



Descubrimiento de fármacos basado en cribado virtual refinado con enfoques neuronales paralelos



H. Pérez-Sánchez^a, G. Cano^{b,*}, J. García-Rodríguez^b y J.M. Cecilia^a

^a Departamento de Informática, Universidad Católica San Antonio de Murcia (UCAM), Campus de los Jerónimos s/n, E30107, Guadalupe, Murcia, España

^b Departamento de Tecnología Informática y Computación, Universidad de Alicante, Apartado 99, E03080, Alicante, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 31 de enero de 2014

Aceptado el 18 de junio de 2014

On-line el 3 de diciembre de 2014

Palabras clave:

Redes neuronales
Investigación clínica
Descubrimiento de fármacos
Cribado virtual
Computación paralela
Inteligencia computacional

R E S U M E N

En el campo de la medicina clínica es crucial poder determinar la seguridad y la eficacia de los fármacos actuales y además acelerar el descubrimiento de nuevos compuestos activos. Para ello se llevan a cabo ensayos de laboratorio, que son métodos muy costosos y que requieren mucho tiempo. Sin embargo, la bioinformática puede facilitar enormemente la investigación clínica para los fines mencionados, ya que proporciona la predicción de la toxicidad de los fármacos y su actividad en enfermedades nuevas, así como la evolución de los compuestos activos descubiertos en ensayos clínicos.

Esto se puede lograr gracias a la disponibilidad de herramientas de bioinformática y métodos de cribado virtual por ordenador (CV) que permitan probar todas las hipótesis necesarias antes de realizar los ensayos clínicos, tales como el *docking* estructural, mediante el programa BINDSURF.

Sin embargo, la precisión de la mayoría de los métodos de CV se ve muy restringida a causa de las limitaciones presentes en las funciones de afinidad o *scoring* que describen las interacciones biomoleculares, e incluso hoy en día estas incertidumbres no se conocen completamente. En este trabajo abordamos este problema, proponiendo un nuevo enfoque en el que las redes neuronales se entrenan con información relativa a bases de datos de compuestos conocidos (proteínas diana y fármacos), y se aprovecha después el método para incrementar la precisión de las predicciones de afinidad del método de CV BINDSURF.

© 2014 CIMNE (Universitat Politècnica de Catalunya). Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Virtual screening refined with parallel neural approaches improves drug discovering

A B S T R A C T

In the field of clinical research, it is crucial to determine the safety and efficacy of current drugs and further accelerate the discovery of new active compounds. The traditional methods performed expensive laboratory tests. These methods are very costly and require a long time. However, bioinformatics can greatly facilitate the clinical research for the above purposes, providing the prediction of drug toxicity and activity in new diseases and evaluating the validity of active compounds discovered in clinical trials.

This can be achieved through the availability of bioinformatics tools and methods of computer virtual screening (VS) that allow the test all the necessary hypotheses before doing the clinical trials, such as structural docking using the BINDSURF program.

However, the accuracy of most VS methods is severely restricted due to constraints in the affinity or scoring functions describing biomolecular interactions, and even today these uncertainties are not fully known. In this paper we address this problem by proposing a new approach in which neural networks are trained with information on databases of known compounds (target proteins and drugs), for further exploiting the method to improve the accuracy of VS BINDSURF predictions of affinity.

© 2014 CIMNE (Universitat Politècnica de Catalunya). Published by Elsevier España, S.L.U. All rights reserved.

Keywords:

Neural networks
Clinical research
Drug discovery
Virtual screening
Parallel computing
Computational intelligence

* Autor para correspondencia.

Correos electrónicos: horacio@ucam.edu (H. Pérez-Sánchez), gcano@dtic.ua.es (G. Cano), jgarcia@dtic.ua.es (J. García-Rodríguez), jmcecilia@ucam.edu (J.M. Cecilia).

1. Introducción

En la investigación clínica es crucial determinar la seguridad y la eficacia de los fármacos actuales y acelerar el descubrimiento de compuestos activos, especialmente para procesar grandes conjuntos de datos que describen estructuras de proteínas conocidas en bases de datos biológicas, tales como bases de datos de proteínas (PDB, por sus siglas en inglés) [1], y también derivados de los datos genómicos utilizando técnicas como el modelado de homología [2]. Los ensayos de laboratorio y optimización de compuestos son métodos caros y lentos. Sin embargo, la bioinformática puede facilitar enormemente la investigación clínica para los fines mencionados, proporcionando la predicción de la toxicidad de fármacos y su actividad en enfermedades nuevas, así como la evolución de los compuestos activos descubiertos en ensayos clínicos.

Esto se puede lograr gracias a la disponibilidad de herramientas de bioinformática y métodos de cribado virtual (CV) que permitan probar todas las hipótesis necesarias antes de realizar los ensayos clínicos. Los métodos de CV, como el acoplamiento molecular (*docking*), fallan en la predicción de la toxicidad y las predicciones de actividad, ya que están limitados por el acceso a recursos computacionales, e incluso los métodos más rápidos de CV no pueden procesar grandes bases de datos biológicas en un tiempo razonable. Por lo tanto, estas restricciones imponen serias limitaciones en muchas áreas relacionadas de la investigación.

El uso de arquitecturas hardware masivamente paralelas y orientadas al rendimiento, tales como las unidades de procesamiento gráfico (GPU, por sus siglas en inglés), pueden ayudar a superar este problema. Las GPU han ganado popularidad en el campo de la computación de alto rendimiento (HPC, por sus siglas en inglés) mediante la combinación de una impresionante potencia de cálculo con los requisitos más exigentes de gráficos en tiempo real y el lucrativo mercado masivo que supone la industria del videojuego [3]. Los científicos han aprovechado su potencia de cálculo en el dominio computacional y las GPU se han convertido en un recurso clave de aplicaciones en las que el paralelismo es el denominador común [4]. Para mantener este impulso, NVIDIA ha añadido progresivamente nuevas características de hardware para su gama de GPU, con la arquitectura Kepler [5] como hito más reciente. Por lo tanto, las GPU son muy adecuadas para superar la falta de recursos computacionales en los métodos de CV, permitiendo la aceleración de los cálculos necesarios y la introducción de mejoras en los modelos biofísicos no asequibles en el pasado [6]. Hemos trabajado en esta dirección, que muestra cómo los métodos de CV se pueden beneficiar del uso de la GPU [7–9]. Por otra parte, otra carencia importante de los métodos de CV es que por lo general asumen que el lugar de unión que deriva de una sola estructura cristalina será el mismo para los diferentes ligandos, mientras que se ha demostrado que esto no siempre sucede [10], y por lo tanto es crucial evitar esta simplificación. En este trabajo se presenta una nueva metodología denominada CV BINDSURF que se aprovecha de la alta intensidad de cálculo paralelo masivo de las GPU para acelerar los cálculos requeridos, con máquinas convencionales de bajo consumo y coste, que proporcionan información nueva y útil sobre las proteínas objetivo, con el fin de mejorar las predicciones clave en toxicidad y en grado de actividad. En BINDSURF, una gran base de datos de ligandos se criba simultáneamente sobre toda la superficie de la proteína objetivo. Posteriormente, la información obtenida acerca de los nuevos puntos potenciales de acoplamiento en las proteínas se utiliza para realizar cálculos más detallados utilizando cualquier método de CV, pero solo para un conjunto de ligandos reducido y seleccionado.

Otros autores han realizado estudios de métodos de CV sobre superficies de proteínas enteras [11] utilizando diferentes enfoques y cribando bases de datos de ligandos pequeños pero, por lo que sabemos, ninguno de ellos se ha implementado en GPU ni se ha utilizado de la misma manera que BINDSURF.

Sin embargo, la precisión de la mayoría de los métodos de CV se ve limitada por las limitaciones en las funciones de afinidad o *scoring* que describen las interacciones biomoleculares, e incluso hoy en día estas incertidumbres no se conocen completamente. En este trabajo abordamos este problema, proponiendo un nuevo enfoque en el que las redes neuronales se entrenan con bases de datos de compuestos activos (fármacos) e inactivos conocidos y posteriormente se utilizan para mejorar las predicciones mediante BINDSURF.

El resto del trabajo se organiza de la siguiente manera. El segundo apartado introduce brevemente el conocimiento previo para comprender mejor el resto del artículo; la tercera parte presenta nuestra propuesta de uso de redes neuronales para mejorar las predicciones de CV; finalmente, se muestran las conclusiones y posibles direcciones para el trabajo futuro.

2. Metodología

En este apartado se describen los métodos que hemos utilizado para la predicción de la afinidad de la proteína-ligando: el método de cribado virtual BINDSURF, una red neuronal entrenada con datos de similitud química de los compuestos activos e inactivos conocidos y la similitud molecular que conforman las variables de entrada de la red neuronal.

2.1. Cribado virtual con BINDSURF

La principal idea que subyace en este método de cribado virtual con BINDSURF es la de una técnica de detección sobre la superficie de proteínas, ejecutada en paralelo sobre GPU. Esencialmente, los métodos de CV procesan una gran base de datos de moléculas con el fin de encontrar cuál encaja mejor con algunos criterios establecidos [12]. En el caso del descubrimiento de nuevos fármacos potenciales, optimización de compuestos, valoración de la toxicidad y las etapas adicionales del proceso de descubrimiento de fármacos, examinamos una gran base de datos de compuestos para encontrar una pequeña molécula, que interactúa de una manera deseada con uno o varios receptores. Entre los muchos métodos de CV disponibles para este propósito hemos decidido utilizar el acoplamiento proteína-ligando [13,14]. Estos métodos tratan de obtener predicciones rápidas y exactas de la conformación 3D que adopta un ligando cuando interactúa con una determinada proteína objetivo, y también la fuerza de esta unión, en términos del valor de su función de afinidad. Normalmente, las simulaciones de acoplamiento se llevan a cabo en una parte muy concreta de la superficie de la proteína en métodos como Autodock [15], Glide [16] y Dock [17], por nombrar algunos. Esta región se deriva comúnmente desde la posición de un ligando particular en la estructura cristalina, o de la estructura cristalina de la proteína sin ligando. El primero se puede realizar cuando la proteína se ha cocrystalizado con el ligando, pero puede ocurrir que no haya una estructura cristalina del par ligando-proteína que se encuentre disponible. Sin embargo, el principal problema es suponer, una vez que se especifica el sitio de unión, que muchos ligandos diferentes van a interactuar con la proteína en la misma región, descartando completamente las otras áreas de la proteína.

Ante este problema, se propone la división de toda la superficie de la proteína en regiones definidas. Posteriormente, se llevan a cabo simulaciones de acoplamiento para cada ligando en todos los puntos especificados de la proteína simultáneamente. Siguiendo este enfoque, los nuevos puntos de acceso se pueden localizar tras un examen de la distribución de la función de afinidad en toda la superficie de la proteína. Esta información podría conducir al descubrimiento de nuevos lugares de unión. Si comparamos este enfoque con una simulación de acoplamiento típico, realizado solo

Download English Version:

<https://daneshyari.com/en/article/1702524>

Download Persian Version:

<https://daneshyari.com/article/1702524>

[Daneshyari.com](https://daneshyari.com)