BINDSURF: a fast blind virtual screening methodology on GPUs

I. Sánchez-Linares, H. Pérez-Sánchez, J.M. Cecilia, J.M. García

Computer Engineering and Technology Department, University of Murcia, Spain

INTRODUCTION

In clinical research it is very important to determine the safety and effectiveness of current drugs and it implies to be able to process the vast amount of protein structure data available in biological databases like PDB and also derived from genomic data using techniques as homology modelling (Sánchez 1998).
Screenings in lab and compound optimization are expensive and slow methods, but bioinformatics can vastly help clinical research by providing prediction of the toxicity of drugs and activity in non-tested targets. Nevertheless current Virtual Screening (VS) methods like docking fail to make good toxicity and activity predictions since they take the assumption that the binding site derived from the crystal structure will be the same for different ligands, while it has been shown that this not always happens (Branningan 2010) and also because even the fastest VS methods cannot process all the required data in a reasonable time-frame.
In this work in progress we present a new VS methodology called BINDSURF, which takes advantage of the use of GPUs to speed-up the required calculations and to provide new and useful information about targets and thus improving key toxicity and activity predictions.

METHODS

We used the version 4.0 of the CUDA programming model (NVIDIA 2011) in our parallel implementation with a NVIDIA Tesla C2050 GPU. In order to obtain speedup measurements versus the sequential counterpart a core of an Intel Xeon E5450 PC was used. To check reproducibility between sequential and GPU version, calculations for the former were carried out in a 816 node cluster from the Supercomputing Center of Murcia.
For each ligand of the database, a set of 100 different conformations using the docking program FlexScreen (Kokh 2008) is generated. Next, rigid protein-ligand docking simulations are performed over the whole protein surface, divided into spheres of fixed volume, centered around the alpha carbons of each residue. The scoring function uses highly GPU optimized non-bonded interaction kernels (Guerrero 2011) for the description of the electrostatic and Van der Waals interactions between the ligand and the protein. A GPU optimized Monte Carlo minimization algorithm is used to minimize the total energy of the system. In the final output we find for each ligand detailed information about the protein spots where the strongest interactions are found for the different ligand conformations.
This information can be parsed to PyMOL (www.pymol.org) to obtain a graphical depiction of the results. These results can be later used in a more detailed VS methodology to screen only the ligands with the highest estimated affinities in the hotspots found by BINDSURF.

RESULTS AND DISCUSSION

We initially performed redocking simulations for different PDB structures and checked that in most of the tested cases BINDSURF finds efficiently the crystallographic binding mode.
We also obtained concordance with some other methods that try to predict the binding site based on the protein structure alone (Ghersi 2009). In *Figure 1A* we show how the strongest interaction spot (blue sphere) coincides with crystal binding site.
Figure 1B shows how its binding affinity is clearly differentiated from the other low binding affinity interaction spots.
We compared also with the final binding poses obtained using very long trajectories in Molecular Dynamics simulations in Supercomputers (Buch 2011, Ron 2011, Shan 2011), as we can see in *Figure 2.A*, obtaining the same results. It must be noticed that with BINDSURF the dynamical information about the binding process cannot be obtained.
On average, a 600 residues protein is processed on the GPU in 120 minutes, running 64 times faster than the sequential counterpart. In the latter a ligand per sphere is processed in 13 minutes. For sequential speed comparison with other standard single binding site docking methods like Vina or GOLD a ligand per sphere is processed in around 100 seconds.
In cases with proteins whose binding site depends on the ligand our method also works efficiently; in *Figure*

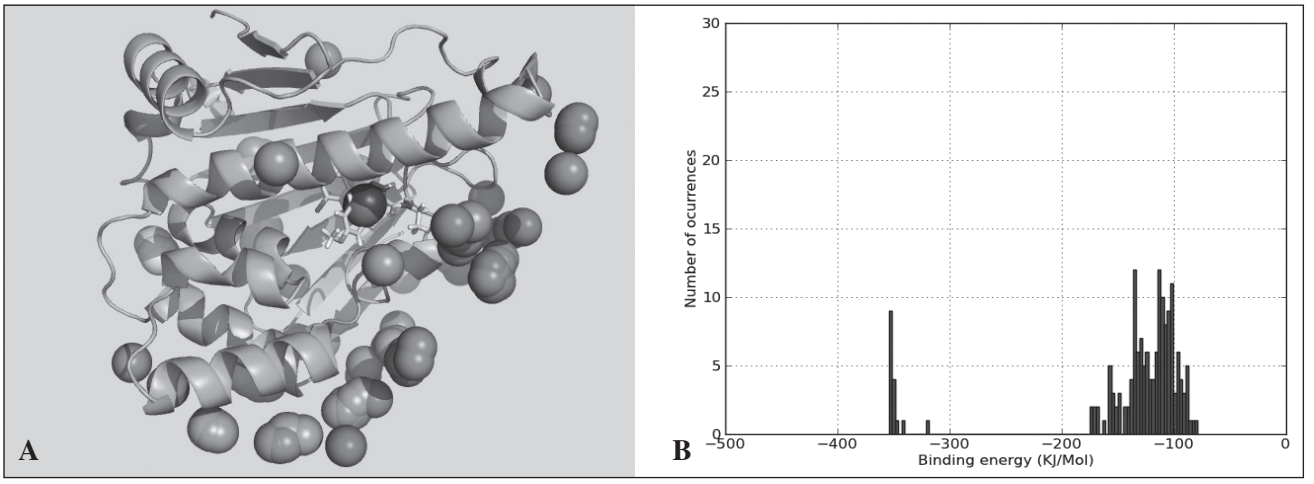

Figure 1 - *(A) PDB: 2BSM. Hotspots found, colors from red to blue denote very low to very high interaction. (B) Distribution of binding energies for the different hotspots.*

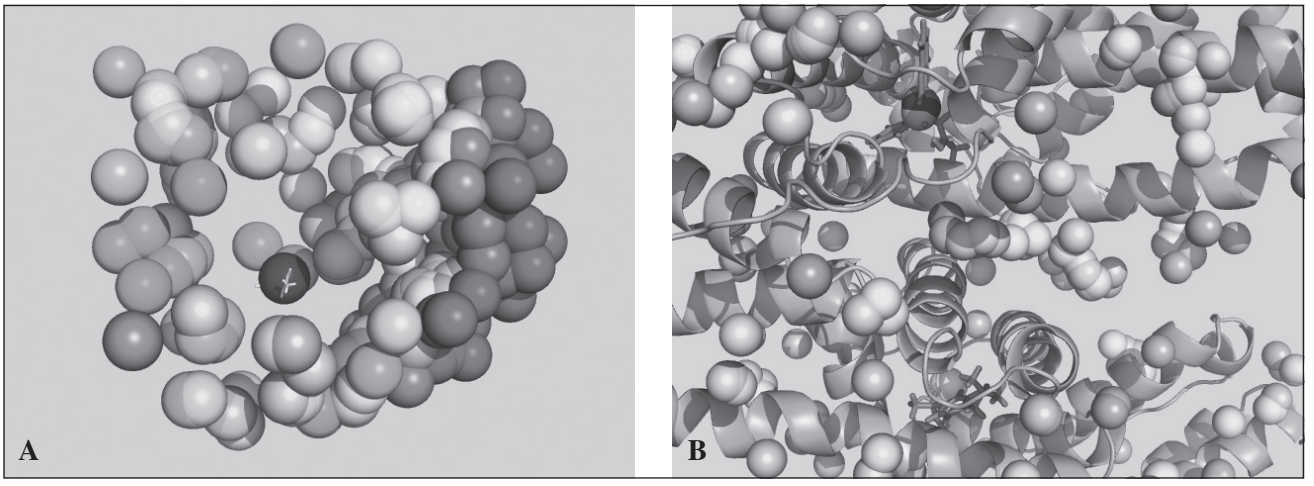

Figure 2 - *(A) PDB: 1QCF (B) PDB: 2BXD.*

2.B the blue sphere represents the strongest interaction spot found in 2BXD (red ligand) which is clearly differentiated from the hotspot obtained in 2BXG (pink ligand) and which does not yield a false positive result.
In view of these first promising results, we conclude that BINDSURF is an efficient and fast methodology for the determination on GPUs of protein binding sites depending on the ligand.
It can be used for fast pre-screening of a large ligand database, and its results can guide posterior detailed application of other VS methods.
Its application can help to improve drug discovery, design, repurposing and therefore help considerably in clinical research.
In the next steps we want to substitute the Monte Carlo minimization algorithm for more efficient alternatives.
We also want to improve the scoring function to treat hydrogen bonds, metals and aromatic interactions.
Finally we are now implementing full ligand and protein flexibility, taking advantage of the computational power of GPUs.

ACKNOWLEDGEMENTS
This research was supported by the Fundación Séneca (Agencia Regional de Ciencia y Tecnología, Región de Murcia) under grants 00001/CS/2007 and 15290/PI/2010, by the Spanish MEC and European Commission FEDER under grants CSD2006-00046 and TIN2009-14475-C04 and a postdoctoral contract from the University of Murcia (30th December 2010 resolution). We also thank Centro de Supercomputación de la Fundación Parque Científico de Murcia for providing supercomputing time.

REFERENCES

Brannigan G, LeBard DN, Hénin J, Eckenhoff RG, Klein ML. Multiple binding sites for the general anesthetic isoflurane identified in the nicotinic acetylcholine receptor transmembrane domain. PNAS. 2010; 107: 14122-14127.

Buch I, et al. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. PNAS. 2011; 108: 10184-10189.

Ghersi D, Sanchez R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. Proteins. 2009; 74: 417-424.

Guerrero GD, Pérez-Sánchez HE, Wenzel W, Cecilia JM, García JM. Effective parallelization of non-bonded interactions kernel for virtual screening on GPUs, Proc. 5th Int. Conf. on Practical Applications of Computational Biology & Bioinformatics (PACBB 2011).

Kokh DB, Wenzel W. Flexible side chain models improve enrichment rates in in silico screening. J Med Chem 2008; 51: 5919-5931.

NVIDIA. NVIDIA CUDA Programming Guide 4. 2011.

Ron O, et al. Pathway and mechanism of drug binding to G protein-coupled receptors. PNAS, 2011; 108: 13118-13123.

Pérez-Sánchez H, Wolfgang Wenzel. Optimization methods for virtual screening on novel computational architectures. Current computer-aided drug design. 2011; 7: 44-52.

Sánchez R, Sali A. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. PNAS. 95, 1998; 13597-13602.

Shan Y, et al. How does a drug molecule find its target binding site? J Am Chem Soc 110505112729011, 2011.