pp. 61–88, 2009.
- [39] S. Wu, J. Skolnick, y Y. Zhang, "Ab initio modeling of small proteins by iterative TASSER simulations," *BMC Biology*, vol. 5, no. 17, pp. 1–10, 2007.