

Structural bioinformatics

ProtPOS: a python package for the prediction of protein preferred orientation on a surface

Jimmy C. F. Ngai¹, Pui-In Mak² and Shirley W. I. Siu^{1,*}

¹Department of Computer and Information Science and ²State-Key Laboratory of Analog and Mixed-Signal VLSI, Faculty of Science and Technology, University of Macau (UM), Avenida da Universidade, Taipa, Macau, China

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on November 16, 2015; revised on March 10, 2016; accepted on March 3, 2016

Abstract

Summary: Atomistic molecular dynamics simulation is a promising technique to investigate the energetics and dynamics in the protein–surface adsorption process which is of high relevance to modern biotechnological applications. To increase the chance of success in simulating the adsorption process, favorable orientations of the protein at the surface must be determined. Here, we present ProtPOS which is a lightweight and easy-to-use python package that can predict low-energy protein orientations on a surface of interest. It combines a fast conformational sampling algorithm with the energy calculation of GROMACS. The advantage of ProtPOS is it allows users to select any force fields suitable for the system at hand and provide structural output readily available for further simulation studies.

Availability and Implementation: ProtPOS is freely available for academic and non-profit uses at <http://cbbio.cis.umac.mo/software/protpos>

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

Contact: shirleysiu@umac.mo

1 Introduction

Understanding the mechanisms of protein adsorption on surfaces is important for many biotechnological applications, such as the design of medical implants, protein biochips, drug delivery systems and point-of-care devices, etc. By virtue of molecular modeling and simulations, the protein adsorption processes, its dynamics and subsequent protein–surface interactions can be studied at the molecular or atomic level of detail. However, a successful simulation of the adsorption process depends largely on the initial protein orientation relative to the surface (Wei *et al.*, 2012). To avoid sampling protein trajectory which will eventually fail to adsorb, a work-around is to firstly determine the preferred orientations of the protein and use them as starting structures in molecular simulations.

Essentially, a search for the low-energy protein orientation on a given surface is an optimization process which includes a conformation generation strategy and a scoring function. For the former, a systematic rotation (and translation) of the protein or a Monte Carlo simulation is commonly used (Hsu *et al.*, 2008; Xie *et al.*, 2010; Zheng *et al.*, 2004). For this structure determination problem, a simplified scoring function

accounting for only protein–surface interactions is usually preferred for higher computational efficiency. Interaction parameters can be taken from an empirical force field or modification from the scoring function of a docking method (Sun *et al.*, 2005). Interestingly, unlike other structure prediction problems such as protein–ligand docking or protein–protein docking where numerous tools exist to simplify the prediction task, there are only limited choices available for fast protein–surface structure prediction. One such option is RosettaSurface (Makrodimitris *et al.*, 2007) which is included in the Rosetta molecular modeling package, and hence the use of RosettaSurface entails a deep learning curve. Other prediction methods mentioned above, to our knowledge, are unavailable publicly for general use.

Hence, to fill this gap, we have developed a self-contained, lightweight and easy-to-use software package called prediction of PROtein Preferred Orientation on a Surface (ProtPOS) based on the protein–surface docking method described in (Ngai *et al.*, 2015). It combines particle swarm optimization (PSO) algorithm (Kennedy and Eberhart, 1995) with energy minimization to perform global and local search over six degrees of freedom of the protein translation and rotation on

a given surface. The advantage of ProtPOS is that it allows users to select any force fields suitable for the system at hand, to be provided as inputs in the format of GROMACS topology together with the protein and surface coordinates. In this way, predicted low energy protein conformations are compatible to the selected force field and readily available for further molecular dynamics (MD) simulations.

2 Implementation

ProtPOS is a package of open-source shell scripts and Python programs utilizing the PyMOL library for conformation generation and the popular MD simulation software GROMACS for energy minimization and scoring. The workflow of ProtPOS is explained in Fig. 1. To facilitate ease-of-use, a template wrapper script is provided through which users can customize parameters for the search algorithm, perform prediction and clustering analysis. During the search, all rotational angles around the protein center of mass are allowed. For surfaces with repeating surficial patterns, lateral translation (X and Y directions) of the protein can be restrained to defined ranges, typically the dimensions of the surface unit cell; whereas the translation along the surface normal (Z) is set to be [1.0, 5.5] Å from the surface, i.e. the interfacial region where protein-surface interactions are significant. If contacting residues are known a priori, they can be given to ProtPOS to limit further the

search on the specified protein surface. It should be mentioned that, for better computational efficiency, the Generalized Born implicit solvent model (GBSA) is used during energy minimization and only nonbonded interactions between the protein and the surface are included in scoring.

3 Validation and performance

ProtPOS was tested by predicting the initial adsorption orientations of lysozyme on a hydrophobic surface of perfluorodecane molecules. The results were compared to experiments and previous computational studies qualitatively and found good agreement (See [Supplementary Table S1](#)). Furthermore, the selected low-energy structures were subjected to 50 ns MD simulations in order to assess the stability of the protein orientations. In three out of four cases, the proteins maintained the predicted orientations and stably attached to the surface ([Supplementary Table S2](#)) whereas in one case the protein was rotated slightly keeping three contacting residues same as in the initial orientation and remain stably attached throughout the remaining simulation time. MD simulations of randomly selected protein orientations either resulting in the protein converted to predicted orientation or diffused away from the surface, suggesting that ProtPOS predicted structures serve as good starting structures for further simulation studies.

Regarding the time performance, our test case running on a consumer-grade desktop computer (with Intel i7-4790 3.6 GHz CPU and 16 GB memory) took about 1 day to complete one ProtPOS run, with an average of 81.5 iterations to reach convergence in the PSO search. The most expensive calculation is the energy minimization step which is done for all PSO particles in each iteration. Future improvement on ProtPOS should consider pre-filtering of highly unfavorable orientations before performing energy minimization, as well as parallelizing the PSO algorithm by multicore CPU and advanced GPU techniques.

Funding

This project was supported by the research grant of University of Macau (grant number MYRG2014-00104-FST). The authors would like to thank ICTO for their support of computing facilities.

Conflict of Interest: none declared.

References

- Hsu, H.J. et al. (2008) Preferred orientation of albumin adsorption on a hydrophilic surface from molecular simulation. *Colloids Surf. B Biointerf.*, **67**, 183–191.
- Kennedy, J. and Eberhart, R. (1995) Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, pp. 1942–1948.
- Makrodimitris, K. et al. (2007) Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite. *J. Am. Chem. Soc.*, **129**, 13713–13722.
- Ngai, J.C.F. et al. (2015) Predicting favorable protein docking poses on a solid surface by particle swarm optimization. In *Proceedings of CEC 2015*, pp. 2745–2752.
- Sun, Y. et al. (2005) Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation. *Langmuir*, **21**, 5616–5626.
- Wei, T. et al. (2012) Molecular dynamics simulation of lysozyme adsorption/desorption on hydrophobic surfaces. *J. Phys. Chem. B*, **116**, 10189–10194.
- Xie, Y. et al. (2010) Parallel tempering monte carlo simulations of lysozyme orientation on charged surfaces. *J. Chem. Phys.*, **132**, 065101.
- Zheng, J. et al. (2004) Molecular simulation study of water interactions with oligo (ethylene glycol)-terminated alkanethiol self-assembled monolayers. *Langmuir*, **20**, 8931–8938.

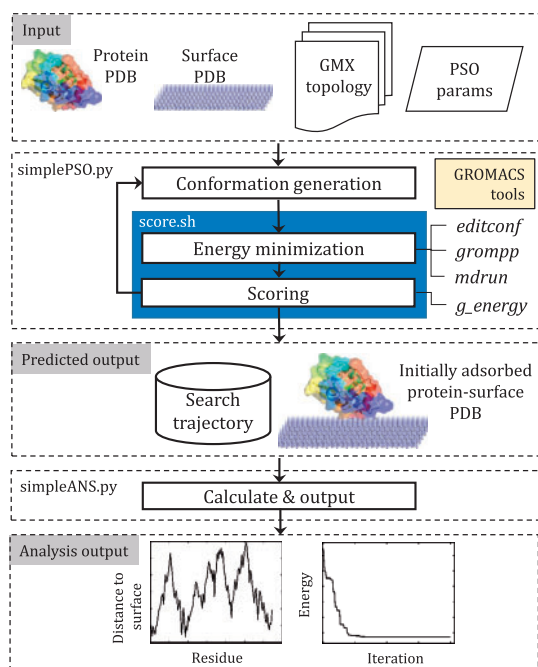


Fig. 1. The workflow of ProtPOS: First, the user has to prepare the initial protein and surface coordinates in two separate PDB files. The molecular topologies in GROMACS format can be generated directly using *pdb2gmh* and the automated topology building software such as ATB web server. The Python program *simplePSO.py* generates a population of different protein positions and orientations with respect to the surface; each of them will undergo energy minimization and scoring until better protein orientation cannot be found within defined number of iterations or the maximum number of iterations is reached. The lowest energy structure (*gbest.pdb*) is then reported together with the search trajectory. The latter can be fed to *simpleANS.py* to produce the protein-surface minimum distance profile and the plot of energy evolution. Therefore, each complete run reports one lowest energy structure. Due to the stochastic nature of the PSO algorithm, the user should repeat ProtPOS 10–15 times (see [Supplementary information](#)) and perform clustering analysis to identify unique low-energy protein orientations (Color version of this figure is available at *Bioinformatics* online.)