

## Resumen

GRAPHICPRO es una aplicación informática que permite graficar modelos de proteínas como apoyo a la investigación de profesionales de Bioinformática y Biología Computacional en la resolución de problemas relacionados al proceso de repliegado de las proteínas. Para el desarrollo de la herramienta, se consideró el modelo HP (Hidrofóbico-Polar) que reduce considerablemente la complejidad de las estructuras proteicas, aprovechando características físico-químicas de los aminoácidos que las conforman. Este modelo se aplica generalmente sobre reticulados que restringen el espacio de conformaciones a un grupo reducido de movimientos discretos. La funcionalidad más importante de GRAPHICPRO es la visualización de proteínas considerando los reticulados: cuadrado y triangular 2D, y cubo 3D, y las coordenadas de movimientos en el espacio: absolutas y relativas. Dados los siguientes datos: (i) una instancia de proteína según el modelo HP; (ii) una conformación espacial de la instancia codificada usando letras que indican la secuencia de movimientos en el espacio; (iii) el tipo de reticulado e (iv) el tipo de coordenadas; el programa muestra el gráfico correspondiente.

## Un Modelo para representar Proteínas

Una proteína es una secuencia lineal de aproximadamente 50 a 1000 unidades, denominadas aminoácidos, que bajo ciertas condiciones físicas se repliega conformando una estructura funcionalmente única llamada su estructura terciaria o estado nativo (ver Fig. 1). Este estado constituye a su vez la clave para comprender el funcionamiento de las proteínas dentro de un organismo vivo.

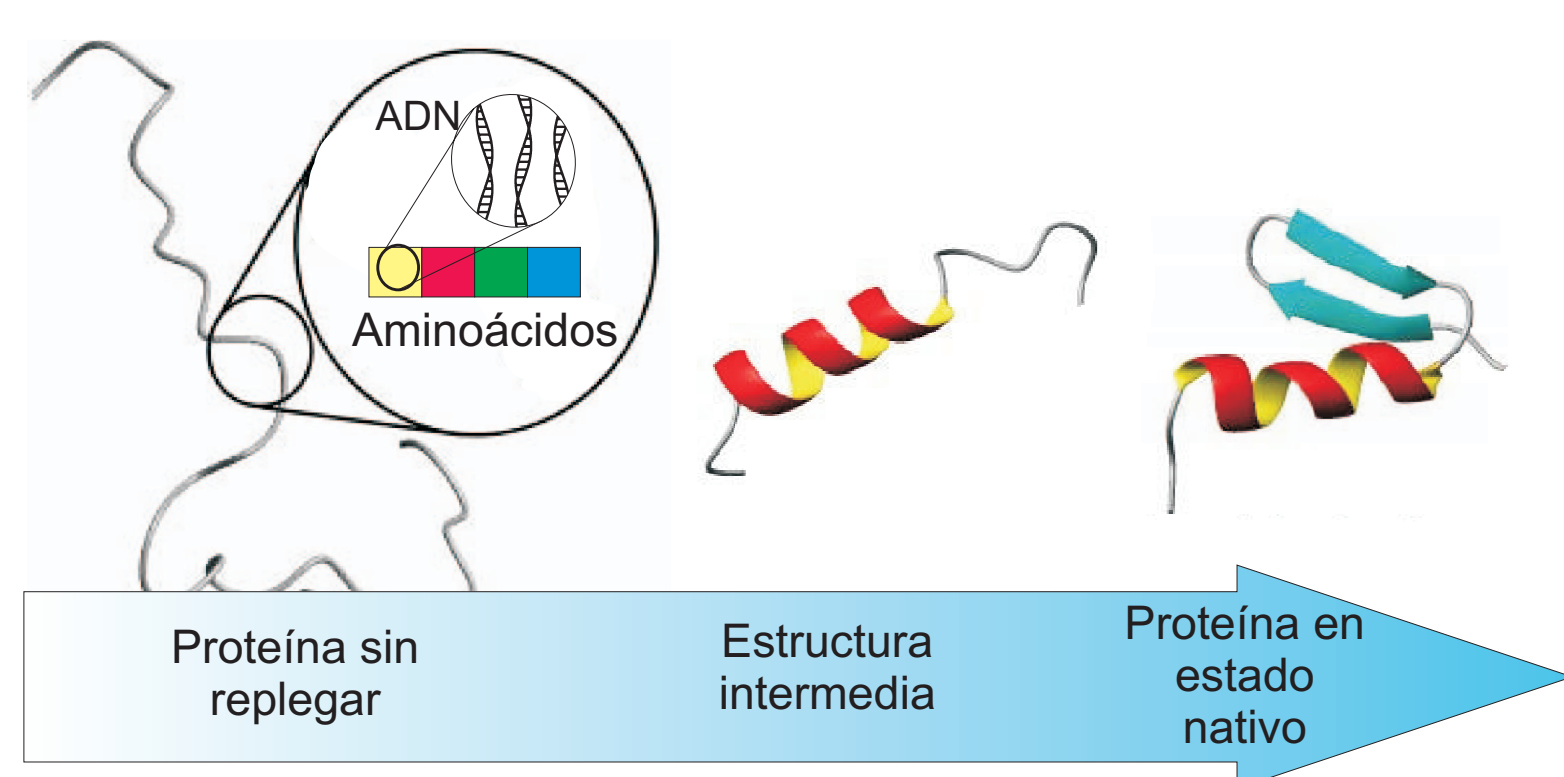


FIGURA 1: El proceso de repliegado de las proteínas

Las proteínas representan, aún en la actualidad, sistemas muy complejos para permitir una simulación exacta del proceso de repliegado. Por lo tanto, los métodos computacionales que permiten indagar tanto en las propiedades cuantitativas como cualitativas de las proteínas, deben ser realizados sobre modelos reducidos. En el modelo HP [1], los 20 aminoácidos naturales se clasifican en dos tipos: hidrofóbicos (H) e hidrofílicos o polares (P). Cada aminoácido se representa por una esfera o cuadrado. El espacio de conformaciones válidas está restringido a caminos que no se entrecruzan en un reticulado específico (cuadrado, triangular, cúbico, diamante, etc.), con cada aminoácido alojado en un vértice. Para designar los movimientos sobre el reticulado se pueden considerar coordenadas absolutas o relativas. Las interacciones H-H que se generan al replugar la proteína sobre el reticulado son las que se consideran para calcular la energía de una instancia, aportando cada una un valor de  $-1$ . Aunque este modelo es muy sencillo captura las características esenciales del repliegado de proteínas.

La figura 2 muestra dos ejemplos de la aplicación de este modelo en un reticulado cuadrado, 2D, utilizando coordenadas absolutas y relativas. Explicaremos a continuación la representación en coordenadas relativas. La

estructura terciaria de la instancia consiste en un vector  $c \in \{L, R, F\}^{n-2}$  que contiene una codificación formada por  $L$ ,  $R$  y  $F$  para denotar un giro a la izquierda (Left), giro a la derecha (Right) y hacia delante (Forward), respectivamente. Dada una cadena  $s \in \{H, P\}^n$ , el aminoácido  $i + 1$  ( $i = 2, \dots, n$ ) se coloca usando alguno de éstos términos relativos a la dirección del enlace de los aminoácidos  $i - 1$  e  $i$ . Los dos primeros aminoácidos ( $i = 1, 2$ ) se mantienen fijos para definir el sistema de coordenadas relativas. La proteína del ejemplo tiene una longitud de 20 aminoácidos y está dada por HPHPPHHPHPPHPPHPPH (designando los H's con cuadros negros), y una conformación RFRLLRLRRFRLLRRFR (conocida como óptima para esta instancia), formando una energía de  $-9$ , es decir, con 9 enlaces H-H (líneas puntuadas).

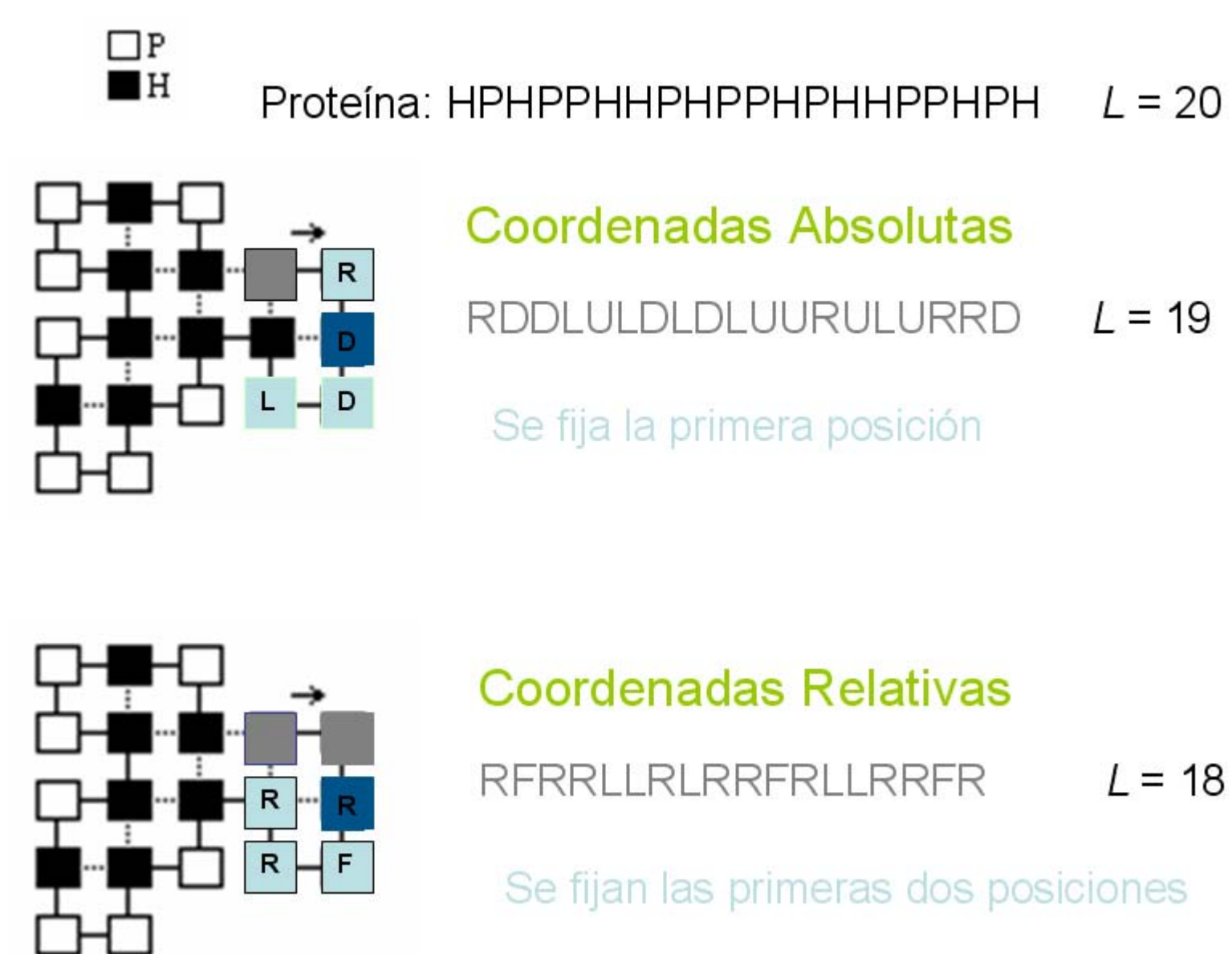


FIGURA 2: Ejemplos de representación del repliegado de una instancia HP

Existen varios problemas abiertos relacionados con el repliegado de las proteínas. Uno de ellos, el problema de predicción de la estructura terciaria, consiste en predecir la configuración terciaria o estado nativo que tendrá la proteína, dada su secuencia primaria. Debido a que el espacio de posibles conformaciones puede resultar bastante grande para proteínas de más de 25 aminoácidos (y crece exponencialmente a medida que la longitud de la proteína es mayor) la mayoría de los algoritmos que se han propuesto sólo obtienen soluciones aproximadas. Precisamente, la idea de este trabajo fue desarrollar una herramienta que permitiera visualizar los resultados intermedios o finales que generan estos algoritmos. Aun cuando existen aplicaciones gráficas de gran potencia para visualizar proteínas, como Rasmol y Chime, se requería una herramienta más sencilla que soportara el modelo HP, ya que es uno de los modelos más utilizados para tratar computacionalmente el problema de predicción de la estructura terciaria de las proteínas.

## Características de la herramienta gráfica

La herramienta GRAPHICPRO está basada en la filosofía de orientación al objeto: la especificación del diseño fue realizada utilizando diagramas UML (Lenguaje de Modelado Unificado) y el código se desarrolló usando el Lenguaje de Programación C/C++ y funciones gráficas de la librería OpenGL. A continuación se listan las características más destacadas de GRAPHICPRO :

- Presenta en una interfaz gráfica los elementos que permiten al usuario navegar entre las funcionalidades disponibles e introducir datos fácilmente (ver Figuras 3 y 4).



FIGURA 3: Menú Nuevo de la herramienta gráfica

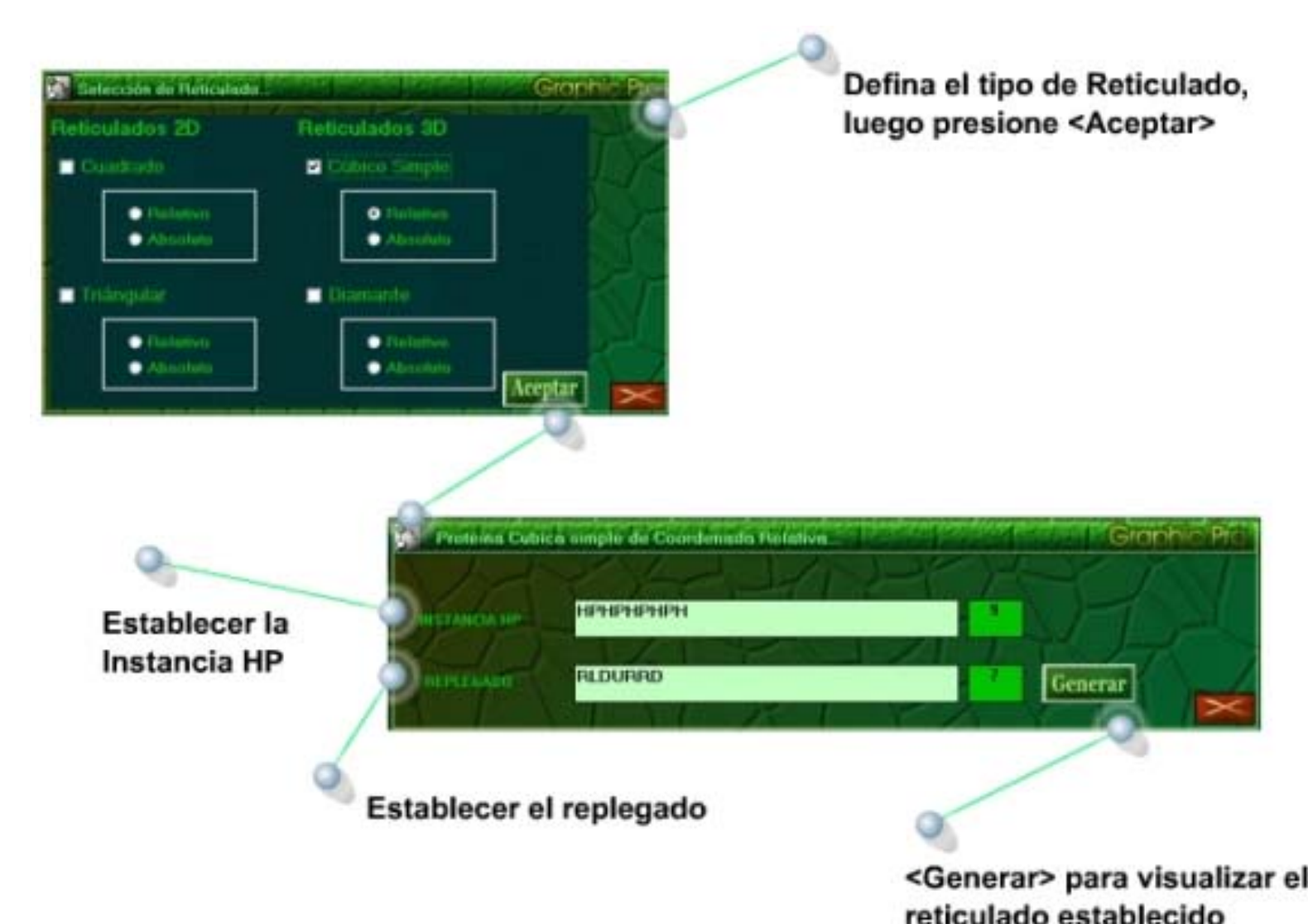


FIGURA 4: Pantalla que permite capturar los datos de entrada

- Permite visualizar diferentes representaciones de la estructura de instancias HP en 2 y 3 dimensiones y aplicar sobre ellas diferentes esquemas y patrones de color. (ver Figuras 5, 6 y 7)



FIGURA 5: Gráfico generado para una proteína de secuencia primaria aleatoria

- Rotar, desplazar y cambiar el tamaño de la representación en pantalla de la proteína.
- Además del gráfico, se presenta el número de enlaces H-H, que representa la energía del repliegado según el modelo HP, así como también el número de cruces que marcan puntos de intersección en conformaciones no válidas.
- Cuenta con una interfaz que le permite al usuario almacenar y recuperar datos acerca de las proteínas graficadas.

- Adicionalmente, el programa puede convertir la secuencia primaria (de aminoácidos) de una proteína real en su codificación según el modelo HP, de acuerdo a la configuración que establezca el usuario.
- GRAPHICPRO cuenta además con una ayuda gráfica que explica su utilización.



FIGURA 6: Instancia de prueba sobre un reticulado cúbico 3D

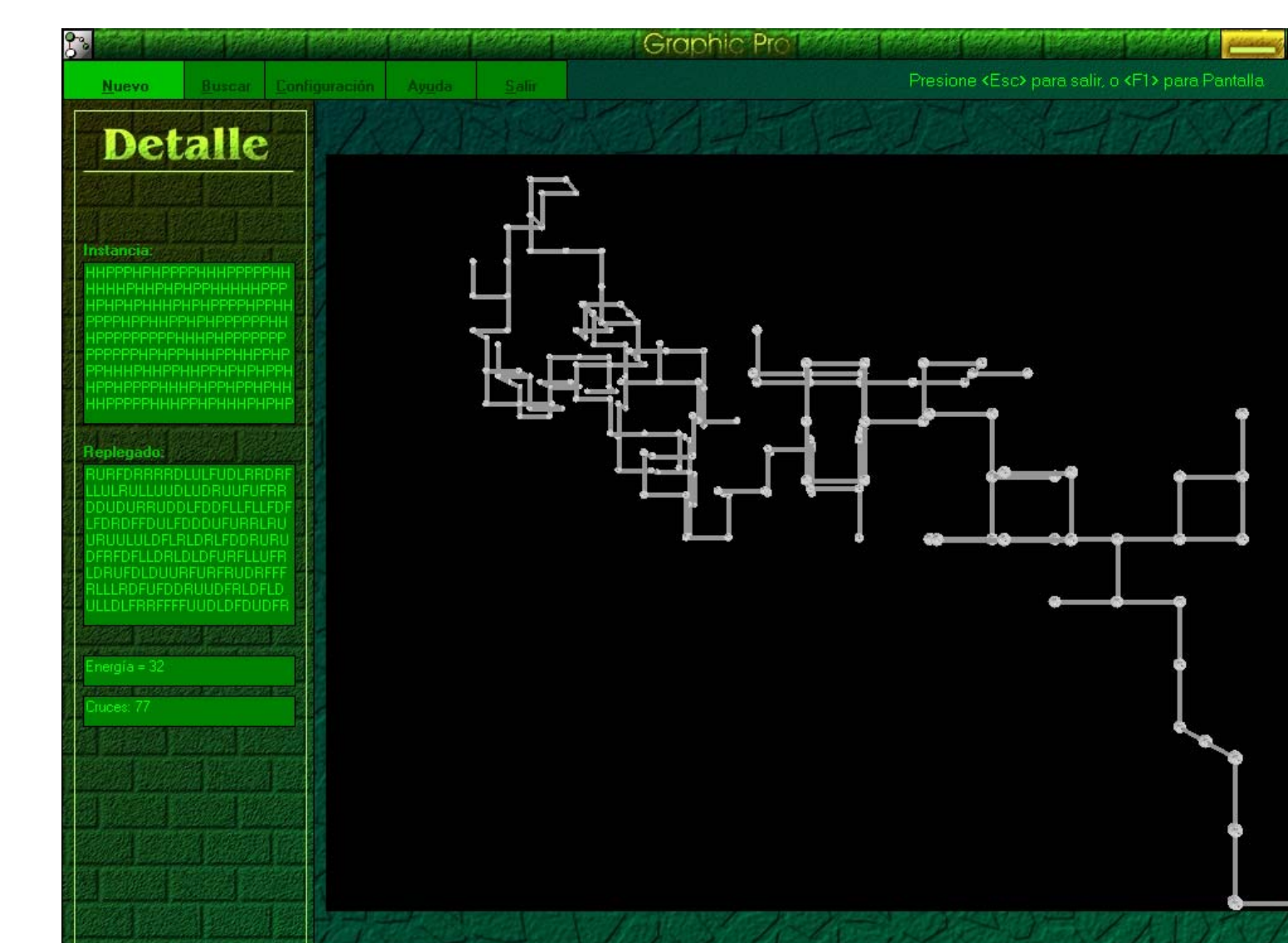


FIGURA 7: Vista de una instancia donde se presentan sólo las interacciones entre aminoácidos

## Conclusiones

GRAPHICPRO es una aplicación informática de ambiente gráfico que toma los datos de las estructuras primarias de proteínas según el modelo HP y las conformaciones de estructuras terciarias, y muestra los resultados en reticulados de 2 y 3 dimensiones. Este software no realiza predicciones de las conformaciones que presentan las proteínas repliegadas, sólo permite graficar los resultados obtenidos por otras aplicaciones. Esto permite a los investigadores concentrarse más en mejorar las estrategias de predicción de la estructura terciaria de las proteínas, en vez de tratar con el componente gráfico de la aplicación.

## Referencias

- [1] Ken A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24:1501, 1985.