

Evolution: Una infraestructura conceptual y computacional para el estudio y exploración de problemas de la biología estructural

Demostración de Operación

Pedro Pablo González Pérez

Arturo Rojo Domínguez

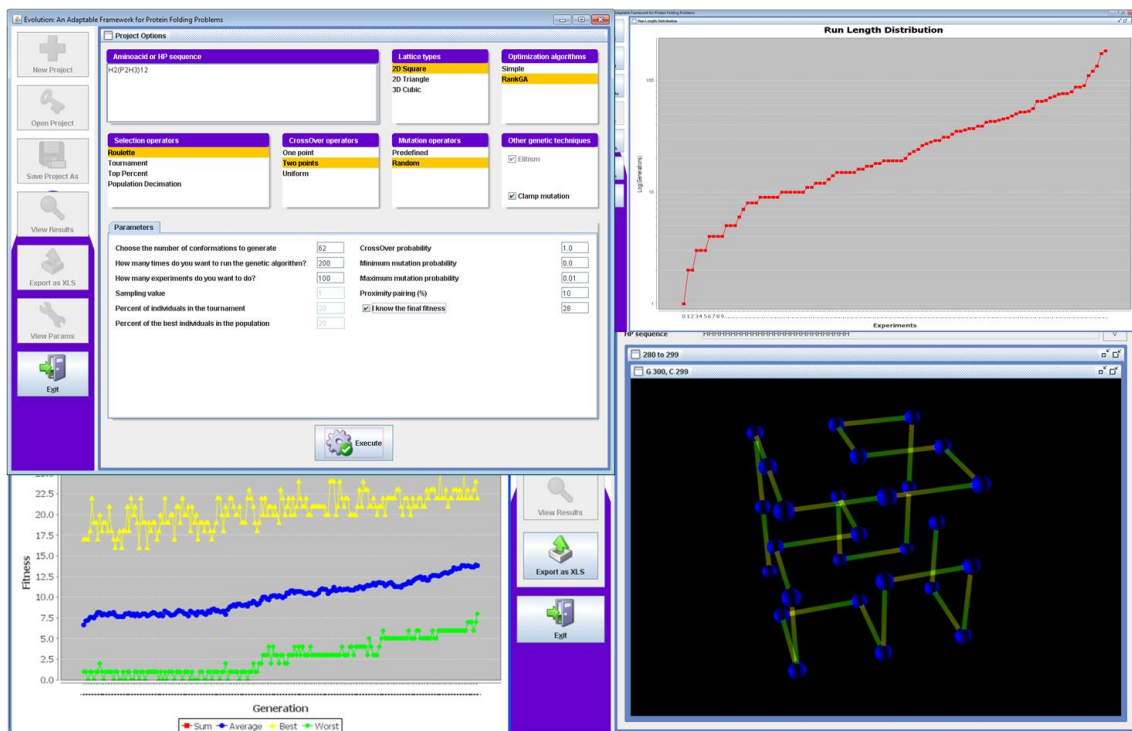
Hírám Isaac Beltrán Conde

Jorge Cervantes Ojeda

Máximo Eduardo Sánchez Gutiérrez

Versión 3.0 del 2010

Universidad Autónoma Metropolitana, Unidad Cuajimalpa



DEMOSTRACIÓN DE OPERACIÓN

Para mostrar las diferentes funcionalidades, flexibilidad y tipos de interacción con el usuario que proporciona la plataforma bioinformática Evolution, desarrollaremos a continuación dos experimentos representativos, a través de dos algoritmos evolutivos diferentes, los cuales nos permitirán familiarizarnos con sus interfaces gráficas, menús de opciones, visualización y manipulación 2D/3D y resúmenes tabulares.

Los experimentos tratan con la exploración y análisis de una secuencia 27mero completamente hidrofóbica, a través de dos técnicas de optimización, cada una de las cuales explota un tipo particular de algoritmo evolutivo. Como parte de los experimentos se mostrará la sucesión completa de interfaces gráficas, visualizaciones 2D/3D y resúmenes tabulares originados durante el desarrollo de los mismos, de forma tal que éstos constituyan una guía sobre cómo interactuar con la plataforma bioinformática.

Experimento: Exploración de las capacidades y flexibilidad de la plataforma bioinformática Evolution en el plegamiento de una secuencia 27mero hidrofóbica, cuya estructura tridimensional óptima viene dada por un cubo de tres capas, con 9 aminoácidos ubicados en cada capa y un valor de fitness igual a 28. .

Algoritmo Genético Simple

Secuencia: [H27](#)

Características del experimento

- **Algoritmo de optimización:** Simple
- **Tipo de representación:** 3D Cubic
- **Tamaño de la población de conformaciones:** 300
- **Número de generaciones para el algoritmo evolutivo:** 500
- **Número de experimentos en el proyecto:** 5
- **Operadores genéticos:** Selección según el método *Roulette selection*, Cruce en dos puntos
- **Otras técnicas genéticas a considerar:** Mutación aleatoria y Elitismo activado
- **Valor de “Sampling”:** 1
- **Probabilidad de cruce:** 1.0

En la figura 1 se puede apreciar la interfaz principal de la plataforma bioinformática Evolution. Desde esta primera interfaz se accede a un menú principal el cual proporciona al usuario las siguientes opciones:

- ***New Project***
- ***Open Project***
- ***Save Project As***
- ***View Results***
- ***Export as XLS***

- ***View Params***
- ***Exit***

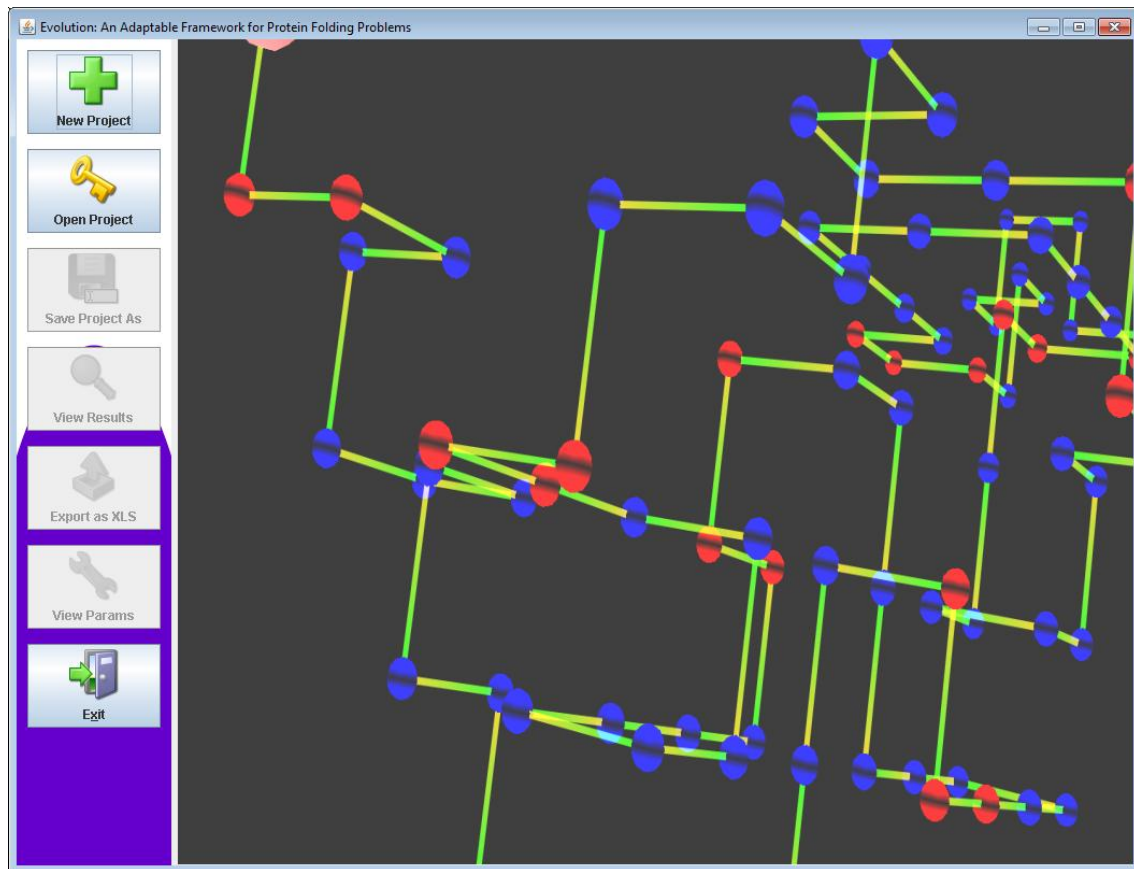


Figura 1. Interfaz y menú principal de la plataforma bioinformática Evolution.

El menú brinda al usuario todas las funcionalidades necesarias para la creación, almacenamiento, recuperación y exportación de proyectos. Éste menú es precisamente el que hace que la plataforma bioinformática Evolution se muestre al usuario como un verdadero Laboratorio Virtual, donde los experimentos pueden ser iniciados, desarrollados, almacenados, recuperados, exportados, modificados, reiniciados y vueltos a almacenar.

Cuando del menú se selecciona la opción *New Project*, entonces viene presentada al usuario una nueva interfaz, tal como la que se muestra en la figura 2, a través de la cual se introduce la secuencia o segmento de secuencia que se desea explorar y analizar, se selecciona el Algoritmo Genético (GA), se configuran los parámetros de dicho algoritmo y se ejecuta. La secuencia introducida puede ser formulada como una cadena de aminoácidos o como una cadena HP (donde la H denota un elemento no polar y la P uno polar).

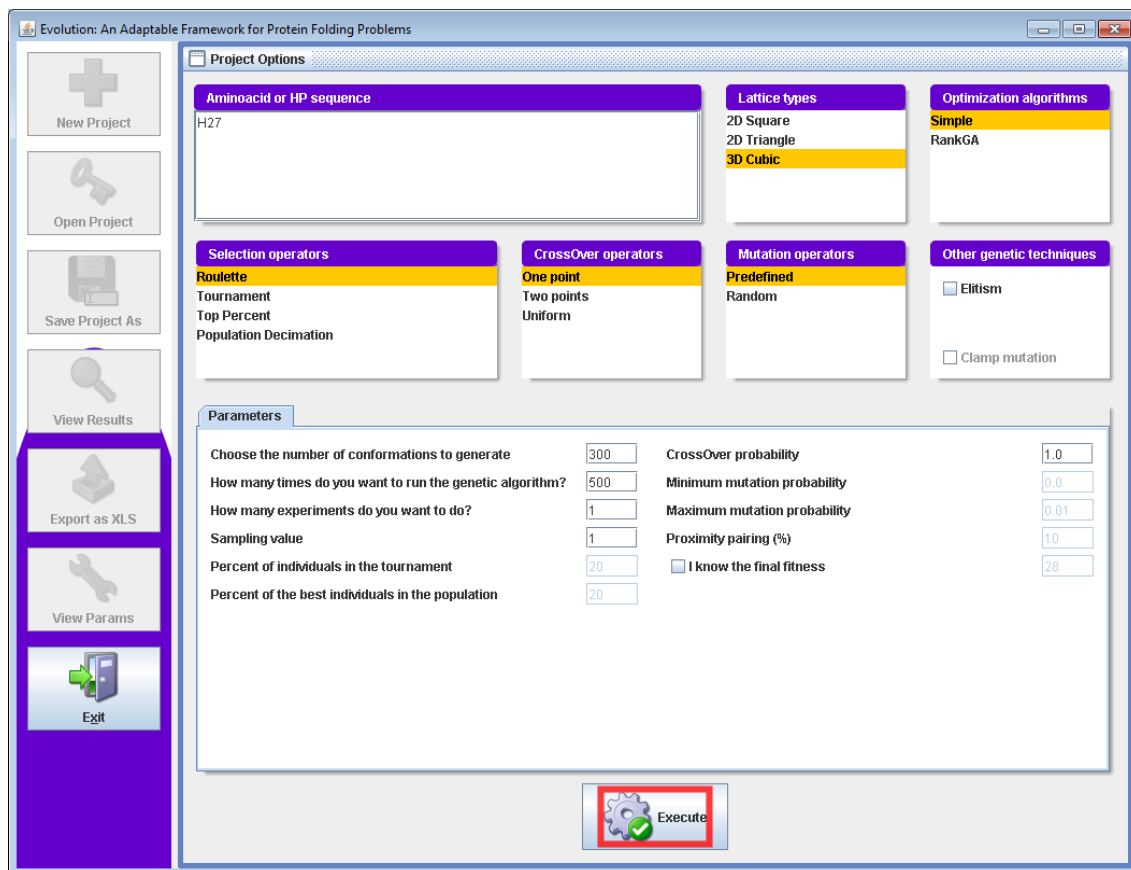


Figura 2. Interfaz para introducir la secuencia a explorar. Tal secuencia puede ser descrita en términos de aminoácidos o como secuencia HP, la herramienta detectará automáticamente de cuál se trata. Además, en esta interfaz, se pueden modificar los parámetros necesarios para la ejecución del Algoritmo Genético seleccionado.

La interfaz mostrada en la figura 2 permitirá al usuario seleccionar el modelo *lattice* 2D o 3D para generar la población inicial de conformaciones en el plano o en el espacio, respectivamente, y el algoritmo de aproximación para hacer evolucionar dicha población inicial hacia una generación de conformaciones cercanas al óptimo. Cabe señalar que en esta versión 2.4 de la plataforma bioinformática se ha agregado un algoritmo de aproximación denominado RankGA, el cual es un algoritmo evolutivo con una variada gama de operadores genéticos y técnicas genéticas a disposición del usuario. También debemos puntualizar que el criterio seguido para el proceso de optimización (función de *fitness*) es la energía mínima libre calculada entre elementos H que en la conformación generada resulten vecinos topológicos y no vecinos en la secuencia.

Como se puede apreciar en la figura 3, existen varios modelos Lattice disponibles al usuario para generar una población de conformaciones 2D/3D a partir de la secuencia HP introducida como entrada. En la actual versión de Evolution, los modelos Lattice disponibles son los siguientes:

- **2D Square**
- **2D Triangle**
- **3D Cubic**

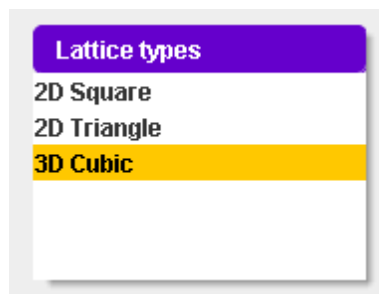


Figura 3. A través de la interfaz *New Project* se selecciona el modelo *Lattice* a ejecutar para generar la población de conformaciones aleatorias a partir de la secuencia HP proporcionada como entrada. En el experimento en ejecución se ha seleccionado el modelo *Lattice 3D Cubic*.

Después de haber seleccionado el modelo *Lattice* a ejecutar, se debe especificar el algoritmo de optimización que se va a emplear como se muestra en la figura 4.

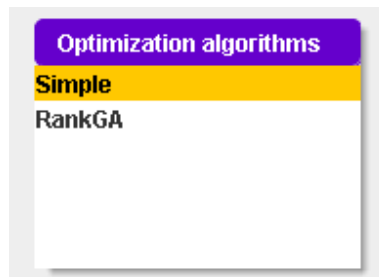


Figura 4. A través de la interfaz *New Project* se selecciona el algoritmo de optimización a ejecutar para generar la población de conformaciones. En el experimento en ejecución se ha seleccionado el *algoritmo Simple*.

Una vez seleccionado el algoritmo de optimización a ejecutar, se deben especificar los tipos de operadores y parámetros que caracterizarán al algoritmo, como por ejemplo el *Operador de Selección*, el *Operador de Cruce* y el *Operador de Mutación*, como se puede apreciar en la figura 5.

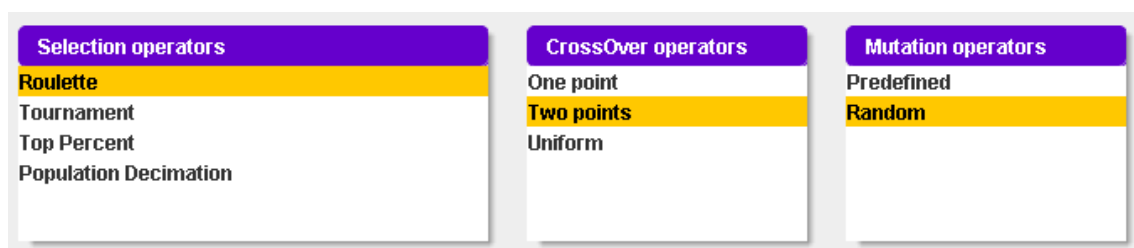


Figura 5. A través de la interfaz *New Project* se seleccionan los operadores de selección, cruce y mutación. En el experimento en ejecución se ha seleccionado el operador de selección *Roulette*, el operador de cruce *Two points* y el operador de mutación *Random*.

Además de la selección de los tipos de operadores, se pueden seleccionar otras técnicas genéticas como se ejemplifica en la figura 6.

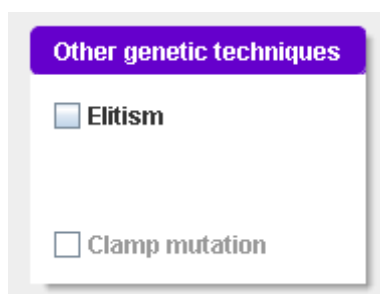


Figura 6. A través de la interfaz *New Project* se pueden seleccionar otras técnicas genéticas como el *Elitism* y la *Clamp mutation*.

Una vez seleccionados los tipos de operadores del algoritmo genético, se debe proporcionar el número de conformaciones con las que contará la generación inicial, el número de generaciones que se ejecutaran, el número de veces que será creada la población inicial (experimentos), el "Sampling" (o muestreo) de conformaciones y la probabilidad de cruce, tal como se muestra en la figura 7.

Parameters	
Choose the number of conformations to generate	300
How many times do you want to run the genetic algorithm?	500
How many experiments do you want to do?	1
Sampling value	1
Percent of individuals in the tournament	20
Percent of the best individuals in the population	20
CrossOver probability	1.0
Minimum mutation probability	0.0
Maximum mutation probability	0.01
Proximity pairing (%)	10
☒ I know the final fitness	28

Figura 7. A través de la interfaz *New Project* se permite ajustar los parámetros del algoritmo como; número de conformaciones, de generaciones, de experimentos, de "Sampling" y la probabilidad de cruce. En el experimento en ejecución se han asignado los valores 300, 500, 1, 1 y 1.0 respectivamente.

Una vez introducida la secuencia de aminoácidos o secuencia HP, habiendo seleccionado el tipo de lattice y configurado los operadores y parámetros del algoritmo genético seleccionado, se pulsa el botón *Execute* para iniciar el proceso de generación de la población inicial de conformaciones y la optimización de las estructuras moleculares que la componen.

Como se ilustra en la figura 8, la opción *View Results* en el menú principal se activa tan pronto el proceso antes descrito ha concluido. La selección de la opción *View Results* tras haber ejecutado el algoritmo genético producirá como resultado las visualizaciones y gráficos disponibles para su manipulación.

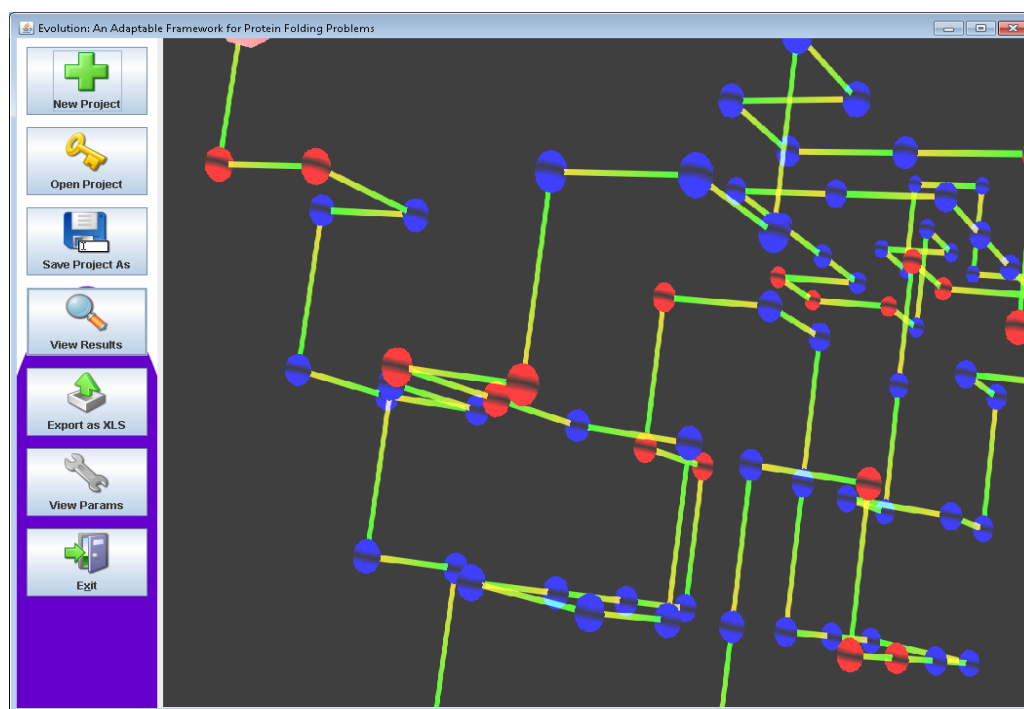


Figura 8. Activación de la opción View Results en el menú de la interfaz principal.

Como se muestra en las figuras 9, 10, 11 y 12, las opciones de visualización e interacción con el espacio conformacional son amplias.

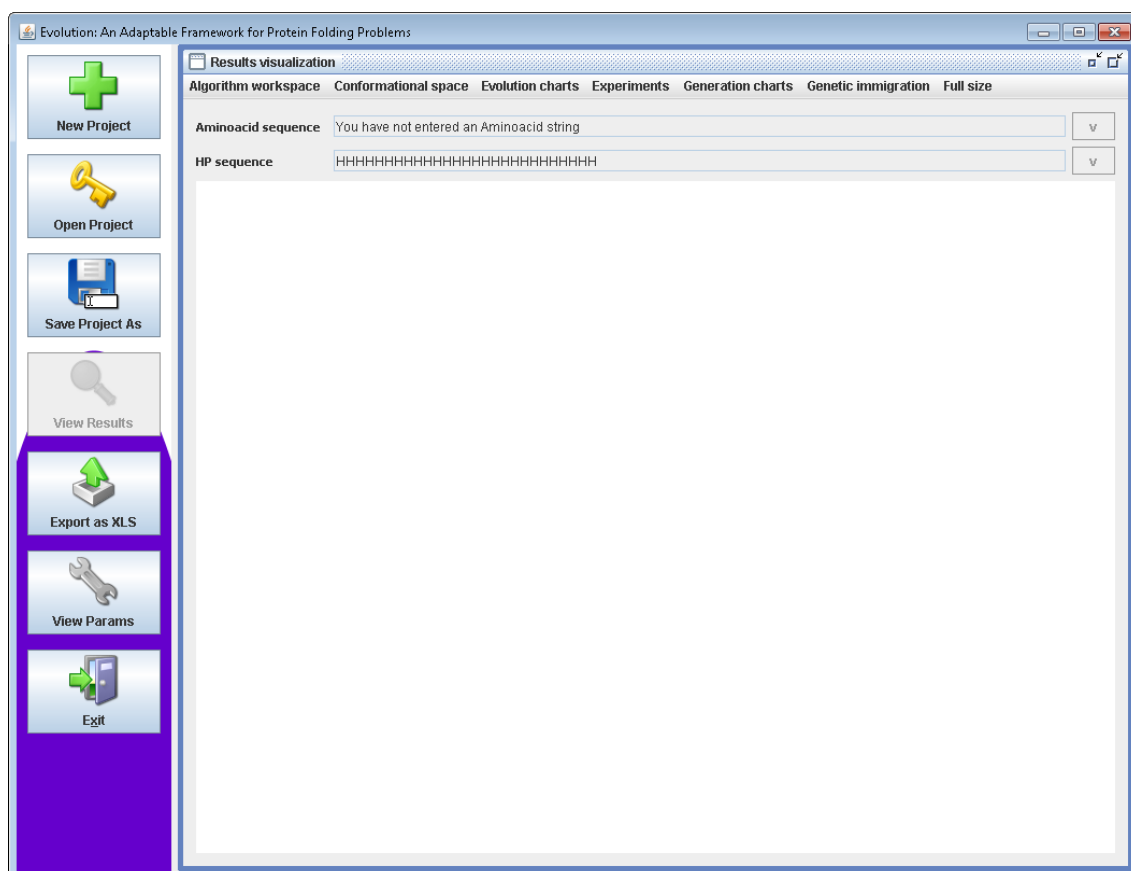


Figura 9. Interfaz *View Results*, a través de la cual resulta posible la visualización 2D/3D de las conformaciones que integran el espacio conformacional.

Para visualizar el espacio conformacional inicial se debe seleccionar el submenú *Show space* del menú *Conformational Space* en la interfaz *View Results* (figura 9). Posteriormente aparecerá la interfaz *Generation number* (figura 10), a través de la cual se deberá especificar el número de la generación a visualizar, que en este caso corresponderá a la generación 0.

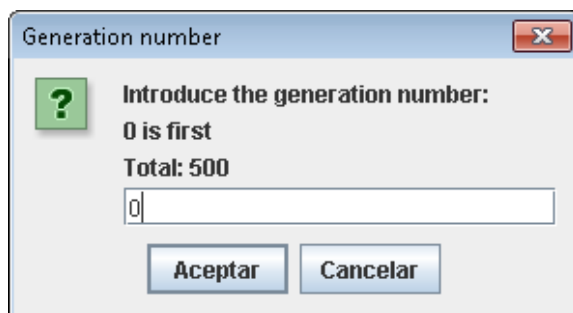


Figura 10. Interfaz *Generation number*, a través de la cual se especifica el número de la generación de la que se desean visualizar sus conformaciones. En este caso, seleccionaremos la generación inicial; 0.

Una vez presionado el botón *Aceptar* en la interfaz *Generation number*, se desplegarán en el lienzo de la interfaz *Results visualization* un conjunto de cajones (botones), cada uno de los cuales agrupa un número determinado de conformaciones (ver figura 12). Se abre el cajón deseado y se seleccionan las conformaciones que se desean visualizar sobre el lienzo (ver figura 12).

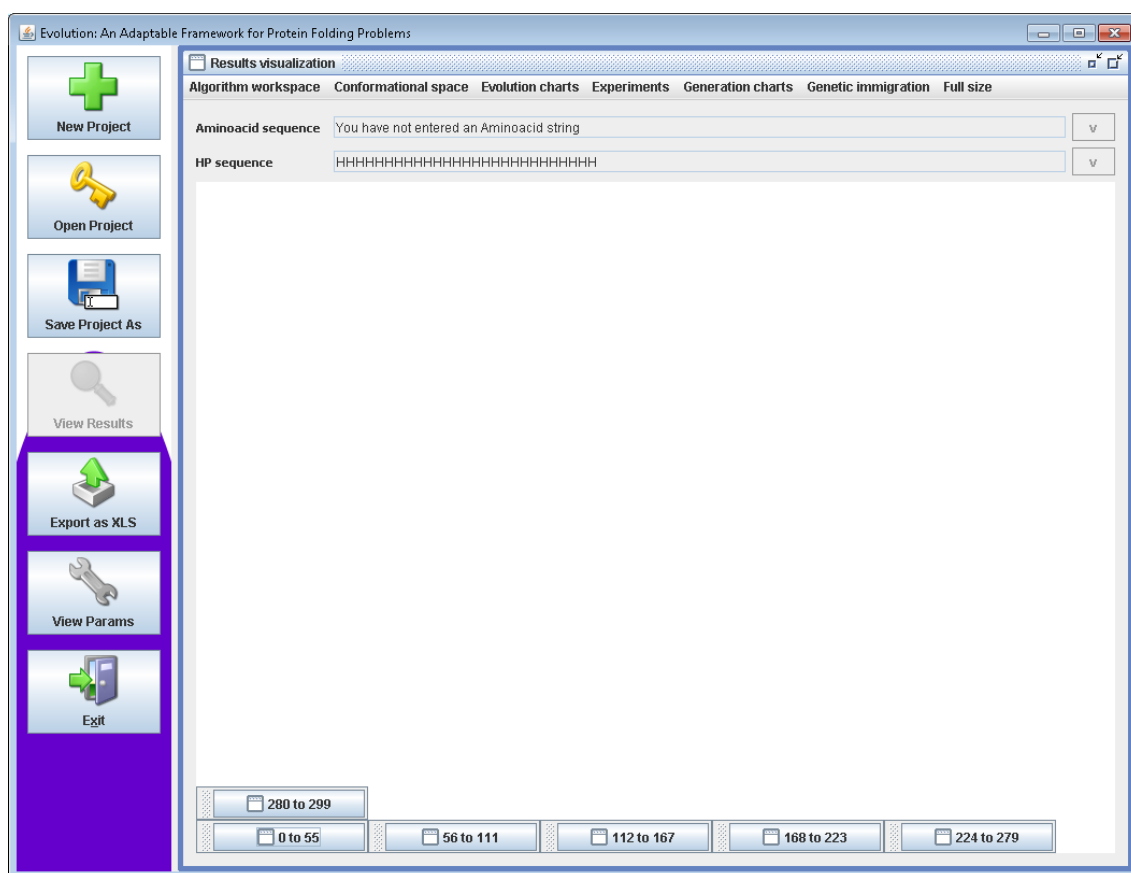


Figura 11. La interfaz *Results visualization* muestra el conjunto de cajones que agrupan en su interior las conformaciones a visualizar.

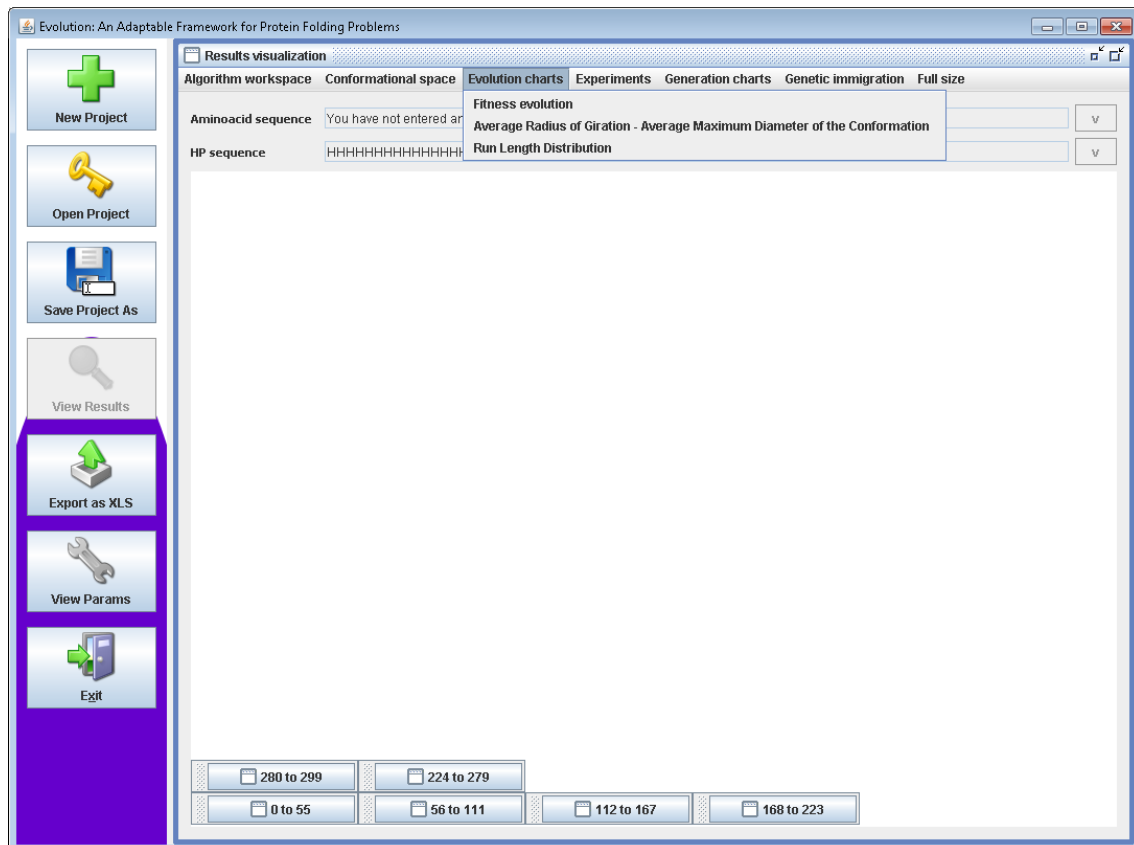


Figura 13. Visualización de los gráficos correspondientes a la historia evolutiva del algoritmo de optimización a través de la interfaz *View Results*.

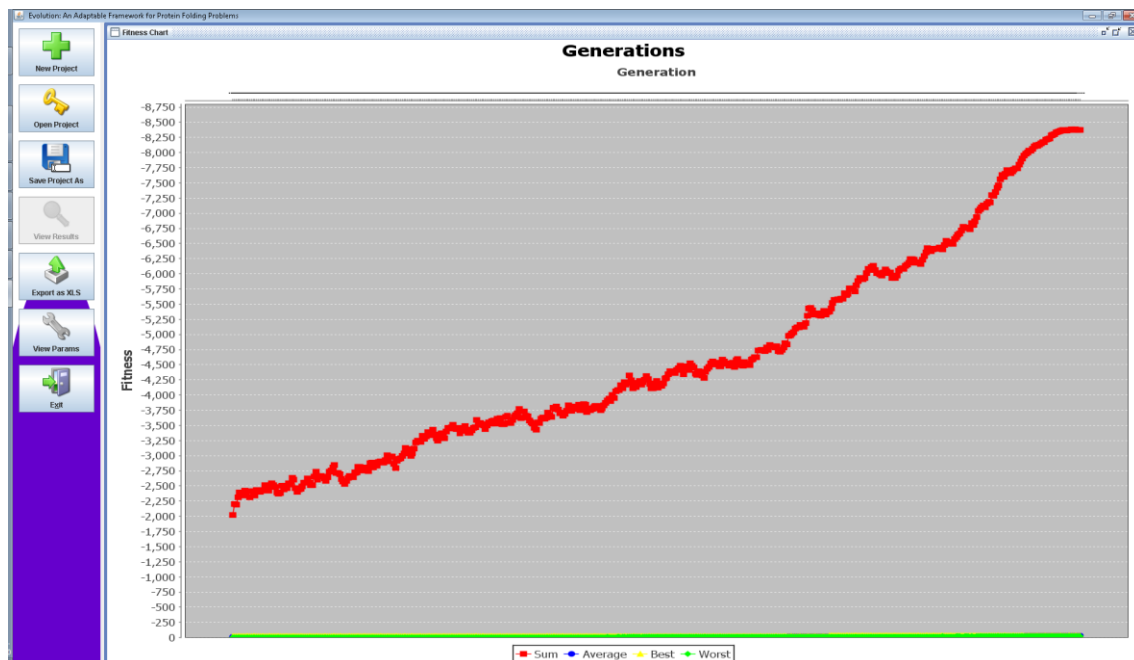


Figura 14. Gráfico *Fitness Evolution*, el cual muestra el mejor valor de fitness, el peor valor de fitness, el fitness promedio y la suma de fitness, para cada una de las generaciones de conformaciones.

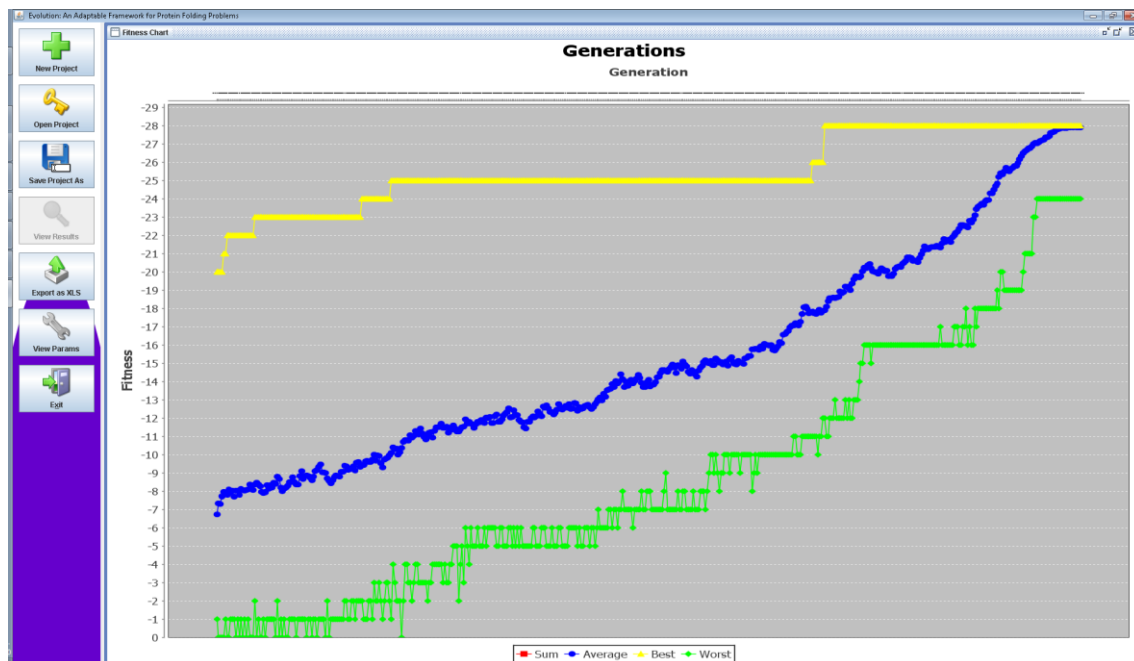


Figura 15. Gráfico *Fitness Evolution*. Amplificación de la representación del mejor valor de fitness, el peor valor de fitness y el fitness promedio para cada una de las generaciones de conformaciones.

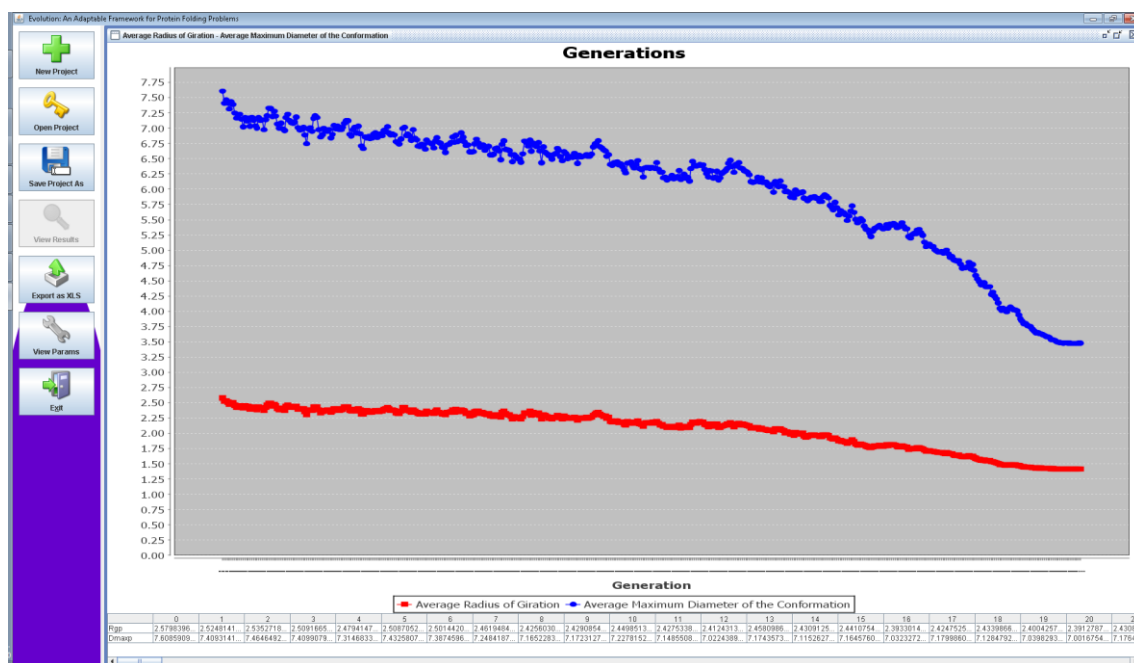


Figura 16. Gráfico *Average Radius of Giration / Average Maximum Diameter of the Conformation* mostrando la evolución del radio de giro promedio y diámetro máximo promedio a través de las generaciones.

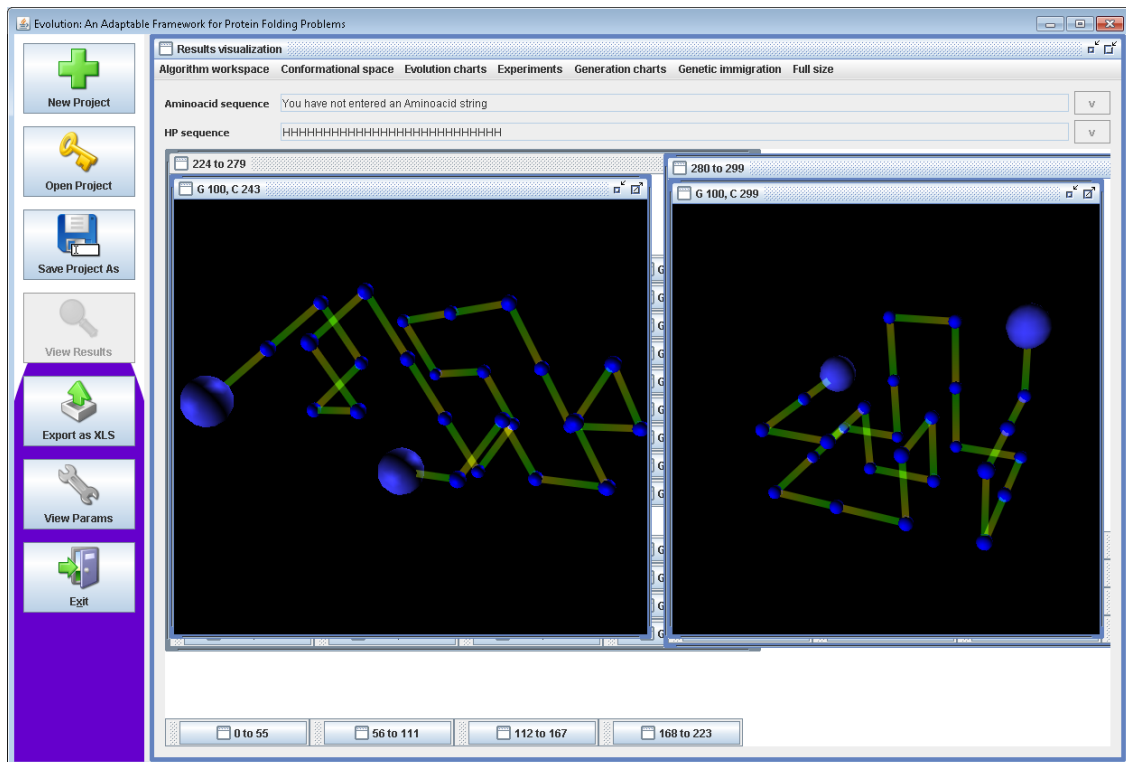


Figura 17. Visualización de dos conformaciones 3D pertenecientes a la generación 100. Notemos como ambas conformaciones 3D producidas después de 99 generaciones comienzan a "auto-organizarse" en una estructura mucho más cercana a un cubo.

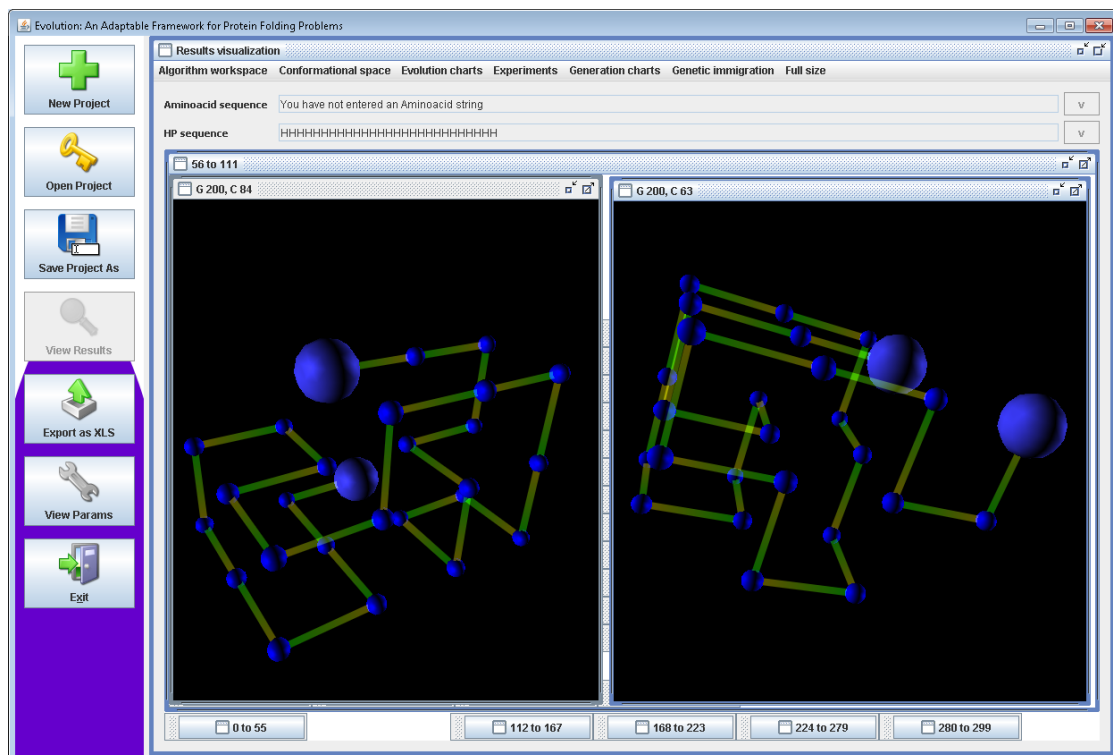


Figura 18. Visualización de dos conformaciones 3D pertenecientes a la generación 200. Las conformaciones tienden a ser más compactas y la aproximación a la forma del cubo en ambos casos comienza a ser más evidente.

Ahora, nuevamente, a través de la interfaz *View Results*, visualizaremos algunas de las conformaciones generadas, pero en esta ocasión serán aquellas correspondientes a las generaciones producidas por la ejecución del algoritmo genético. En las figuras 17, 18, 19, 20 y 21 se pueden apreciar la evolución de las conformaciones hacia un espacio de conformaciones óptimas o cercanas al óptimo, tomando en consideración las generaciones 100, 200, 300, 400 y 500, respectivamente. La conformación óptima, como ya habíamos mencionado al inicio, debe corresponder a un cubo de tres capas con 9 aminoácidos ubicados en cada capa.

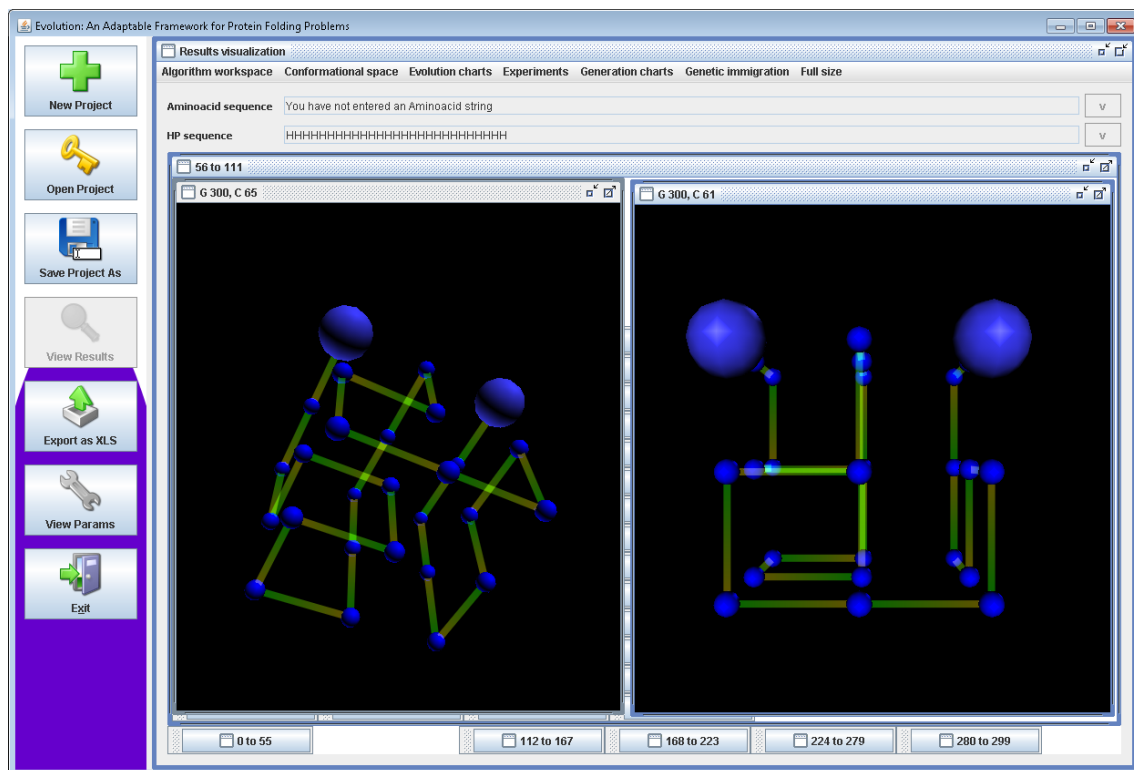


Figura 19. Visualización de dos conformaciones 3D pertenecientes a la generación 300. Las conformaciones continúan con una tendencia a ser más compactas y la aproximación a la forma del cubo en ambos casos resulta más cercana.

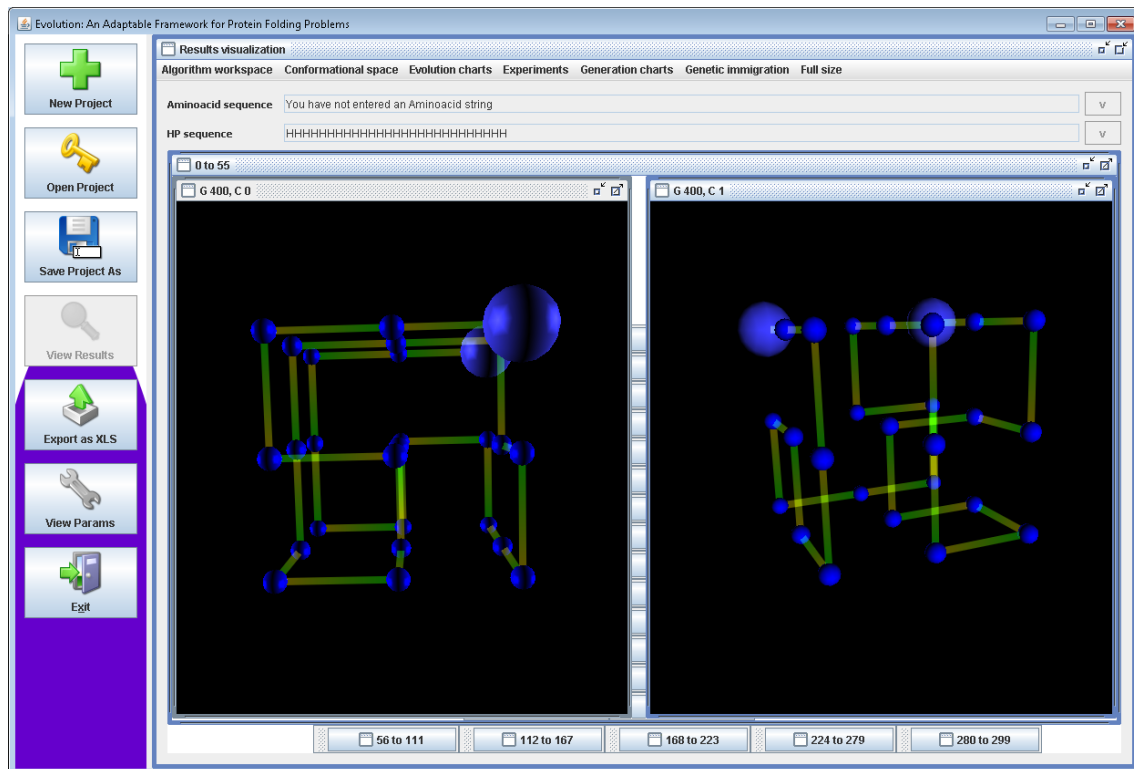


Figura 20. Visualización de dos conformaciones 3D pertenecientes a la generación 400. El proceso de optimización de las conformaciones a través del algoritmo evolutivo ha producido en esta generación conformaciones muy próximas ya a un cubo de tres capas.

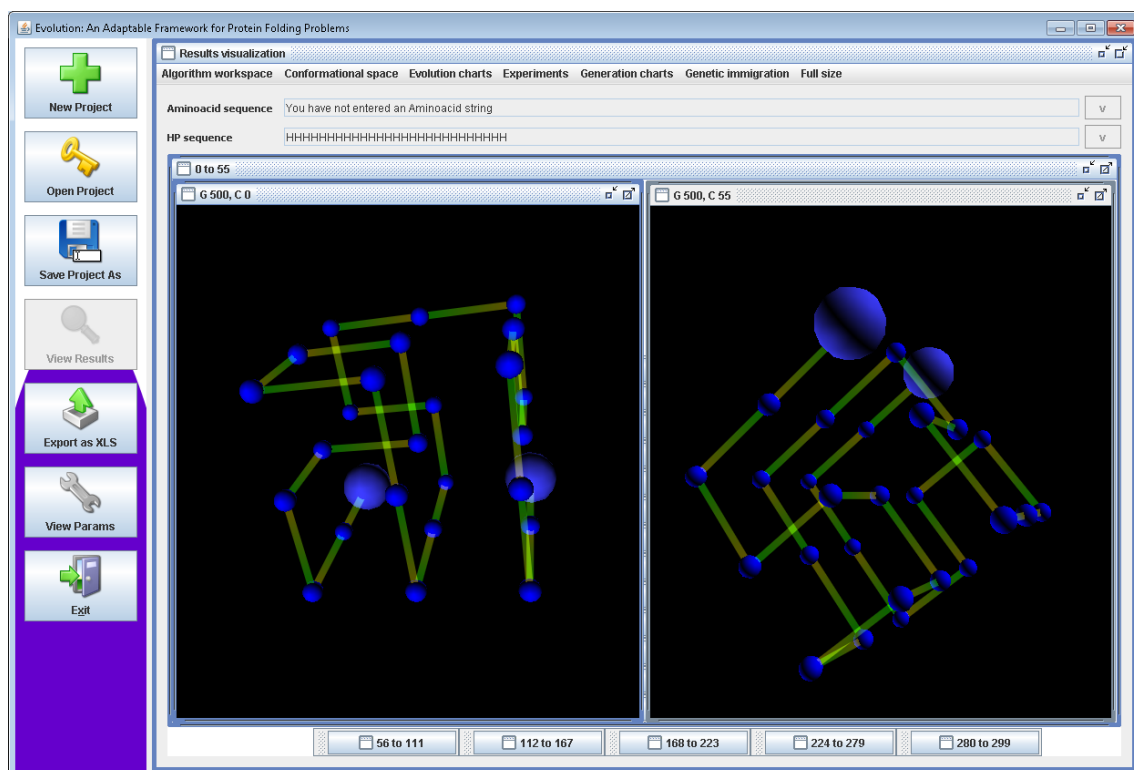


Figura 21. Visualización de dos conformaciones 3D pertenecientes a la generación 500. En esta generación ya se han producido conformaciones óptimas como las buscadas; un cubo de tres capas con 9 aminoácidos ubicados en cada capa.

A través del submenú *Generation charts* de la interfaz *View Results*, se pueden visualizar diferentes tipos de gráficos que ilustran al usuario las características del conjunto de conformaciones de una generación particular, considerando medidas tales como radio de giro, diámetro máximo de la conformación y relación entre los valores de fitness entre conformaciones padres y conformaciones hijos.

En las figuras 22 y 23 se presentan gráficamente los valores de fitness y relación entre los valores de radio de giro y diámetro máximo, respectivamente, para las conformaciones que componen la generación 100. Los valores de fitness y relación entre los valores de radio de giro y diámetro máximo para las conformaciones que componen la generación 500 se muestran en las figuras 24 y 25, respectivamente.

Nótese en la figura 24 como los valores de fitness se ubican más cercanos al valor de fitness 28, el cual resulta ser el valor de fitness que corresponde a la conformación óptima buscada; el cubo de tres capas. Por otra parte, en la figura 25 se puede apreciar una clara tendencia a los valores de radio de giro y diámetro máximo asociados al valor de fitness 28, los cuales son 1.4142 y 3.4641, respectivamente.

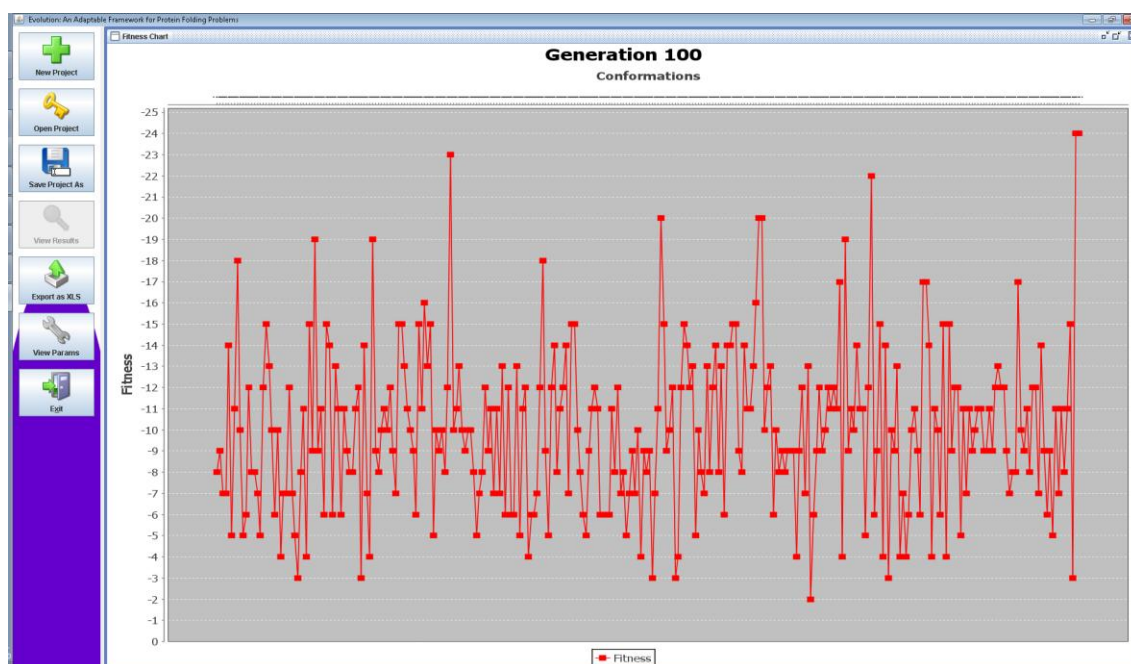


Figura 22. Valores de fitness para las conformaciones que componen la generación 100

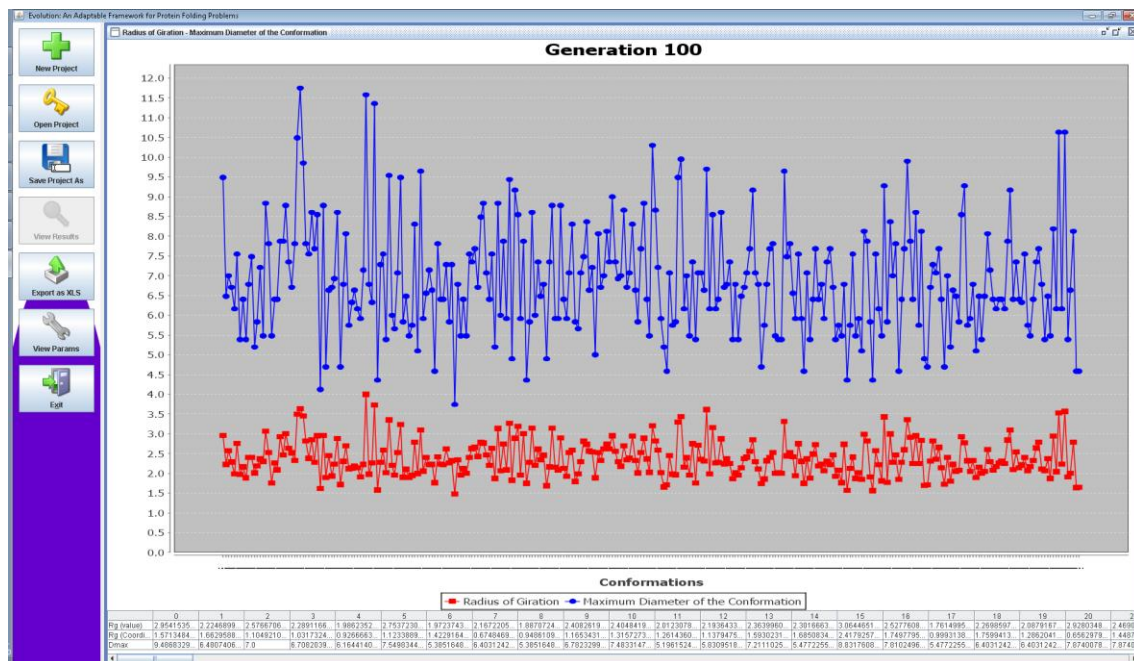


Figura 23. Relación entre los valores de radio de giro y diámetro máximo para las conformaciones que componen la generación 100

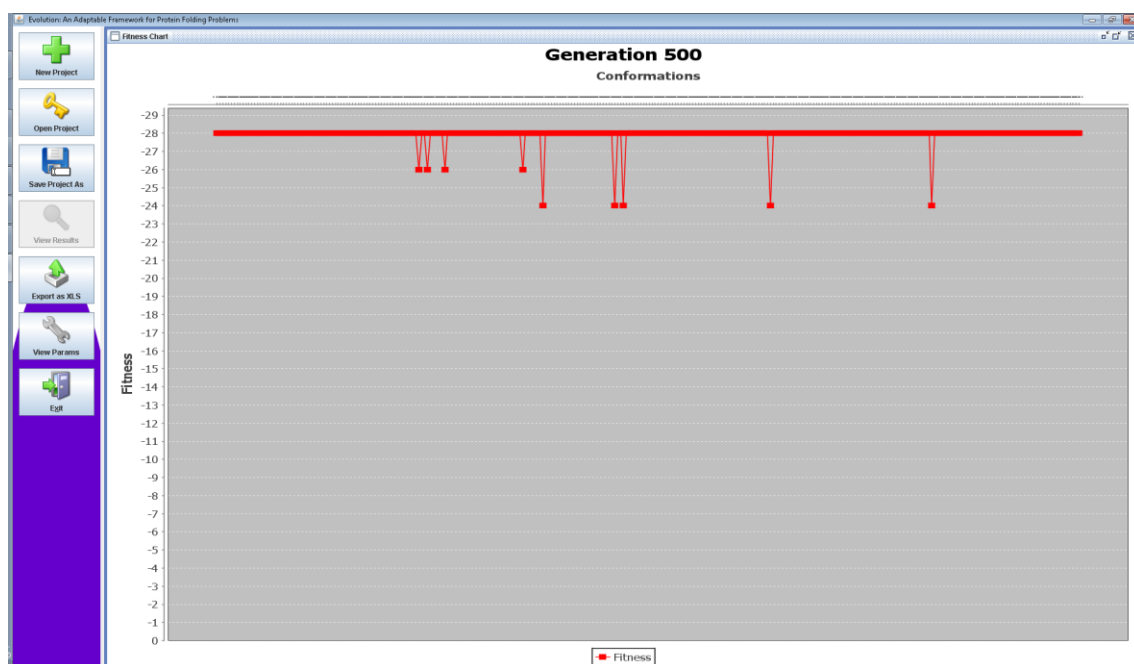


Figura 24. Valores de fitness para las conformaciones que componen la generación 500.

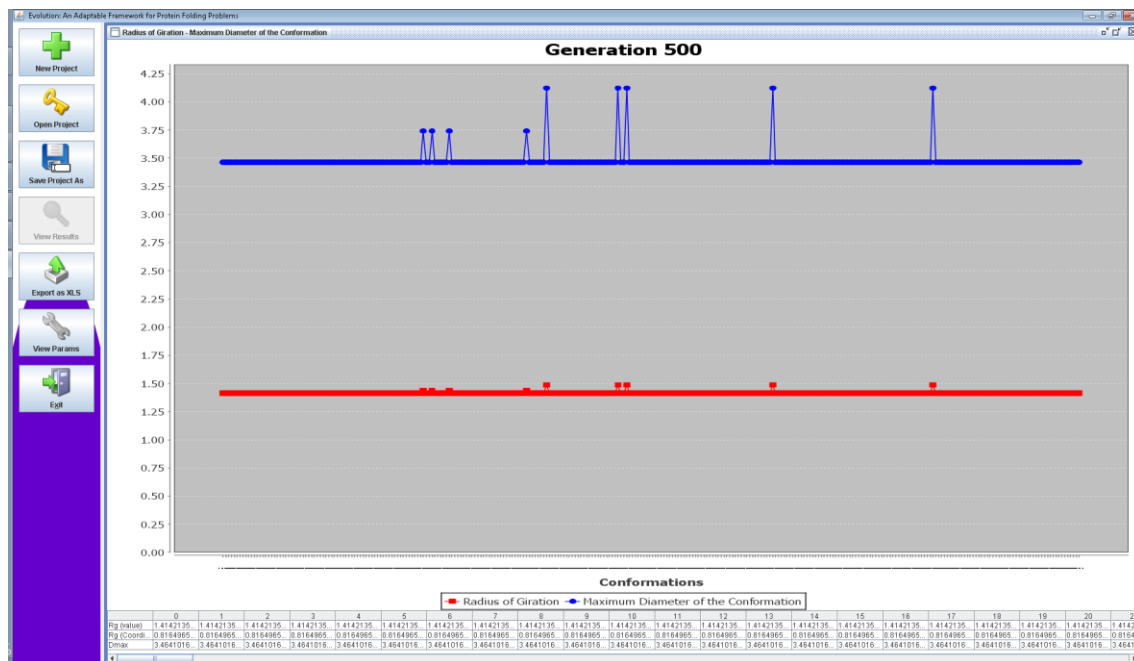


Figura 25. Relación entre los valores de radio de giro y diámetro máximo para las conformaciones que componen la generación 500.

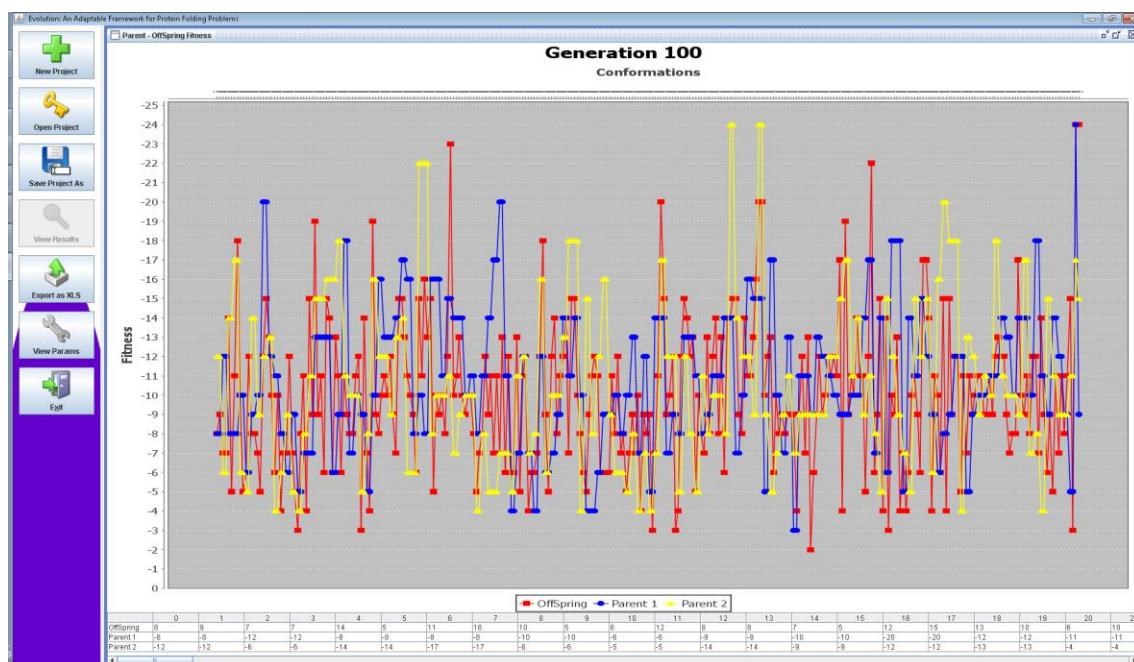


Figura 26. Relación entre los valores de fitness de las conformaciones padres con el valor de fitness de la conformación hijo en la generación 100.

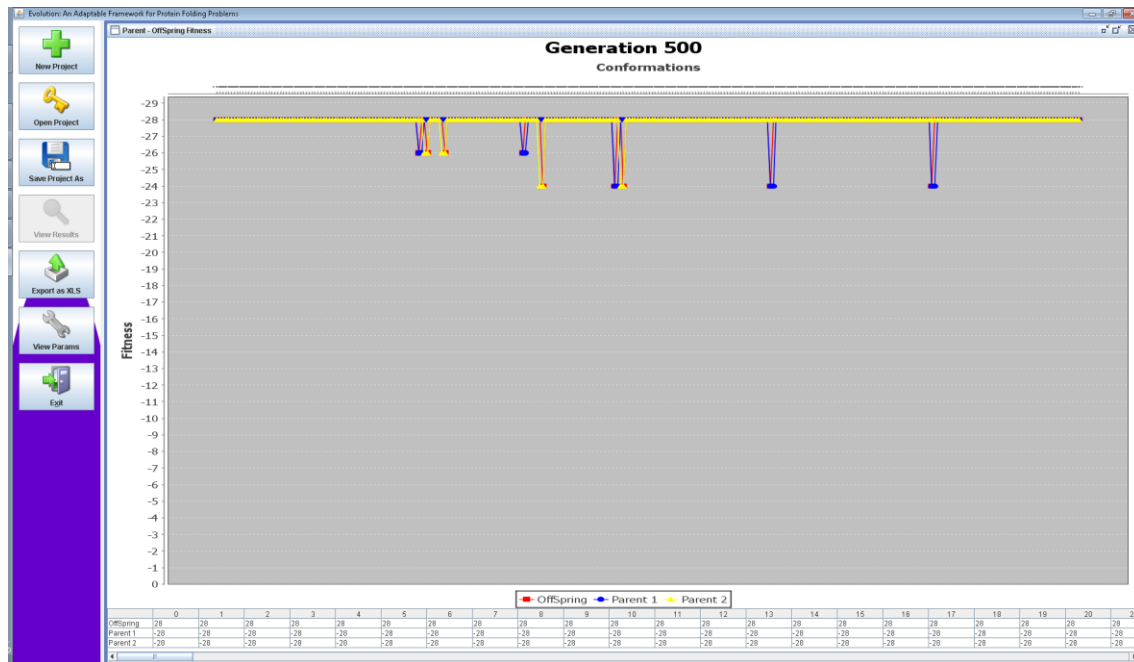


Figura 27. Relación entre los valores de fitness de las conformaciones padres con el valor de fitness de la conformación hijo en la generación 500.

Así mismo, en las figuras 26 y 27 se puede apreciar la relación entre los valores de fitness de las conformaciones padres y la conformación hijo, para cada una de las conformaciones de una generación particular. En la figura 26, el gráfico corresponde a las conformaciones de la generación 100, mientras que la figura 27 hace referencia a la generación 500. Nótese en la figura 27 la gran cercanía entre los valores de fitness de los padres y el valor de fitness del hijo para la gran mayoría de las conformaciones de esta generación. También cabe destacar que en muchos casos (nos referimos a la figura 27) el valor de fitness del hijo coincide con el mayor de los dos valores de fitness de los padres. Este hecho anticipa el arribo a un estado de convergencia en el proceso de optimización. Es decir, un estado en el cual no habrá diferencias entre los valores de fitness de las conformaciones de una generación.

De la misma manera la figura 28 muestra el menú *Experiments* donde se encuentra la lista de todos los experimentos ejecutados, que en caso de volverse muy extensa, se podrá acceder a ellos gracias a un *Scroll* que aparecerá automáticamente (véase figura 29). También muestra un submenú *Project Charts* que muestra los menús que permiten visualizar la evolución del fitness y el radio de giro promedio/distancia máxima promedio entre extremos a través de los experimentos (figuras 31 y 32).

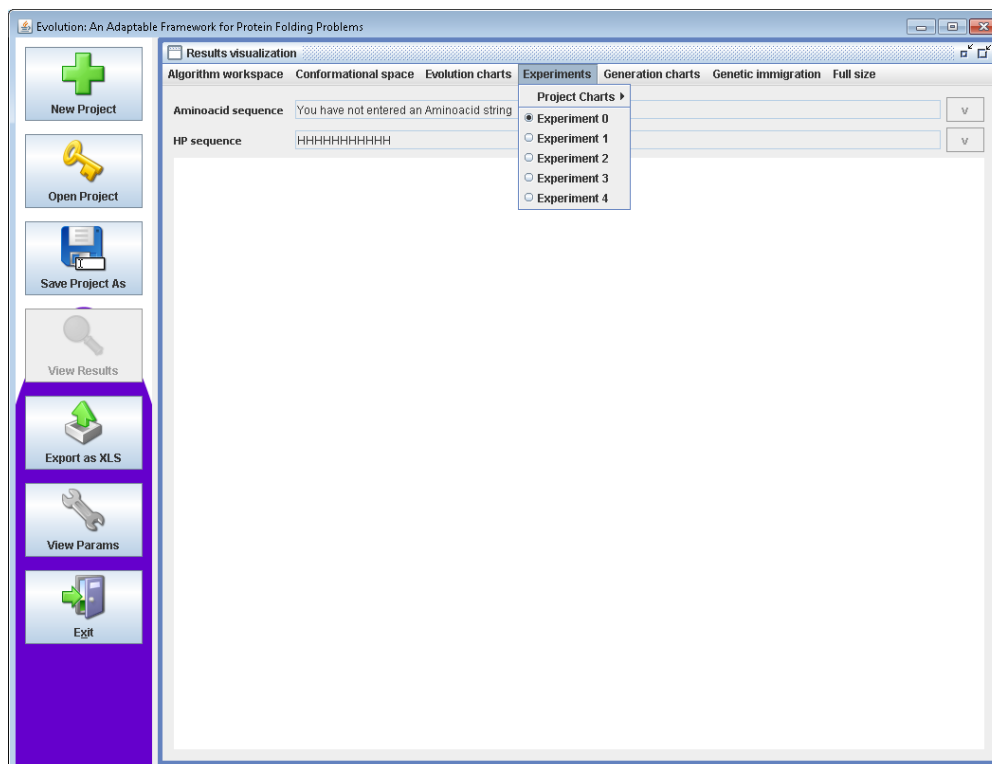


Figura 28. Lista de los diferentes experimentos pertenecientes al proyecto. Así como el submenú *Project Charts*.

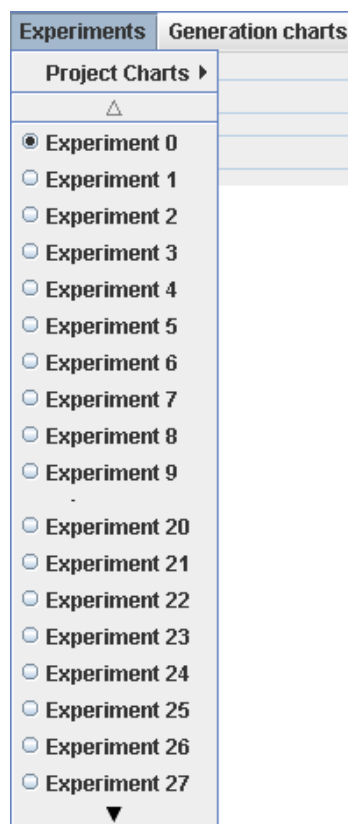


Figura 29. Lista de los diferentes experimentos pertenecientes al proyecto. Así como el submenú *Project Charts*.

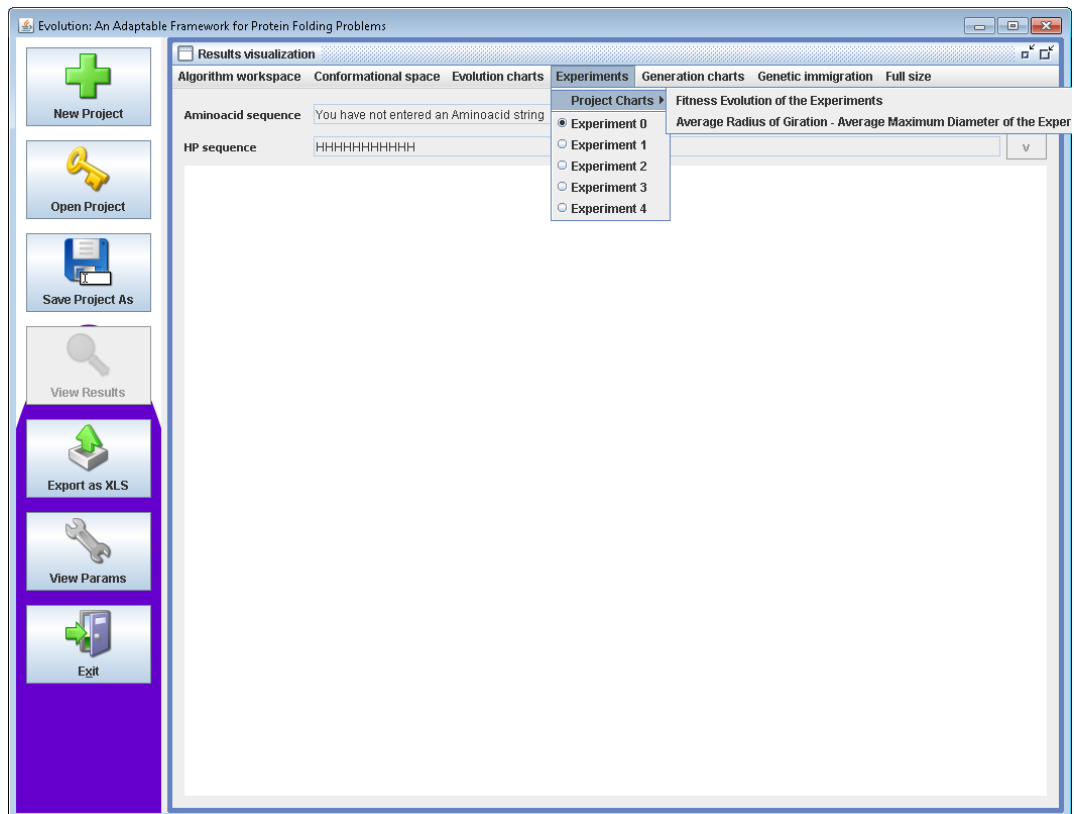


Figura 30. Submenú *Project Charts* que muestra los menús que permiten visualizar la evolución del fitness y el radio de giro promedio/distancia máxima promedio entre extremos a través de los experimentos.

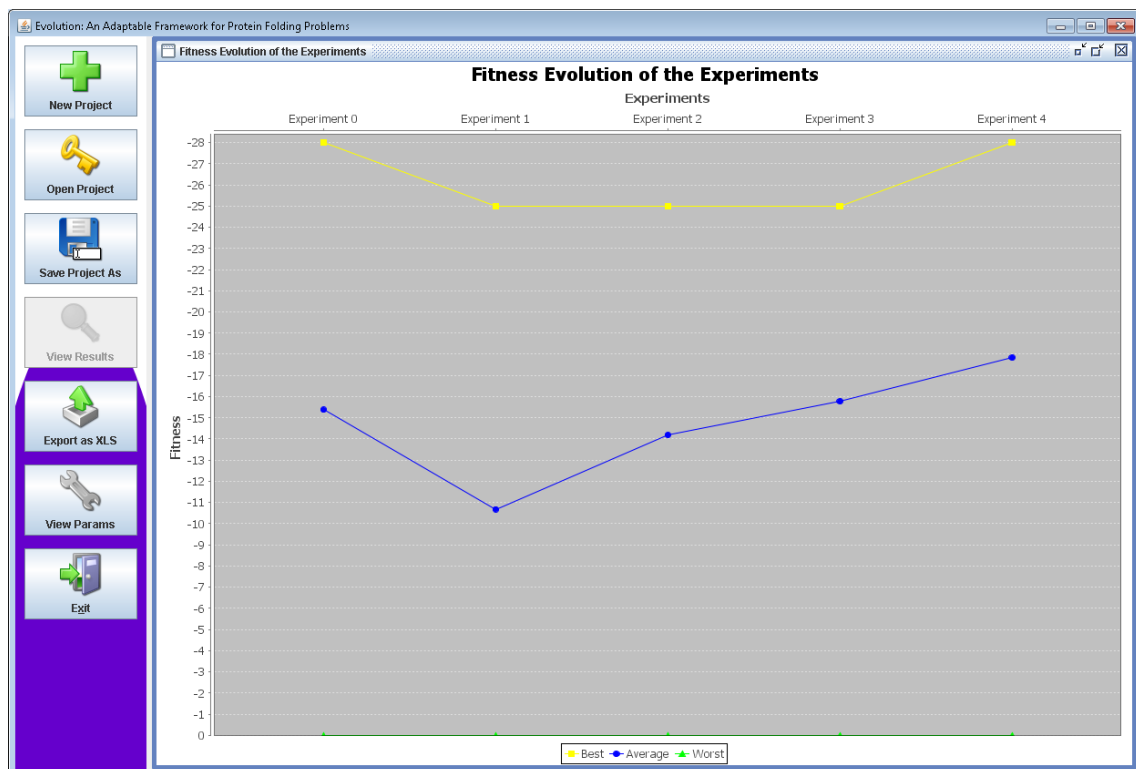


Figura 31. Gráfico *Fitness Evolution of the Experiments*. Representación del mejor valor de fitness, el peor valor de fitness y el fitness promedio para cada uno de los experimentos en el proyecto.

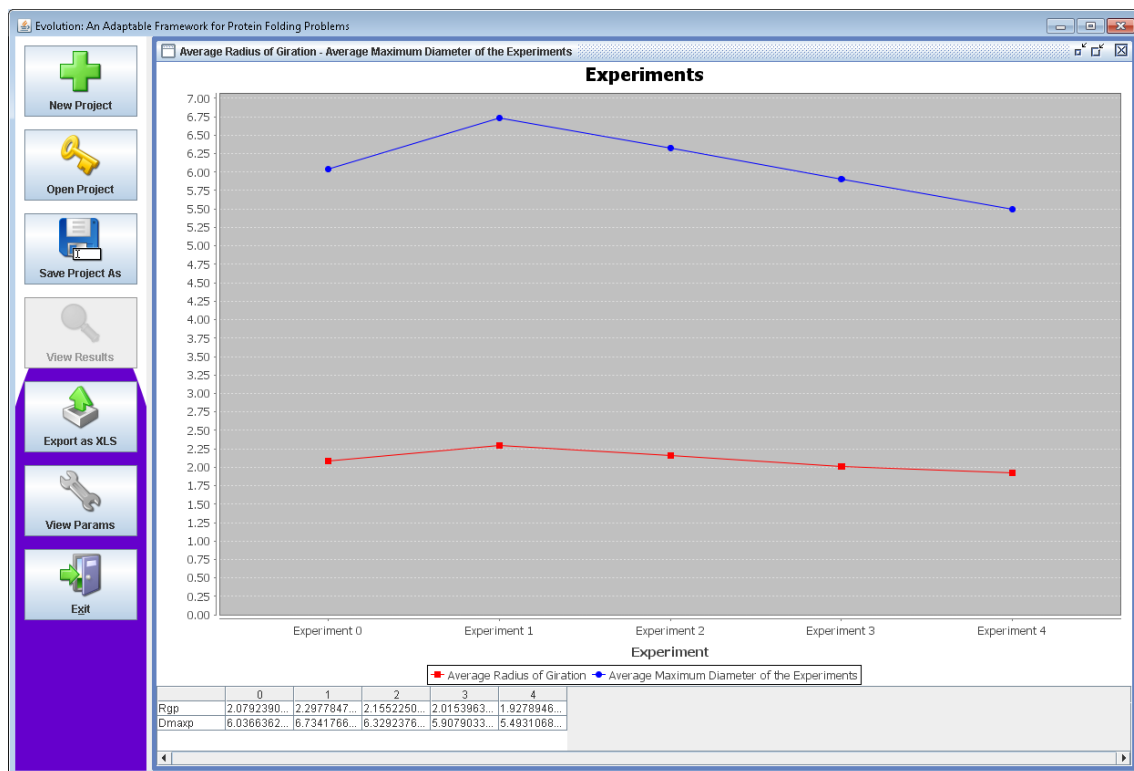


Figura 32. Gráfico *Average Radius of Giration / Average Maximun Diameter of the Conformation* mostrando la evolución del radio de giro promedio y diámetro máximo promedio a través de los Experimentos.

Algoritmo Genético RankGA

Secuencia: **H27**

Características del experimento

- **Algoritmo de optimización:** RankGA
- **Tipo de representación:** 3D Cubic
- **Tamaño de la población de conformaciones:** 28
- **Número de generaciones para el algoritmo evolutivo:** 500
- **Número de experimentos en el proyecto:** 100
- **Operadores genéticos:** selección según el método *Roulette selection*, cruce en dos puntos y mutación aleatoria.
- **Otras técnicas genéticas a considerar:** elitismo activado y Clamp mutation
- **Probabilidad de cruce:** 1.0
- **Probabilidad de mutación mínima:** 0.0
- **Probabilidad de mutación máxima:** 0.4
- **Vecindad:** 10%
- **Fitness final:** 28

Una vez que se selecciona del menú principal (ver figura 1) la opción *New Project*, entonces viene presentada al usuario una nueva interfaz, tal como la que se muestra en la figura 2, a través de la cual se introduce la secuencia o segmento de secuencia que se desea explorar y analizar, en este caso nuevamente se trata de la secuencia H27.

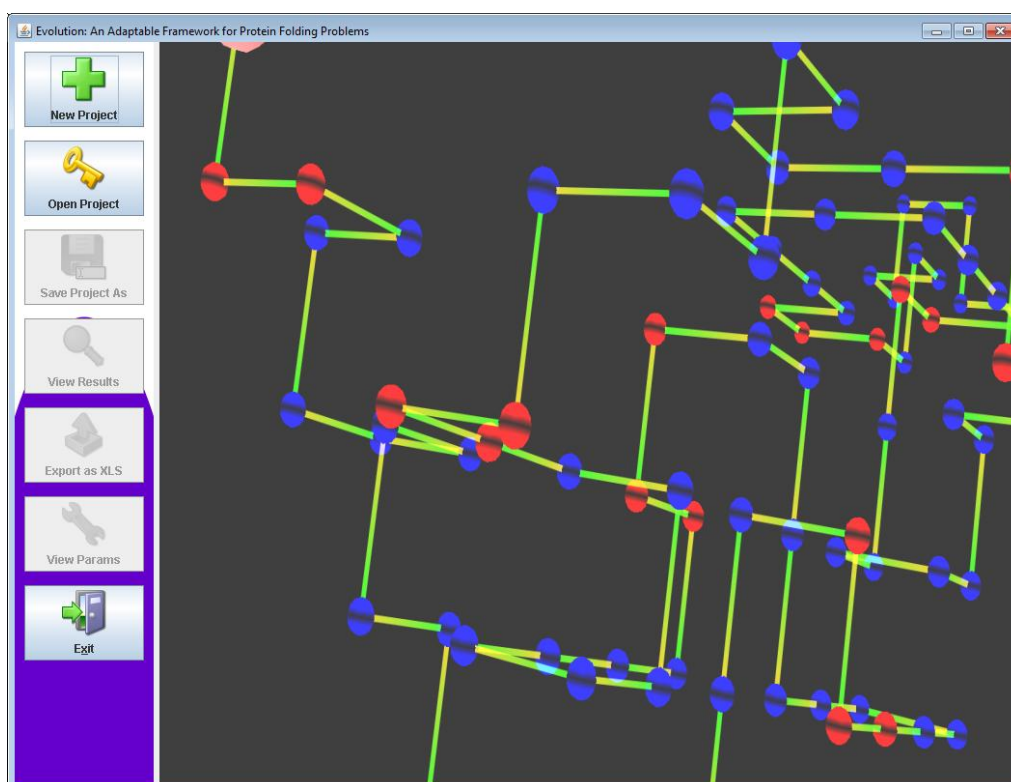


Figura 1. Interfaz y menú principal de la plataforma bioinformática Evolution.

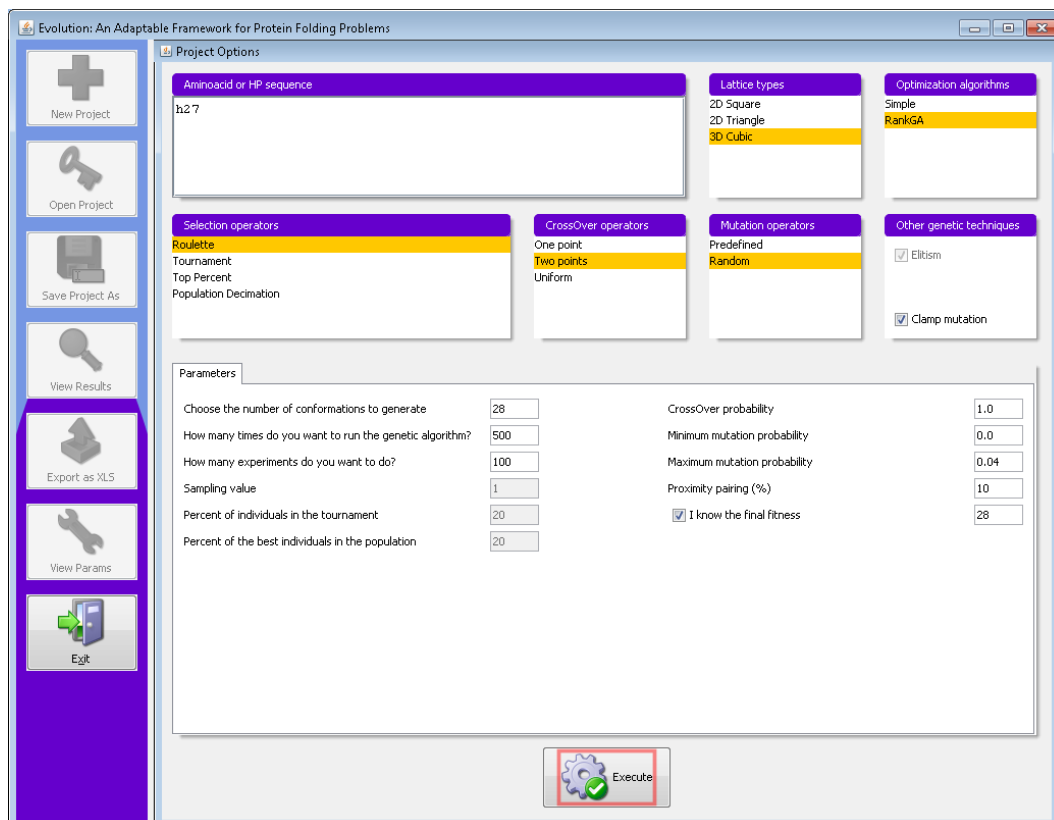


Figura 2. Interfaz para introducir la secuencia a explorar. Tal secuencia puede ser descrita en términos de aminoácidos o como secuencia HP, la herramienta detectará automáticamente de cuál se trata. Además, en esta interfaz, se pueden modificar los parámetros necesarios para la ejecución del Algoritmo Genético seleccionado.

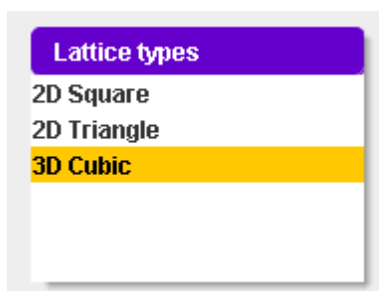


Figura 3. A través de la interfaz *New Project* se selecciona el modelo *Lattice* a ejecutar para generar la población de conformaciones aleatorias a partir de la secuencia HP proporcionada como entrada. En el experimento en ejecución se ha seleccionado el modelo *Lattice 3D Cubic*.

Después de haber seleccionado el modelo *Lattice* a ejecutar (ver figura 3), se debe especificar el algoritmo de optimización que se va a emplear, en este caso el algoritmo RankGA, tal como se muestra en la figura 4.

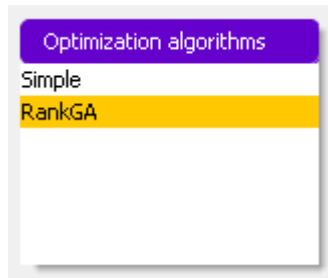


Figura 4. A través de la interfaz *New Project* se selecciona el algoritmo de optimización a ejecutar para generar la población de conformaciones. En el experimento en ejecución se ha seleccionado el *algoritmo RankGA*.

Una vez seleccionado el GA a ejecutar, se deben especificar los parámetros comunes con los que se desea trabajar como el *Operador de Selección*, el *de cruce* y el *de mutación* como se ve en la figura 5.

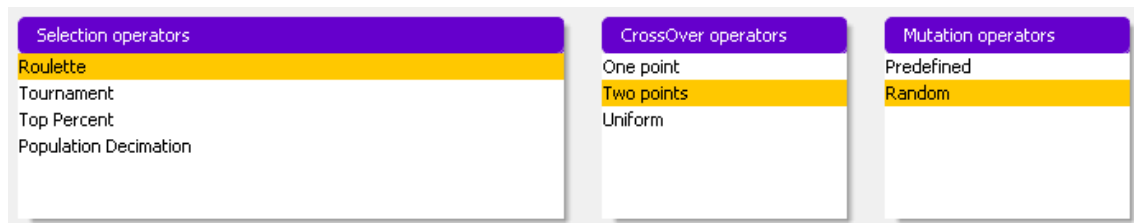


Figura 5. A través de la interfaz *New Project* se seleccionan los operadores de selección, cruce y mutación. En el experimento en ejecución se ha seleccionado el operador de selección *Roulette*, el operador de cruce *Two points* y el operador de mutación *Random*.

Además, se pueden seleccionar otras técnicas genéticas como se ejemplifica en la figura 6.

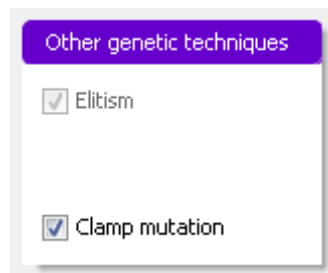


Figura 6. A través de la interfaz *New Project* se pueden seleccionar otras técnicas genéticas la *Clamp mutation*.

Una vez que se han seleccionado los tipos de operadores y parámetros que caracterizarán el algoritmo genético, se debe proporcionar el número de conformaciones con las que contará la generación inicial, el número de generaciones que se ejecutaran, el número de veces que será creada la población inicial, el “Sampling” de conformaciones y la probabilidad de cruce, como se muestra en la figura 7.

Parameters	
Choose the number of conformations to generate	28
How many times do you want to run the genetic algorithm?	500
How many experiments do you want to do?	100
Sampling value	1
Percent of individuals in the tournament	20
Percent of the best individuals in the population	20
CrossOver probability	1.0
Minimum mutation probability	0.0
Maximum mutation probability	0.04
Proximity pairing (%)	10
☒ I know the final fitness	28

Figura 7. A través de la interfaz *New Project* se permiten ajustar los parámetros del algoritmo como; número de conformaciones, de generaciones, de experimentos, la probabilidad de cruce, las probabilidades mínimas y máximas, el porcentaje de vecindad y el fitness final. En el experimento en ejecución se han asignado los valores 28, 500, 100, 1.0, 0.0, 0.04, 10 y 28 respectivamente.

Una vez introducida la secuencia de aminoácidos o secuencia HP, habiendo seleccionado el tipo de lattice y parametrizado el GA seleccionado, se pulsa el botón *Execute* para iniciar el proceso de generación de la población inicial de conformaciones y la optimización de las estructuras moleculares que la componen.

Como se ilustra en la figura 8, la opción *View Results* en el menú principal se activa tan pronto el proceso ha concluido. La selección de la opción *View Results* tras haber ejecutado el algoritmo genético producirá como resultado las visualizaciones y gráficos disponibles para su manipulación.

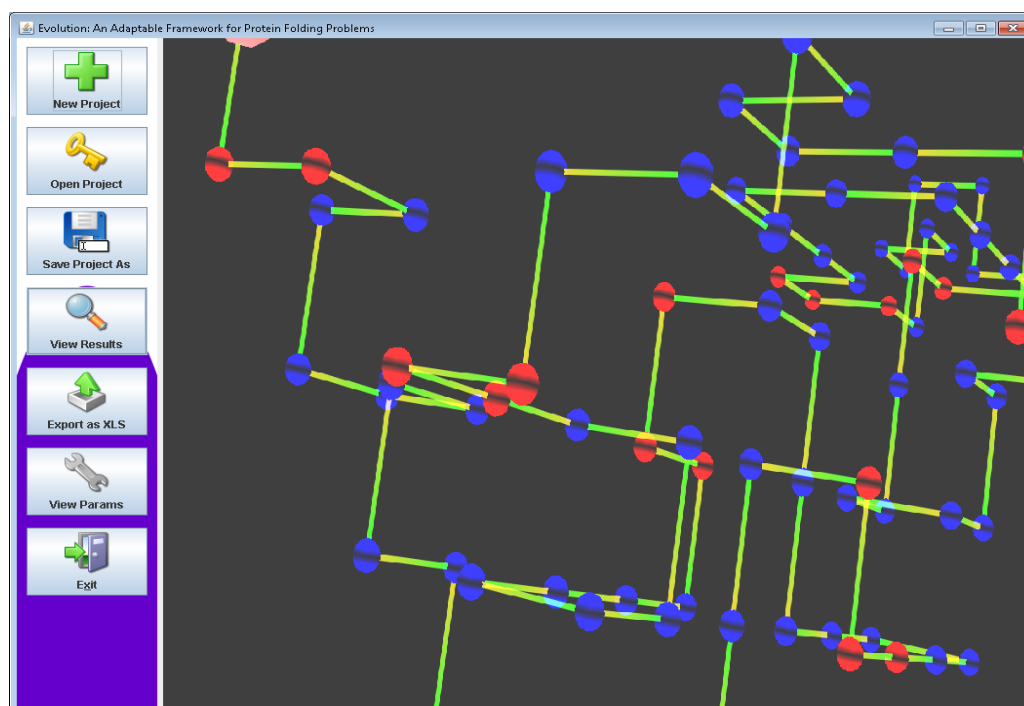


Figura 8. Activación de la opción *View Results* en el menú de la interfaz principal.

Como se muestra en las figuras 9, 10, 11 y 12, Evolution proporciona al usuario una amplia gama de servicios para la visualización e interacción con el espacio conformacional.

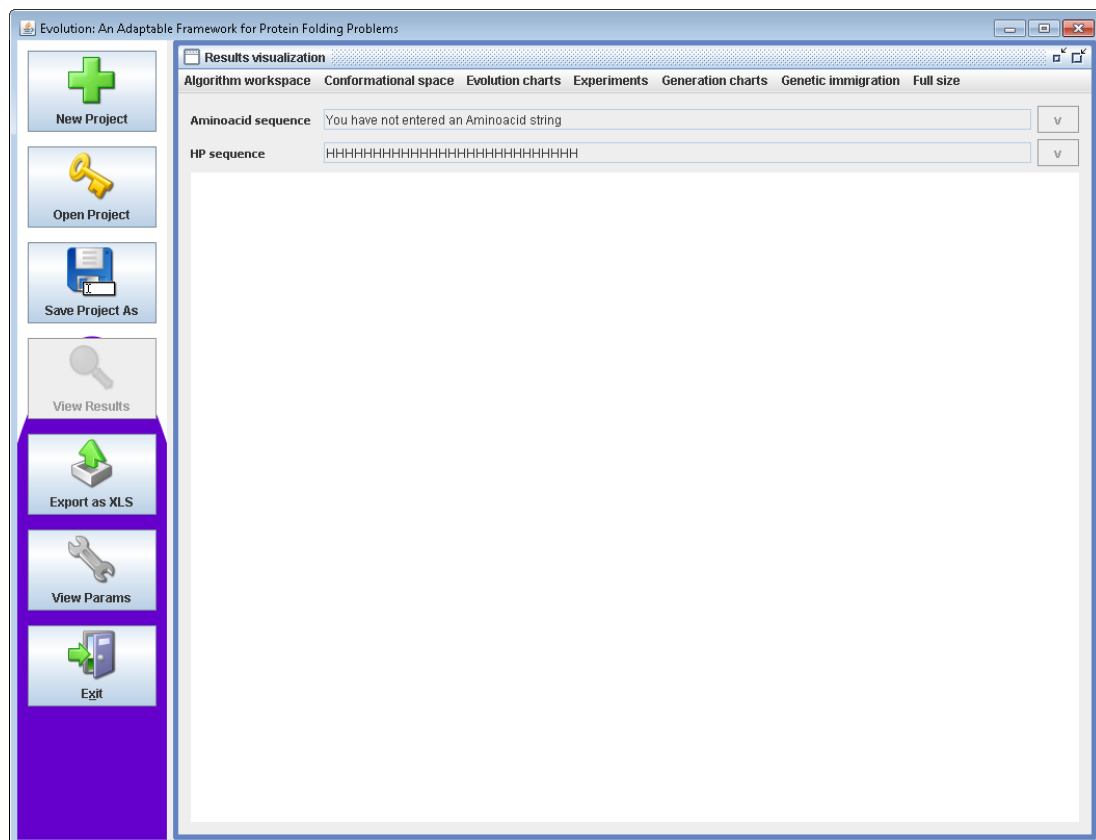


Figura 9. Interfaz *View Results*, a través de la cual resulta posible la visualización 2D/3D de las conformaciones que integran el espacio conformacional.

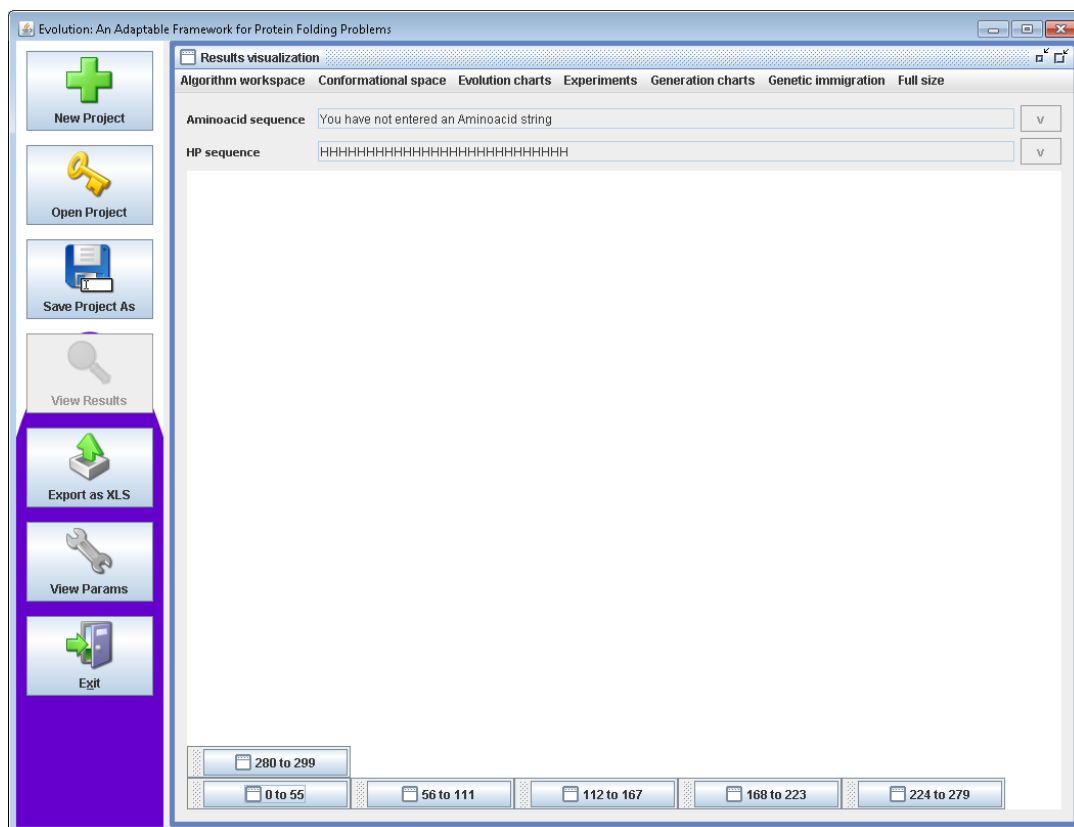


Figura 10. La interfaz *Results visualization* muestra el conjunto de cajones que agrupan en su interior las conformaciones a visualizar.

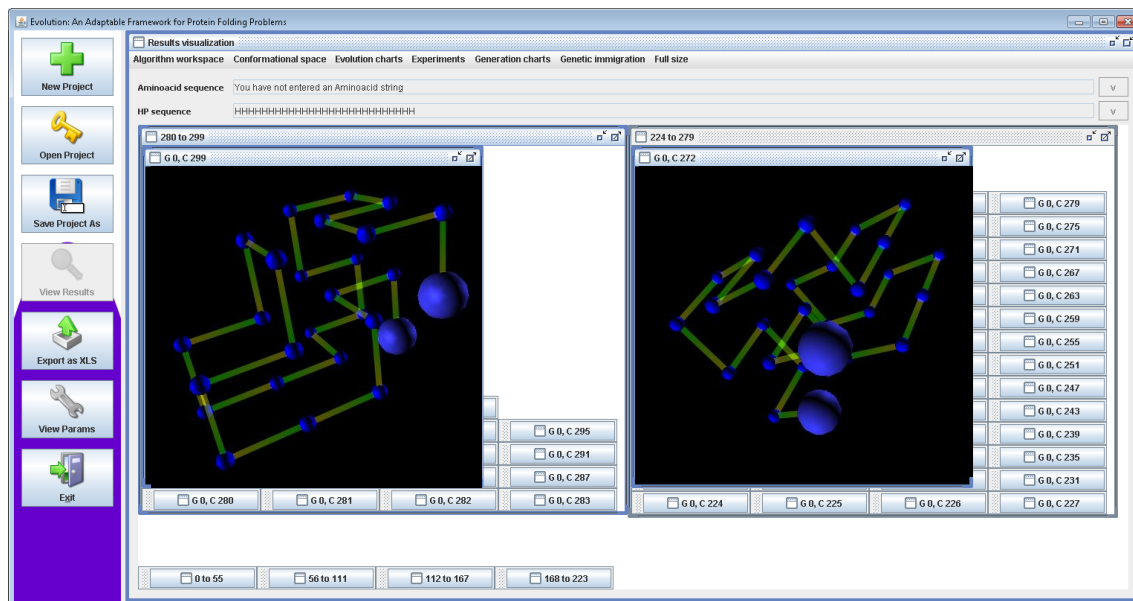


Figura 11. Visualización de dos de las 500 conformaciones que integran el espacio conformacional final.

Como se puede apreciar en la figura 12, las conformaciones que integran el espacio conformacional inicial están cerca en su estructura de aquellas a las cuales nos deberá aproximar el algoritmo de optimización, es decir, el cubo de tres capas.

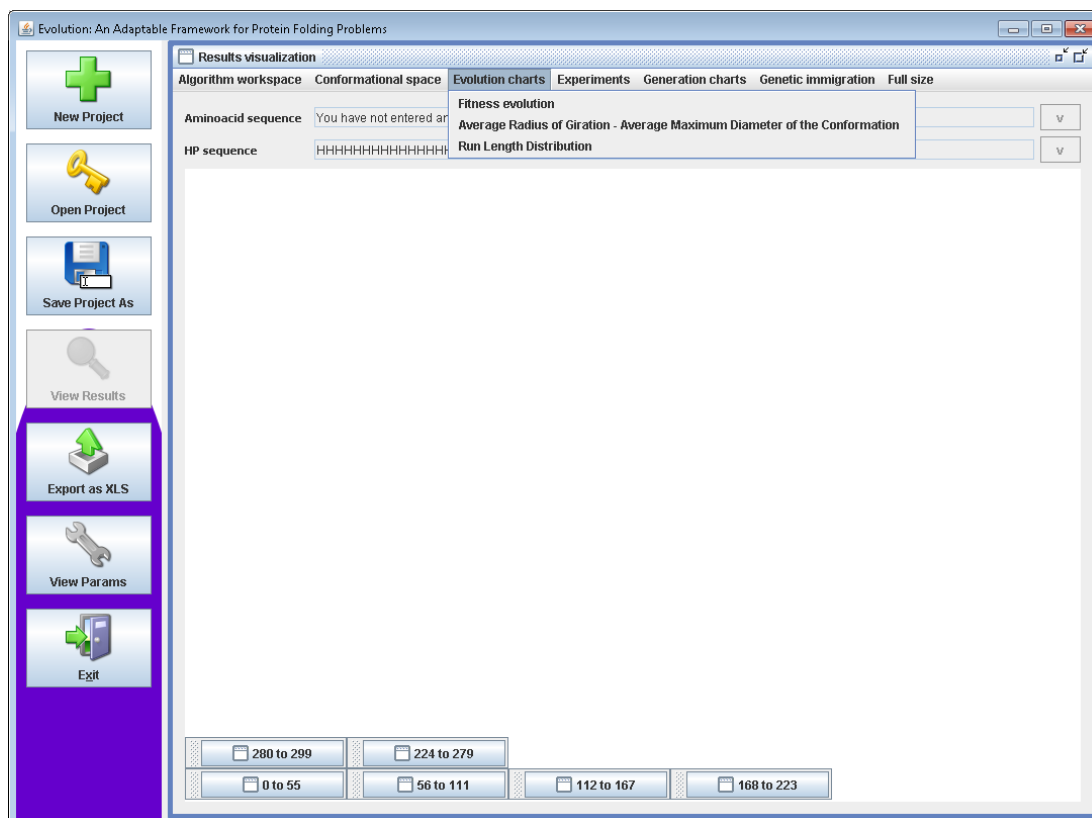


Figura 12. Visualización de los gráficos correspondientes a la historia evolutiva del algoritmo de optimización a través de la interfaz *View Results*.

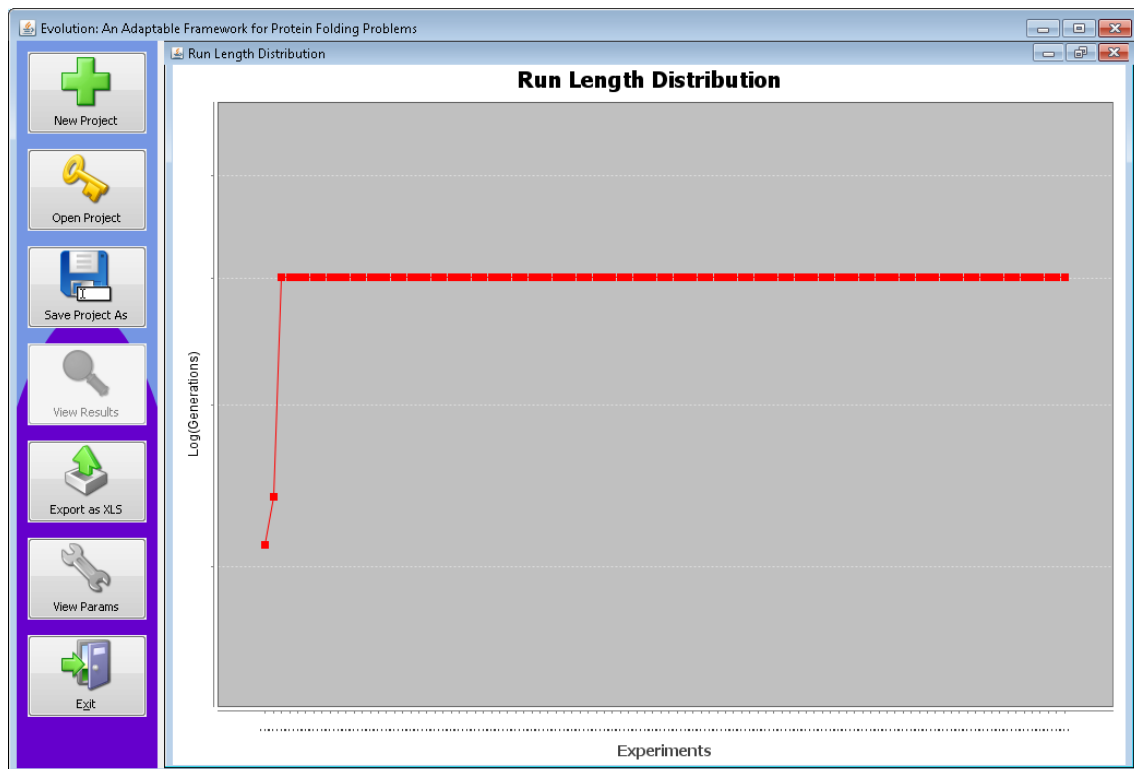


Figura 13. Gráfico *Run Length Distribution*, el cual muestra las generaciones requeridas por cada experimento para alcanzar el fitness que hemos definido.

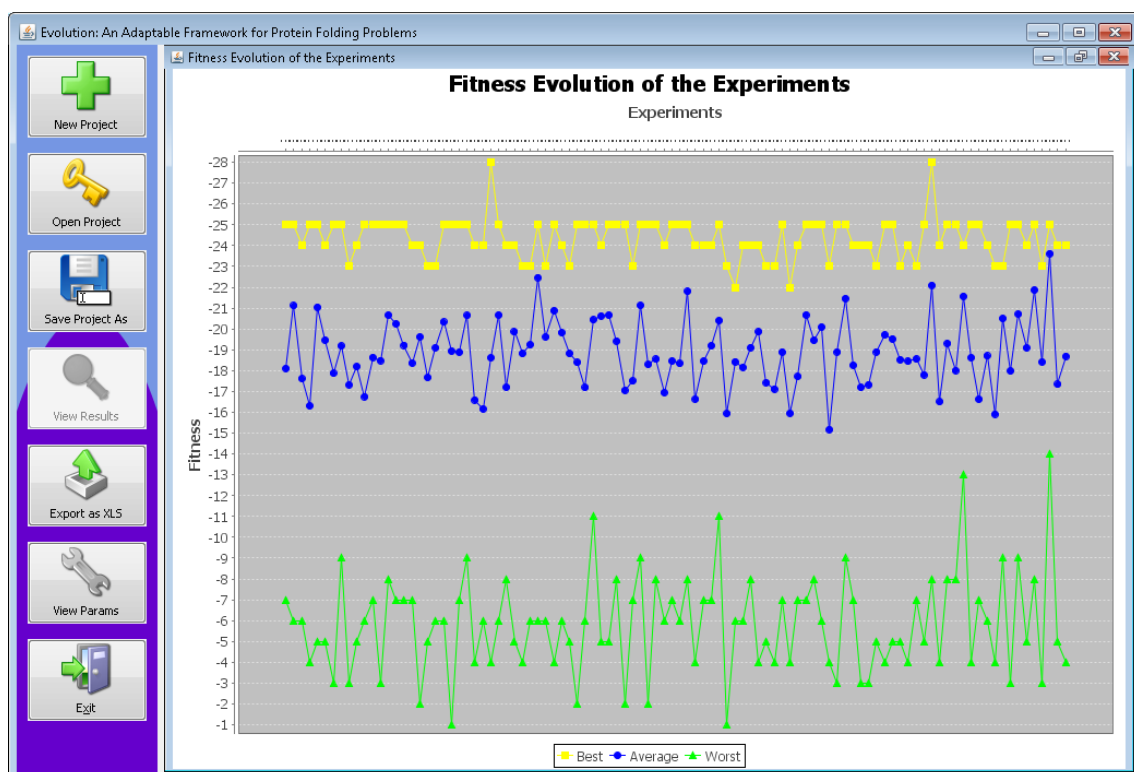


Figura 14. Visualización de la evolución del mejor, peor y del fitness promedio a través de los experimentos en el proyecto.

Además de los gráficos mostrados en el primer experimento (Algoritmo Genético Simple), en la figura 13 se muestra la gráfica *Run Length Distribution* que muestra las generaciones requeridas por cada experimento para alcanzar el fitness que hemos definido (ver figura 7). Se puede acceder a ella mediante el menú *Evolution charts* en la figura 12.

En la figura 13 se muestran dos experimentos que han alcanzado el fitness deseado antes de recorrer el número máximo de generaciones previamente especificado, en este caso 500.

Una vez conocido que existen al menos dos experimentos que han alcanzado el fitness esperado procedemos a visualizar, como se muestra en la figura 14, el gráfico *Fitness Evolution of the Experiments* que muestra los experimentos que han obtenido el fitness 28 y que se muestran en la figura 15.

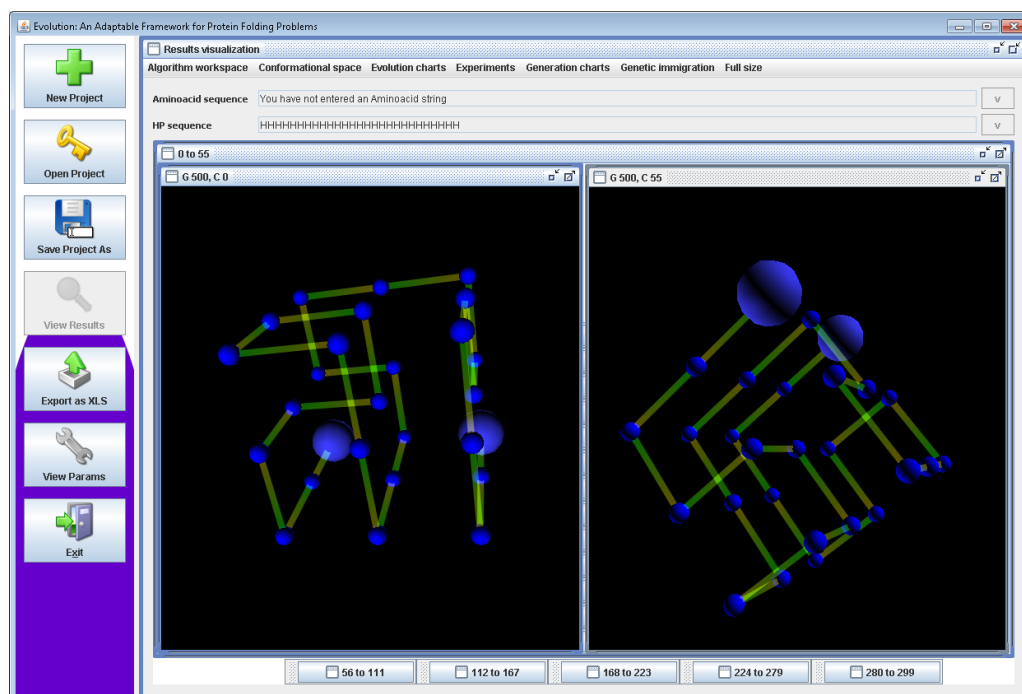


Figura 15. Visualización de dos conformaciones 3D pertenecientes al experimento 26. En éste experimento ya se han producido conformaciones óptimas como las buscadas; un cubo de tres capas con 9 aminoácidos ubicados en cada capa.