

Estudio de Estrategias de Archivo en PSO Multi-Objetivo para el Docking Molecular

Esteban López-Camacho¹, María Jesús García-Godoy¹, José García-Nieto¹,
Antonio J. Nebro¹, and José F. Aldana-Montes¹

Khaos Research Group
Department of Computer Sciences, University of Málaga, ETSI Informática,
Campus de Teatinos, Málaga, Spain
esteban,mjgarcia,g,jnieto,antonio,jfam@lcc.uma.es

Resumen El acoplamiento molecular es un problema de optimización complejo cuyo objetivo es la predicción de la posición de un ligando en el sitio activo de un receptor con la mínima energía de unión. Este problema puede ser formulado como un problema de optimización de dos objetivos que minimiza la energía de unión y la desviación de la media cuadrática de las posiciones atómicas (RMSD) de los ligandos. En este contexto, el algoritmo multi-objetivo de *swarm-intelligence* SMPSO mostró un rendimiento destacable. SMPSO se caracteriza por usar un archivo externo para almacenar las soluciones no dominadas y como base para estrategia de selección de líder. En este artículo, se analizan diferentes variantes de SMPSO basadas en diferentes estrategias de archivo utilizando un *benchmark* de instancias moleculares. Este estudio revela que la variante SMPSO_{hv} obtiene los mejores resultados.

Keywords: Optimización multi-objetivo, Optimización por enjambre de partícula, Acoplamiento molecular, Estrategias de archivo, Comparación algorítmica

1. Introducción

El acoplamiento molecular (o molecular docking) es un problema de optimización complejo que consiste en predecir la posición de una molécula pequeña (ligando) en el sitio activo del receptor (macromolécula) cuya energía de unión sea mínima. Para resolver el problema de acoplamiento molecular se vienen utilizando tradicionalmente algoritmos metaheurísticos [6] [2], ya que se trata de un problema complejo de optimización, que requiere del ajuste de la posición de las variables correspondientes a las coordenadas de traslación y el movimiento de torsión de la molécula.

En la última década han ido apareciendo una serie de estudios centrados en la aplicación de metaheurísticas mono-objetivo y multi-objetivo al acoplamiento molecular [6] [2] [4] [13] [3], mostrando resultados competitivos para un conjunto de instancias [6] [2] [4] [13] [3]. En estos trabajos, se han propuesto diferentes formulaciones centradas en la optimización de términos de las funciones de energía.

Esteban López-Camacho *et al.*

Recientemente, se ha propuesto una nueva formulación multi-objetivo [7] con un enfoque diferente, en la cual se requiere minimizar dos funciones: la energía de unión y la desviación de la media cuadrática de las posiciones atómicas (RMSD). Este segundo objetivo interviene directamente en la estrategia de búsqueda de los algoritmos cuando se conoce la estructura cristalográfica del complejo ligando-macromolécula.

En [7] se presentó un estudio comparativo de varias metaheurísticas del estado del arte para resolver problemas con esta formulación, cuya conclusión principal fue que el SMPSO [9] sobresalía sobre el resto de técnicas. SMPSO es una variante multi-objetivo de PSO que utiliza un mecanismo para limitar la velocidad de las partículas, restringiendo así su movimiento en regiones dentro de los rangos del problema. SMPSO utiliza un archivo externo que almacena las soluciones no dominadas atendiendo a la distancia de *crowding* [1]; el archivo también se usa para el mecanismo de selección de líder.

La motivación en este artículo es evaluar un conjunto de variantes de SMPSO en el contexto del acoplamiento molecular. Estas variantes difieren sobre todo en el estimador de densidad del archivo externo. En concreto, se han usado: la contribución al hipervolumen, la distancia del coseno, un esquema agregativo y diferentes estrategias en la selección de líderes.

Con este objetivo, se han comparado y analizado las versiones propuestas de SMPSO utilizando 11 complejos moleculares del *benchmark* que se utilizó para testear AutoDock 4.2 [8]. Este conjunto de instancias incluye ligandos flexibles de diferentes tamaños y receptores mediante la aplicación de flexibilidad a sus residuos ARG-8. Para la evaluación de los algoritmos, se han utilizado dos indicadores de calidad que miden la convergencia y la diversidad de las aproximaciones de los frentes de Pareto obtenidos.

Este artículo se organiza de la siguiente forma: La Sección 2 describe el problema del acoplamiento molecular con formulación multi-objetivo. Los algoritmos que se utilizaron se han descrito en la Sección 3. En la Sección 4, se describe la metodología experimental y en la Sección 5 se analizan los resultados obtenidos. Finalmente, la Sección 6 incluye las conclusiones y futuros trabajos.

2. Acoplamiento Molecular

El principal objetivo en el problema del acoplamiento molecular es encontrar una conformación óptima entre el ligando (L) y el receptor (R) que tenga una energía de unión mínima. La interacción entre L y R puede definirse como la función de energía calculada a partir de tres componentes representando los grados de libertad: (1) la translación del ligando, que involucra los tres valores de los ejes (x, y, z) en el espacio de coordenadas cartesianas; (2) la orientación del ligando, modelada como cuatro variables del cuaternión incluyendo el ángulo de la pendiente (θ); y (3) la flexibilidad, representada por la rotación libre de torsión (ángulos diedros) del ligando y las cadenas laterales de ARG-8 del receptor.

- **Codificación de las soluciones:** Cada solución del problema se codifica con un vector de valores reales de $7 + n$ variables. Los tres primeros valores corres-

Estrategias de Archivo en PSO Multi-Objetivo para Docking

ponden a la translación del ligando (x, y, z) con un rango de $[0 \dots 120]$, ya que se han restringido los límites de las coordenadas de un espacio de malla previamente establecido para cada problema. Los cuatro siguientes valores corresponden a la orientación del ligando y/o receptor, y los n valores restantes son los ángulos diedros de torsión del ligando. Estas variables se miden en radianes, por lo que se codifican en un rango de $[-\pi, \pi]$.

- **Funciones de fitness:** la formulación de dos objetivos utilizada consiste en minimizar la $E_{unión}$ y el RMSD. La $E_{unión}$ es la función de energía usada por AutoDock, que se calcula de la siguiente forma:

$$E_{unión} = Q_{bound}^{R-L} + Q_{unbound}^{R-L} \quad (1)$$

donde Q_{bound}^{R-L} y $Q_{unbound}^{R-L}$ son los estados de unión y desunión del complejo ligando-receptor, respectivamente. Cada par de términos de la evaluación de la energía incluye las evaluaciones (Q) de las fuerzas de dispersión/repulsión (vdw), puentes de hidrógeno ($hbond$), interacciones electrostáticas ($elec$) y desolvatación (sol).

El RMSD es una medida de similitud entre la posición real del ligando en el receptor y la posición computada de éste. Una solución de acoplamiento molecular con un valor de RMSD menor de 2Å es considerada como una solución cuya predicción de conformación del ligando es bastante exacta.

El RMSD de dos estructuras idénticas se define de la siguiente manera:

$$RMSD_{ab} = \max(RMSD'_{ab}, RMSD'_{ba}), \text{ siendo } RMSD'_{ab} = \sqrt{\frac{1}{N} \sum_i \min_j r_{ij}^2} \quad (2)$$

La suma es el total de todos los N átomos pesados en una estructura a , el mínimo de todos los átomos de una estructura a con el mismo tipo de elemento que el átomo i en una estructura b .

3. Algoritmos

SMPSO es un algoritmo multi-objetivo de optimización basado en enjambres de partículas (MOPSO) que tiene dos características principales: un mecanismo de restricción de velocidad y un archivo acotado que almacena las soluciones no dominadas encontradas durante la búsqueda [9]. También incorpora un mecanismo de perturbación, implementado como un operador de mutación. El archivo contiene la aproximación actual del frente de Pareto encontrada por el algoritmo, y aplica el estimador de densidad de distancia de *crowding* [1] para decidir qué partícula quitar cuando está lleno. Este archivo también se usa en la estrategia de selección del líder, que consiste en un torneo binario en el que se seleccionan

Esteban López-Camacho *et al.*

aleatoriamente dos soluciones de su interior y se elije la de mayor valor de distancia de *crowding*. La mejor posición local de una partícula i se obtiene aplicando un test de dominancia con el resto de partículas del enjambre.

En [10] se realizó un estudio sobre diferentes mecanismos de selección del líder. En ese trabajo, la variante más prometedora consistía en reemplazar la distancia de *crowding* por los grados de contribución de las soluciones en el archivo externo según el indicador de hipervolumen [16]. De esta forma, la selección del líder se basa en un torneo binario que escoge la partícula con mayor grado de contribución al hipervolumen. Esta versión se llamó SMPSO_{hv} y es el segundo algoritmo escogido para la comparativa de este estudio.

Una nueva variante del SMPSO que proponemos aquí se basa en la distancia del coseno, que es una medida de similaridad entre dos vectores que miden el coseno del ángulo entre ellos. De este modo, dos vectores con la misma dirección tienen un valor de similaridad del coseno igual a cero, mientras que dos vectores perpendiculares tiene un valor igual a uno. Como todas las soluciones en un archivo externo son no dominadas, podemos definir un estimador de densidad fijando un punto de referencia y calculando la similaridad del coseno entre los vectores que forman las soluciones del archivo con respecto a ese punto de referencia. El problema de acoplamiento molecular tiene dos objetivos, así que se pueden ordenar las soluciones en el archivo a partir del primer objetivo y calcular, para cada solución, un valor de densidad sumando la similaridad del coseno de cada punto con respecto a los puntos anterior y posterior. Los puntos de los extremos tienen distancia de similaridad igual a cero. De esta forma, los puntos que tienen un mayor valor de densidad del coseno son los que se encuentran en las regiones más pobladas. El algoritmo resultante se llama SMPSOC. Es importante seleccionar un punto de referencia adecuado. Dado que en el estudio [7] se observó que los frentes tienen una forma convexa, se ha escogido en este estudio una aproximación al punto nadir, tomando los valores de los objetivos más altos de las soluciones en el archivo.

La cuarta versión de SMPSO en nuestro estudio es otra variante nueva que no usa archivo y denominamos SMPsOD. Para poder prescindir del archivo, tomamos una estrategia inspirada por MOEA/D [15], donde un problema multi-objetivo puede descomponerse en varios problemas mono-objetivo que pueden ser optimizados al mismo tiempo. De esta manera, se definen un conjunto de vectores de pesos uniformemente distribuidos $\lambda^1, \lambda^2, \dots, \lambda^N$, siendo N el tamaño del enjambre. Después, cada partícula i tiene asociada el vector λ^i y un vecindario definido por los vectores de pesos más cercanos en $\lambda^1, \lambda^2, \dots, \lambda^N$. La estrategia de escalado sigue el esquema de Techebycheff. La estrategia para conseguir el mejor local de una partícula i es el mismo procedimiento usado por MOEA/D para actualizar el vecindario y consiste en encontrar la mejor solución en el vecindario considerando los valores escalares de las partículas teniendo en cuenta sus vectores de pesos.

Por último, SMPSO está inspirado en el algoritmo OMOPSO propuesto en [12], por lo que se ha incluido también en la comparativa como PSO multi-objetivo de referencia.

Estrategias de Archivo en PSO Multi-Objetivo para Docking

Tabla 1. Instancias estudiadas. El código PDB, la estructura cristalográfica y la resolución obtenidos de la base de datos PDB.

Código PDB	Complejos proteína-ligando	Resolución (Å)
1AJV	HIV-1 protease/AHA006	2.00
1AJX	HIV-1 protease/AHA001	2.00
1BV9	HIV-1 protease/ α -D-glucose	2.20
1D4K	HIV-1 protease/Macrocylic peptidomimetic inhibitor 8	1.85
1G2K	HIV-1 protease/AHA047	1.95
1HIV	HIV-1 protease/U75875	2.00
1HPX	HIV-1 protease/KN1-272	2.00
1HTF	HIV-1 protease/GR126045	2.20
1HTG	HIV-1 protease/GR137615	2.00
1HVH	HIV-1 protease/Q8261	1.80
2UPJ	HIV-1 protease/U100313	3.00

En resumen, en este estudio se han incluido OMOPSO y cuatro variantes SMPSO con diferentes estrategias de archivo: basado en la distancia de *crowding* (SMPSO original), en la contribución del hipervolumen (SMPSO_{hv}), en la distancia del coseno (SMPSOC) y sin ningún archivo (SMPSOD).

4. Experimentación

Para comparar los algoritmos se llevó a cabo un estudio experimental en el que se utilizó un *benchmark* de 11 instancias de flexibilidad en el receptor y ligando. Estos complejos son instancias de acoplamiento molecular de varios grados de complejidad, ya que el ligando presenta un amplio rango de tamaños (desde ligandos pequeños hasta más grandes).

Los estudios de acoplamiento molecular que se llevaron a cabo utilizando estas instancias en [8] mostraron que los problemas más complejos corresponden a aquellos que involucran ligandos más pequeños. Esto se debe a que la flexibilidad añadida a los residuos ARG-8 de los receptores incrementa el espacio de interacción del ligando. Las instancias se obtuvieron de la base de datos PDB¹.

La Tabla 1 resume el conjunto de instancias que se seleccionaron mostrando el código PDB, el nombre de las estructuras cristalográficas y la resolución de la estructura (Å). Para todas las instancias, los grados de torsión para los ligandos y los receptores fueron 10 y 6, respectivamente, seleccionando aquellas torsiones que permitieron el menor número de átomos móviles con respecto al núcleo del ligando.

La metodología que se aplicó consistió en hacer 30 ejecuciones independientes para cada combinación de algoritmo e instancia. A partir de estas ejecuciones, se calcularon la mediana y el rango intercuartílico (IQR) como medidas de tendencia central y dispersión estadística, respectivamente. Se consideraron dos indicadores de calidad para evaluar los algoritmos: el Hipervolumen (I_{HV}) [16] y el indicador Epsilon aditivo ($I_{\epsilon+}$) [17]. El primero tiene en cuenta tanto la convergencia como la diversidad y el segundo proporciona una medida del grado de convergencia de las aproximaciones de los frentes de Pareto obtenidos. En

¹ In URL: <http://www.rcsb.org/pdb/home/home.do>

Esteban López-Camacho *et al.*

este caso, como se está abordando la resolución de problemas reales complejos, sus frentes de Pareto son desconocidos, aunque estos frentes son necesarios para calcular estas dos métricas. Para solucionar este problema, se generó un frente de referencia para cada instancia mediante la combinación de todas las soluciones no dominadas generadas en todas las ejecuciones de todos los algoritmos.

Para los experimentos, se utilizó la implementación de los 5 algoritmos de jMetalCpp [5], en combinación con AutoDock 4.2 para evaluar las soluciones. Como plataforma de ejecución, se utilizó una infraestructura de *grid computing* Condor², como *middleware* que permite coordinar tareas distribuidas utilizando hasta 400 núcleos.

En cuanto a la configuración de parámetros, el tamaño del enjambre fue de 150 y la condición de parada se alcanzó a las 1.500.000 evaluaciones para todos los algoritmos. Estos valores se seleccionaron en línea con aquellos utilizados por defecto en AutoDock [11]. El tamaño del archivo, en aquellas situaciones en las que se aplica, se configuró con tamaño 100. Todas las versiones de SMPSO utilizaron mutación polinomial con índice de distribución $\eta_m = 20$, que se aplica a un sexto de las partículas del enjambre. Los coeficientes de aceleración C_1 y C_2 se configuraron a 1,5 y el peso de inercia a $w = 0,9$.

Con esta configuración de parámetros se intentó realizar una comparación justa, manteniendo el resto de parámetros del SMPSO y OMOPSO con los mismos valores que en los artículos donde se describieron.

5. Resultados y Análisis

El primer análisis corresponde a los resultados según el indicador de hipervolumen I_{HV} . Este indicador calcula la suma de los volúmenes contribuidos por cada punto en el frente de Pareto (soluciones no dominadas) con respecto a un punto de referencia. Por lo tanto, cuanto más alta sea la convergencia y el grado de diversidad de un frente, mayor (y mejor) es el valor I_{HV} .

La Tabla 2 muestra la mediana y el rango intercuartílico de las distribuciones calculadas (a partir de 30 ejecuciones independientes) de I_{HV} , para el conjunto de 11 instancias de acoplamiento molecular y para los cinco algoritmos comparados. SMPSO_{hv} obtiene los mejores valores medianos de I_{HV} para todas las instancias moleculares y SMPSOD es la segunda técnica con mejor rendimiento. Cabe destacar que algunos resultados de OMOPSO tienen un I_{HV} igual a cero. Esto ocurre cuando todos los puntos de los frentes producidos son dominados por el punto de referencia. En cambio, todas las versiones de SMPSO obtuvieron valores de I_{HV} mayores que cero, lo que indica que todas ellas son capaces de producir soluciones dentro de los límites del punto de referencia.

En el caso de $I_{\epsilon+}$, puede extraerse una observación similar a partir de la Tabla 2, es decir, SMPSO_{hv} muestra los mejores resultados para todas las instancias, seguido por SMPSOD y SMPSO (cuanto más bajo es el valor de $I_{\epsilon+}$, mejor es el resultado). Para este indicador, SMPSOC obtuvo la segunda mejor mediana sólo para la instancia 1VB9.

² In URL: <http://research.cs.wisc.edu/htcondor/>

Estrategias de Archivo en PSO Multi-Objetivo para Docking

Tabla 2. Mediana y rango intercuartílico de I_{HV} y $I_{\epsilon+}$ para cada algoritmo e instancia. La mejor y la segunda mejor mediana de los resultados están sombreadas con un gris oscuro y uno claro, respectivamente.

I_{HV}	SMPSO	SMPSO _{hv}	SMPSOD	SMPSOC	OMOPSO
IAJV	3.65e-01 _{3,1e-02}	4.33e-01 _{4,0e-02}	3.63e-01 _{4,6e-02}	3.55e-01 _{4,8e-02}	0.00e+00 _{0,0e+00}
IAJX	4.31e-01 _{2,5e-02}	5.06e-01 _{2,7e-02}	4.74e-01 _{3,7e-02}	4.43e-01 _{3,6e-02}	0.00e+00 _{0,0e+00}
ID4K	6.67e-01 _{1,1e-02}	8.48e-01 _{1,1e-01}	7.11e-01 _{9,4e-02}	7.35e-01 _{9,2e-02}	0.00e+00 _{0,0e+00}
IG2K	8.84e-01 _{3,2e-02}	4.58e-01 _{9,9e-02}	3.82e-01 _{4,1e-02}	3.52e-01 _{5,2e-02}	0.00e+00 _{0,0e+00}
IHV	4.86e-01 _{2,0e-01}	6.74e-01 _{2,9e-02}	5.87e-01 _{7,1e-02}	4.66e-01 _{2,4e-01}	0.00e+00 _{0,0e+00}
IHPX	3.60e-01 _{1,8e-01}	6.30e-01 _{9,7e-02}	4.77e-01 _{1,0e-01}	4.63e-01 _{1,4e-01}	0.00e+00 _{0,0e+00}
IHTF	2.61e-01 _{3,3e-01}	4.17e-01 _{2,4e-01}	3.96e-01 _{7,9e-02}	2.77e-01 _{3,1e-01}	0.00e+00 _{0,0e+00}
IHTG	8.33e-01 _{2,3e-01}	1.46e-01 _{9,6e-02}	1.03e-01 _{8,2e-02}	7.13e-01 _{2,3e-01}	0.00e+00 _{0,0e+00}
IHVH	7.78e-01 _{4,7e-02}	8.69e-01 _{3,3e-03}	7.70e-01 _{2,4e-02}	7.85e-01 _{2,9e-02}	0.00e+00 _{0,0e+00}
IVB9	4.10e-01 _{1,2e-01}	5.09e-01 _{3,6e-02}	4.12e-01 _{1,1e-01}	4.38e-01 _{3,3e-02}	0.00e+00 _{0,0e+00}
2UPJ	5.82e-01 _{9,6e-02}	6.96e-01 _{4,1e-02}	6.27e-01 _{7,4e-02}	6.20e-01 _{6,8e-02}	1.99e-01 _{6,4e-01}
$I_{\epsilon+}$	SMPSO	SMPSO _{hv}	SMPSOD	SMPSOC	OMOPSO
IAJV	5.12e-01 _{1,0e-01}	3.94e-01 _{6,7e-02}	5.35e-01 _{1,0e-01}	5.46e-01 _{1,0e-01}	5.31e+00 _{2,0e+00}
IAJX	2.31e-01 _{1,1e-01}	1.32e-01 _{3,3e-02}	1.94e-01 _{6,1e-02}	2.57e-01 _{9,4e-02}	2.54e+00 _{2,2e+00}
ID4K	2.06e-01 _{8,6e-02}	4.41e-02 _{2,2e-01}	1.54e-01 _{7,3e-02}	1.57e-01 _{8,1e-02}	8.81e+00 _{4,1e+00}
IG2K	4.29e-01 _{1,7e-01}	2.81e-01 _{2,0e-01}	4.75e-01 _{9,7e-02}	5.15e-01 _{1,1e-01}	6.01e+00 _{2,3e+00}
IHV	3.95e-01 _{3,6e-01}	9.03e-02 _{4,4e-02}	2.66e-01 _{1,2e-01}	4.36e-01 _{3,2e-01}	4.91e+00 _{1,1e+00}
IHPX	4.25e-01 _{2,8e-01}	1.30e-01 _{9,2e-02}	2.95e-01 _{1,2e-01}	3.17e-01 _{1,7e-01}	1.13e+01 _{5,7e+00}
IHTF	5.60e-01 _{1,5e+00}	5.46e-01 _{3,7e-01}	5.64e-01 _{1,1e-01}	6.85e-01 _{4,3e-01}	1.49e+00 _{6,2e-01}
IHTG	0.97e-01 _{1,4e-01}	8.35e-01 _{9,3e-02}	8.84e-01 _{8,8e-02}	9.23e-01 _{2,1e-01}	1.21e+01 _{7,7e+00}
IHVH	1.46e-01 _{4,4e-02}	6.12e-02 _{4,8e-03}	1.47e-01 _{3,9e-02}	1.52e-01 _{3,1e-02}	5.11e+00 _{2,4e+00}
IVB9	3.34e-01 _{2,2e-01}	1.96e-01 _{7,7e-02}	3.44e-01 _{1,8e-01}	2.97e-01 _{1,3e-01}	9.31e+00 _{1,6e+00}
2UPJ	2.86e-01 _{9,7e-02}	1.76e-01 _{3,2e-02}	2.25e-01 _{1,4e-01}	2.70e-01 _{5,1e-02}	7.74e-01 _{4,0e+00}

Tabla 3. Media de los rankings de Friedman con los p -valores (0,05) ajustados de Holm de los algoritmos comparados para el conjunto de 11 instancias de acoplamiento molecular. El símbolo * indica el algoritmo de control y la columna de la derecha contiene el ranking global de las posiciones con respecto a I_{HV} y $I_{\epsilon+}$.

Hipervolumen (I_{HV})			Epsilon ($I_{\epsilon+}$)			Global	
Algoritmo	$FriRank$	$HolmAp$	Algoritmo	$FriRank$	$HolmAp$	Algoritmo	Rank
*SMPSO _{hv}	1.01	-	*SMPSO _{hv}	1.00	-	SMPSO _{hv}	2
SMPSOD	2.54	2.18e-02	SMPSOD	2.45	3.09e-02	SMPSOD	4
SMPSOC	3.09	3.85e-03	SMPSO	2.99	6.02e-03	SMPSOC	5
SMPSO	3.36	1.36e-03	SMPSOC	3.54	4.79e-04	SMPSO	5
OMOPSO	4.99	1.19e-08	OMOPSO	4.98	1.19e-08	OMOPSO	10

Estos resultados se comparan además con una confianza estadística (en este estudio, p -valor = 0.05) centrándose en la distribución de cada una de las dos métricas estudiadas. En concreto, hemos aplicado el ranking de Friedman y los tests post-hoc de Holm [14] para saber qué algoritmos son estadísticamente peores que aquella técnica utilizada como control (por ejemplo, la variante con mejor ranking).

De esta manera, como se muestra en la Tabla 3, SMPSO_{hv} es la variante en primera posición del ranking creado por el test de Friedman para los dos indicadores (I_{HV} y $I_{\epsilon+}$), estando SMPSOD en segunda posición. Por lo tanto, SMPSO_{hv} se establece como algoritmo de control en los tests post-hoc de Holm, mediante el cual se compara con el resto de algoritmos. Los p -valores ajustados ($HolmAp$ en la Tabla 3) resultantes de estas comparaciones son, para el resto de variantes (SMPSO, SMPSOD, SMPSO y OMOPSO), más bajos que el nivel de confianza (0.05), significando que SMPSO_{hv} es estadísticamente mejor que el

Esteban López-Camacho *et al.*

resto de algoritmos para las instancias utilizadas. SMPSO y SMPSOC obtuvieron globalmente rendimientos similares, aunque SMPSOC obtuvo mejor posición en los *rankings* que SMPSO según I_{HV} .

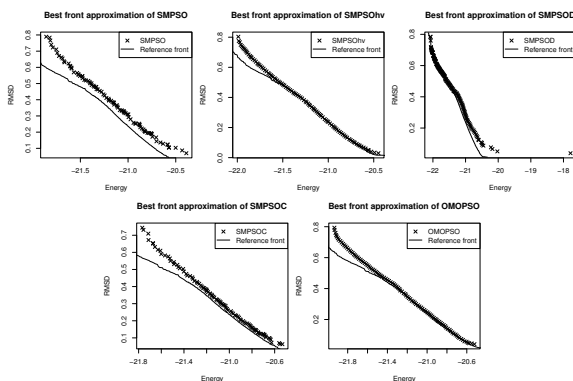


Figura 1. Frentes con los mejores valores I_{HV} sobre el problema 1AJX.

A modo de ilustración, la Figura 1 contiene los frentes encontrados con mejor hipervolumen para el problema 1AJX. Una observación interesante en este sentido consiste en la habilidad de SMPSOD para obtener soluciones no dominadas en la región del frente de Pareto de referencia con baja energía y altos valores de RMSD (gráfico arriba a la derecha de la Figura 1). En contraste con los otros algoritmos comparados, SMPSOD es capaz de cubrir adecuadamente esta área, además de otras áreas con bajos RMSD. Por lo tanto, una implementación híbrida de SMPSO usando una estrategia agregativa (sin archivo) como en MOEA/D podría cubrir el frente de referencia con soluciones no dominadas en los dos extremos de los objetivos. Este es un comportamiento deseable ya que podría arrojar luz en la obtención de nuevos sitios activos de las moléculas estudiadas.

En resumen, SMPSO_{hv} muestra el mejor comportamiento global seguido por SMPSOD. Intuitivamente, el primero obtuvo el mejor I_{HV} , ya que ejecuta un método de selección de líder de soluciones no dominadas (desde el archivo externo) con las mayores contribuciones al hipervolumen. Esto significa que las partículas del enjambre son guiadas por líderes con grandes contribuciones al

Estrategias de Archivo en PSO Multi-Objetivo para Docking

hipervolumen, lo que permite que SMPSO_{hv} obtenga no sólo altos valores de I_{HV} , sino también resultados más precisos en términos de $I_{\epsilon+}$.

Analizando los resultados desde un punto de vista biológico, el conjunto de instancias utilizadas son problemas cuyo receptor HIV-proteasa contiene un túnel en el que se encuentra el sitio catalítico haciendo más compleja la interacción receptor-ligando. En el estudio llevado a cabo por [8] con el mismo conjunto de instancias, los resultados para los valores de energía y RMSD mostraron resultados no favorables a diferencia de los obtenidos en este estudio para el SMPSO_{hv} . En este trabajo, dado que se han minimizado la energía de unión y los valores RMSD, la técnica presentada ha mostrado ser favorable en casos en los que el sitio activo enzimático está localizado en una región en la macromolécula de difícil acceso para el ligando dada la aproximación multi-objetivo. Esta aplicación sería útil en estudios *in silico* en los que se simule la interacción de análogos del ligando co-cristalizado para una determinada diana terapéutica.

6. Conclusiones

En este artículo se analizaron nuevas variantes de SMPSO, una técnica de optimización multi-objetivo de enjambre de partículas basada en estrategias de archivo, aplicándolas a un *benchmark* de complejos receptor-ligando flexibles. El problema se ha formulado como un problema de optimización de dos objetivos en el que se minimiza la energía de unión y la desviación de la media cuadrática de las posiciones atómicas (RMSD) de las coordenadas atómicas del ligando.

Este estudio revela que los frentes obtenidos por SMPSO_{hv} son los de mayor calidad, seguidos del SMPSOD, SMPSOC y SMPSO. SMPSO_{hv} obtiene el mejor I_{HV} dado que implementa un método de selección del líder de aquellas soluciones no dominadas (de un archivo externo) que tienen mayores contribuciones en hipervolumen, lo que parece ser responsable de los mejores valores de convergencia y diversidad en esta comparación. La variante SMPSOD cubre el frente de referencia con soluciones no dominadas en los extremos de los dos objetivos, por ejemplo, con valores de energía y RMSD muy bajos. De esta forma, una implementación de SMPSO utilizando una estrategia agregativa como MOEA/D, debería cubrir el frente de referencia con soluciones no dominadas en los extremos de ambos objetivos. En estos casos, los valores de energía obtenidos serían muy negativos y los valores RMSD más positivos.

Esto abriría una futura línea de investigación para la selección y el estudio de soluciones que pueden ser evaluadas desde un punto de vista farmacológico. Asimismo, como extensión natural de este trabajo, se pretende validar las conclusiones expuestas con un número mayor de instancias y utilizando otros indicadores de calidad.

Agradecimientos. Este trabajo ha sido parcialmente financiado por los proyectos TIN2011-25840 (Ministerio de Ciencia e Innovación) y P11-TIC-7529 y P12-TIC-1519 (Plan Andaluz de Investigación, Desarrollo e Innovación). Este

Esteban López-Camacho *et al.*

trabajo se financia parcialmente por la Acción COST CA15140 del programa de Cooperación Europea en Ciencia y Tecnología.

Referencias

1. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* 6(2), 182–197 (2002)
2. García-Godoy, M.J., López-Camacho, E. and García Nieto, J. and Nebro, A. J. and Aldana-Montes, J. F.: Solving Molecular Docking Problems with Multi-Objective Metaheuristics. *Molecules* 20(6), 10154–10183 (2015)
3. Gu, J., Yang, X., Kang, L., Wu, J., Wang, X.: MoDock: A multi-objective strategy improves the accuracy for molecular docking. *Algorithm Mol. Biol.* 10, 8 (2015)
4. Janson, S., Merkle, D., Middendorf, M.: Molecular docking with multi-objective particle swarm optimization. *Appl. Soft Comput.* 8(1), 666–675 (2008)
5. López-Camacho, E., García-Godoy, M.J., Nebro, A.J., Aldana-Montes, J.F.: jMetalCpp: optimizing molecular docking problems with a C++ metaheuristic framework. *Bioinformatics* 30(3), 437–438 (Feb 2014)
6. López-Camacho, E., García-Godoy, M.J., García-Nieto, J., Nebro, A.J., Aldana-Montes, J.F.: Solving molecular flexible docking problems with metaheuristics: A comparative study. *Appl. Soft Comput.* 28, 379–393 (2015)
7. López-Camacho, E., García-Godoy, M.J., García-Nieto, J., Nebro, A.J., Aldana-Montes, J.F.: A new multi-objective approach for molecular docking based on rmsd and binding energy. In: 3rd Int. Conf. on Alg. for Comp. Bio. pp. In-Press (2016)
8. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J.: AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* 30(16), 2785–2791 (2009)
9. Nebro, A., Durillo, J., García-Nieto, J., Coello Coello, C.A., Luna, F., Alba, E.: SM-PSO: A new PSO-based metaheuristic for multi-objective optimization. In: *IEEE Symp. on Comp. Intel. in Multi-Criteria Decision-Making*, pp. 66–73 (2009)
10. Nebro, A., Durillo, J., Coello, C.C.: Analysis of leader selection strategies in a mopso. In: *IEEE Cong. on Evol. Comp. (CEC)*, pp. 3153–3160 (June 2013)
11. Norgan, A.P., Coffman, P.K., Kocher, J.P.A., Katzmann, D.J., Sosa, C.P.: Multi-level Parallelization of AutoDock 4.2. *J. Cheminform.* 3(1), 12 (2011)
12. Reyes, M., Coello Coello, C.: Improving PSO-based multi-objective optimization using crowding, mutation and ϵ -dominance. In: 3th Int. Conf. on Evol. Multi-Criterion Opt., (EMO 2005). LNCS, vol. 3410, pp. 509–519. Springer (2005)
13. Sandoval-Perez, A., Becerra, D., Vanegas, D., Restrepo-Montoya, D., Niño, F.: A multi-objective optimization energy approach to predict the ligand conformation in a docking process. In: *EuroGP*, pp. 181–192 (2013)
14. Sheskin, D.J.: *Handbook of Parametric and Nonparametric Statistical Procedures*. Chapman & Hall/CRC (2007)
15. Zhang, Q., Li, H.: MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Trans. on Evol. Comp.* 11(6), 712–731 (2007)
16. Zitzler, E., Thiele, L.: Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach. *IEEE Trans. on Evol. Comp.* 3(4), 257–271 (1999)
17. Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., da Fonseca, V.G.: Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Trans. on Evol. Comp.* 7(2), 117 – 132 (2003)