# Impact of implicit solvation models on database enrichment in GPU based blind Virtual Screening

Ginés D. Guerrero[1], José M. Cecilia[2], José M. García[1], Horacio Pérez-Sánchez[2*]

[1]*Computer Engineering Department, University of Murcia*
*30100 Murcia, Spain*
{gines.guerrero, jmgarcia}@ditec.um.es

[2]*Computer Science Department, Catholic University of Murcia (UCAM)*
*30107 Murcia, Spain*
jmcecilia@ucam.edu, horacio.ucam@gmail.com *(corresponding author)

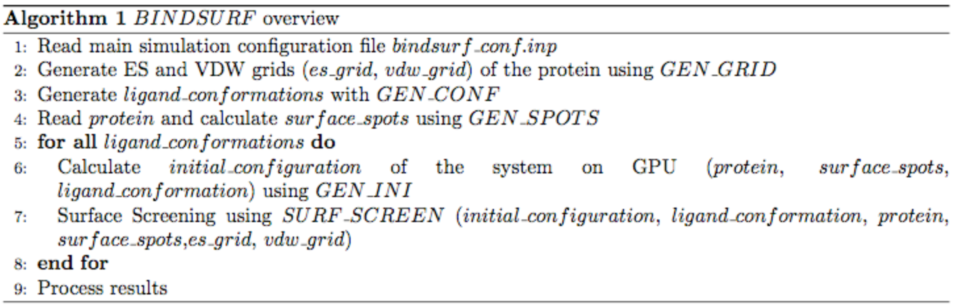
**Abstract.** Virtual Screening (VS) methods can considerably aid clinical research, predicting how ligands interact with drug targets. We present a novel VS methodology that uses implicit solvation models to scan the whole protein surface in order to find new hotspots depending on each ligand, and avoiding bias present in many current VS methods, since they assume same binding site for different ligands. Furthermore, our methodology is completely designed from scratch on last generation massively parallel GPU hardware, running up to 64 times faster than in a desktop computer and allowing fast processing of large ligand databases over the whole protein surface.

## 1 Introduction

Current Virtual Screening (VS) methods such as docking fail to make good toxicity and activity predictions since they do not take into account the effect of the solvent, and because 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 supposition is wrong [1]. We present a new VS methodology that avoids the previous assumption screening the whole protein surface and which incorporates an implicit solvation model, to provide new and useful information about targets and thus improving key toxicity and activity predictions. Furthermore, our methodology is completely designed for GPUs [8], allowing fast processing of large ligand and protein databases.

## 2 Methodology

$$S = \sum_{protein} \sum_{lig,flSC} \left( \frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{r_{ij}} \right) + \sum_{h-bonds} \cos\Theta_{ij} \left( \frac{\widetilde{R}_{ij}}{r_{ij}^{12}} - \frac{\widetilde{A}_{ij}}{r_{ij}^{6}} \right) + \sum_{SASA} \sigma_i A_i \quad (1)$$



**Algorithm 1** *BINDSURF* overview

```
Read main simulation configuration file $bindsurf\_conf.inp$
Generate ES and VDW grids ($es\_grid$, $vdw\_grid$) of the protein using $GEN\_GRID$
Generate $ligand\_conformations$ with $GEN\_CONF$
Read $protein$ and calculate $surface\_spots$ using $GEN\_SPOTS$
for all $ligand\_conformations$ do
    Calculate $initial\_configuration$ of the system on GPU ($protein$, $surface\_spots$, $ligand\_conformation$) using $GEN\_INI$
    Surface Screening using $SURF\_SCREEN$ ($initial\_configuration$, $ligand\_conformation$, $protein$, $surface\_spots$,$es\_grid$, $vdw\_grid$)
end for
Process results
```

**Fig. 1.** Pseudoce for the difrent stages involved in BINDSURF

We used the version 4.0 of the CUDA programming model [9] in our parallel implementation with a NVIDIA Tesla C2050 GPU. Rigid protein-ligand docking simulations were performed over the whole protein surface, divided into spheres of fixed volume, centered around the alpha carbons of each residue. The scoring function (see Eq. 1) uses highly GPU optimized non-bonded interaction kernels [5], implemented previously in the program BINDSURF [10], for the description of the electrostatic, Van der Waals and hydrogen bond interactions between the ligand and the protein. Solvent contribution is calculated using an implicit salvation model, which calculates solvent accessible surface area (SASA) [3], employing a highly optimized GPU kernel [2]. A Monte Carlo algorithm [6] optimized for GPU is used to minimize the total energy of the system, as depicted in Figure 1. 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, such as Molecular Dynamics, to screen only the ligands with the highest estimated affinities in the hotspots found by our methodology.

## 3. Results

| Atom number | ES | VDW | SASA | SOLVATION | HBOND |
|---|---|---|---|---|---|
| 1 | 15.9435 | -1.14305 | 0.0172082 | 0.00275332 | 0 |
| 2 | 17.4201 | -0.206061 | 0 | 0 | 0 |
| 3 | 20.4906 | -2.84008 | 0.154874 | 0.0247799 | 0 |
| 4 | 9.638 | -0.356836 | 0 | 0 | 0 |
| 5 | 42.9991 | -3.14011 | 0.0516247 | 0.00825996 | 0 |
| 6 | 11.473 | -0.336246 | 0 | 0 | 0 |
| 7 | 36.3463 | -1.39851 | 0 | 0 | 0 |
| 8 | 7.58693 | -0.0112397 | 0 | 0 | 0 |
| 9 | 37.2497 | -4.71891 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... |
| 32 | -214.723 | -0.387638 | 0.219912 | -0.0131947 | 0 |
| 33 | -296.406 | -5.3169 | 0.314159 | -0.0188496 | -0.0651586 |
| 34 | -248.987 | -0.334133 | 0.502655 | -0.0301593 | 0 |
| 35 | -235.443 | -0.908453 | 0.455531 | -0.0273319 | -0.407141 |
| 36 | -252.355 | -0.964031 | 0.455531 | -0.0273319 | -2.12521 |
|  | *-1756.04* | *-76.1119* | *5.50158* | *-0.280879* | *-36.5779* |

**Table 1.** Detailed information for the individual terms of the scoring function for each atom obtained for the top pose in antithrombin-TMI docking [7]. Columns denote atom number of the ligand, electrostatic interactions term, Van der Waals interaction term, SASA value, salvation term and hydrogen bond term. Final row shows total summation taking into account all ligand atoms.

We initially performed redocking simulations for different PDB structures (such as 2BSM, 1QCF) and checked that in most of the tested cases our implementation finds efficiently the crystallographic binding mode with a RMSD value less than two Angstroms. In cases with proteins whose binding site depends on the ligand our method also works efficiently, such as for crystal structures 2BXG and 2BXD, in good agreement with previous results obtained by other authors [4]. We could also reproduce docking results obtained previously by other docking methodologies [7], for more detailed information see Table 1.

## 4. CONCLUSION

We have developed an efficient methodology based on BINDSURF that uses an implicit solvation model 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.

## ACKNOWLEDGEMENT

This research was supported by the Fundación Séneca (Agencia Regional de Ciencia y Tecnología, Región de Murcia) under grant 15290/PI/2010, by the Spanish MEC and European Commission FEDER under grants CSD2006-00046 and TIN2009-14475-C04.

## References

1. Brannigan G. et al, 2010. Multiple binding sites for the general anesthetic isoflurane identified in the nicotinic acetylcholine receptor transmembrane domain. *PNAS*, Vol. 107, pp 14122–14127.
2. Cepas-Quiñonero E et al, 2011. MURCIA: Fast parallel solvent accessible surface area calculation on GPUs and application to drug discovery and molecular visualization. *In: NETTAB 2011 workshop focused on Clinical Bioinformatics*, pp. 52-55.
3. Eisenberg D. et al, 1986. Solvation energy in protein folding and binding. *Nature*, Vol. 319, pp. 199–203.

4. Ghuman J. et al, 2005. Structural basis of the drug-binding specificity of human serum albumin. *Journal of Molecular Biology*, Vol. 353, pp. 38–52.

5. Guerrero G.D. et al, 2012. Parallelization of Virtual Screening in Drug Discovery on Massively Parallel Architectures. *In: 20th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing PDP 2012*, pp. 588-595.

6. Metropolis N. et al, 1953. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, Vol. 21, pp. 1087–1092.

7. Navarro-Fernández J. et al, 2012. In Silico Discovery of a Compound with Nanomolar Affinity to Antithrombin Causing Partial Activation and Increased Heparin Affinity. Journal of Medicinal Chemistry, (in press).

8. Pérez-Sánchez H. et al, 2011. *Optimization methods for virtual screening on novel computational architectures*. Current Computer Aided Drug Design, Vol. 7, pp. 44-52.

9. NVIDIA, 2011. NVIDIA CUDA Programming Guide 4.

10. Sánchez-Linares I. et al, 2012. High-throuput parallel blind Virtual Screening using *BINDSURF*. In: BMC Bioinformatics, Vol 13 (Suppl 14):S13.