Predicting favorable protein docking poses on a solid surface by particle swarm optimization

Jimmy C. F. Ngai*, Pui-In Mak†, Shirley W. I. Siu*
*Computational Biology and Bioinformatics Lab, Faculty of Science and Technology, University of Macau (UM), Avenida da Universidade, Taipa, Macau, China
† State-Key Laboratory of Analog and Mixed-Signal VLSI, and Faculty of Science and Technology, University of Macau (UM), Avenida da Universidade, Taipa, Macau, China
jimmycfngai@gmail.com, pimak@umac.mo, shirleysiu@umac.mo

***Abstract*—Protein adsorption at solid surfaces has received intense focus due to its high relevance to biotechnological applications. In alternative to experimental approaches, computational methods such as molecular dynamics (MD) simulations are frequently employed to simulate the protein adsorption process and to study molecular interactions at the interfacial region. However, a successful simulation of the adsorption process depends largely on the initial adsorbed protein orientation on the surface. To avoid sampling protein trajectory which will eventually fail to adsorb, a workaround is to first determine the preferred orientations of the protein relative to the surface and use them as starting structures in MD simulations. Here, we present the first application of particle swarm optimization (PSO) to search for the low energy docking poses of a protein molecule on a solid surface. Performing rigid-body translation and rotation of the protein with energy minimization and empirical scoring function, our search algorithm successfully located the low energy orientations of the lysozyme molecule on a hydrophobic PTFE surface. Nine out of ten predicted docking poses are energetically more favorable than all poses sampled using a brute-force search. Three sets of major adsorption sites are identified for the lysozyme and they are in good agreement to results obtained by long MD simulations; novel adsorption sites are also identified from the lowest energy docking pose. Our method provides a reliable way to predict the optimal protein orientations useful for computational studies of protein-surface interactions.**

***Keywords*—protein adsorption, particle swarm optimization, hydrophobic solid surface, PTFE, lysozyme**

I. Introduction

Protein adsorption to surfaces is a fundamental biological phenomena which is of high relevance to modern biotechnological applications such as the design of biochips, pharmaceutical drug products, and biomaterials [1], [2]. Common for these applications is the need to understand the protein adsorption process, how proteins adsorb, whether the adsorption process alters proteins' functions, or alters the efficacy of bioproducts. Such valuable knowledge provide guidance to the ultimate goal of rationally designing the proteins or the interfacial material.

In addition to physical experiments, computational methods such as molecular dynamics (MD) simulations are widely used to study protein-surface adsorption due to the high-level of detail that the methods can provide. In MD, the dynamics of the protein is modeled explicity following the Newton's law of motion, hence, conformational changes of the protein upon adsorption on surface can be closely followed. Some atomistic and coarse-grained MD studies have been published recently including different types and numbers of proteins on physically, chemically, or structurally different surfaces [3]–[7].

978-1-4799-7492-4/15/$31.00 ©2015 IEEE

Nevertheless, using MD to observe slow biological processes which occur on the microseconds-and-above timescale is difficult. As pointed out by Wei and co-workers [3], the simulation time required in the protein adsorption studies tend to be rather long because the surface-induced dehydration process of the approaching protein is very slow, taking around 70 ns for a small protein like lysozyme. Not only that, the protein needs to be in the proper spatial orientation towards the surface in order to overcome the barrier of electrostatic repulsion [8]; otherwise, the protein is likely to diffuse away from the surface.

To successfully simulate the protein adsorption process and to cut down the computing time on the lengthly (and usually uninteresting) rotational diffusion of the protein near the surface, a reliable way is to first predict the preferred orientations of the initial adsorbed protein, and then use these as the starting structures for subsequent MD simulations. A pioneering work was conducted by Zheng et al. [9] who applied Metropolis Monte Carlo (MC) simulations to identify the optimal orientations of the lysozyme protein on different SAM surfaces. Only the non-bonded vdW and coulombic interactions between the protein and the surface were considered in their simulation protocol. Another prediction method was proposed by Makrodimitris et al. [10] which is called RosettaSurface. This method combined the rigid-protein MC with a minimization routine to optimize the interfacial side-chains of the protein. The energy function is a linear combination of five interaction terms including Lennard-Jones, solvation, hydrogen bonding, and electrostatics. Instead of using a random search, Hsu et al. [11] proposed to systematically rotate the molecule stepwise and translate it perpendicularly toward the surface to generate a complete contour map of protein-surface interactions. This guided search was shown to perform faster in locating the energy minimum than standard MC search by approximately 10%. Undoubtedly, more efficient yet reliable methods would always be desirable especially for studies involving large proteins and many different surface

models. Recently, Wei et al. [12] presented a use of the genetic algorithm (GA) as the global optimization method to solve the protein-surface prediction problem. GA is a popular evolutionary algorithm that mimics the process of natural selection. Using a genetic representation of the search (solution) space called chromosome, GA applies genetic operators such as crossover and mutation on selected parent solutions to generate child solutions for the new generation. Their proposed hybrid multi-loop GA method showed good agreement with a brute-force search and it has been sucessfully applied in later MD studies to predict the protein initial adsorbed orientations [3], [13].

The particle swarm optimization (PSO) is yet another popular search algorithm for solving complex optimization problem. Also inspired by nature, the idea of the PSO is to simulate the social behavior of bird flocking in looking for food by iteratively updating the bird's position based on the knowledge of the swarm and own's flying experience. It is well-known for its simplicity and flexibility but the applicability of it in protein-surface prediction problem has not been investigated.

Here, we present the first application of PSO algorithm on the protein-orientation prediction problem. The paper is organized as follows. In Section II, we present the computational methods including the protein-surface model, the PSO algorithm for protein-surface docking, the energy function, and the software used in the method's implementation. Independent docking runs to predict the docked pose of the lysozyme protein on the hydrophobic PTFE surface were conducted and the results are presented in Section III. Here, the PSO predictions are compared to the brute-force search in terms of protein-surface interaction energy and the predicted protein-surface contacts are compared with published computational studies. Finally, we conclude the paper in Section IV.

II. Computational Methods

Similar to other structure prediction methods, our proposed method for protein docking *pose* prediction on a solid surface consists of two main parts: sampling and scoring. Here, PSO is used to generate a protein pose by exploring the positional and orientational space of the protein on a surface, then the fitness of each pose will be assessed using a force field-based scoring function accounting for the non-bonded interactions between the protein and the surface. To evaluate the performance of our proposed method, we apply it to predict the lysozyme adsorption on a hydrophobic surface and compare the results to a brute-force search.

A. The protein-surface model

The system is a full-atomistic model consisting of a hydrophobic solid surface and a protein molecule. The hydrophobic surface is a layer of perfluorodecane molecules (a short-chained polytetrafluoroethylene, PTFE) in a hexagonal packing arrangement with a total surface area of 225 nm^2. The choice of this surface molecule is based on the fact that fluorocarbon is the main composition of Teflon, the surface coating material popularly used on modern lab-on-a-chip devices. Hence, computational studies of protein adsorption on fluorocarbon surfaces would be interesting to the chip manufacturing industry and surface design research. The surface model used in our case study was tested previously in a MD simulation study of electrowetting of water droplets under the influence of external electric fields [14], and the surface was shown to yield electrowetting property similar to experimental observations [15].

In this case study, the lysozyme molecule is selected as the protein model. Lysozyme is abundant in tears, saliva, human milk, mucus and other secretions of animals and thus it is relevant to many biochemical experiments utilizing lab-on-a-chip systems. It is noteworthy that lysozyme is classified as a *hard* protein meaning that its overall structure cannot be changed easily. Our model structure is the 129-residue hen egg-white lysozyme molecule obtained from the PDB database (PDB code 1AKI, resolution 1.5 Å) [16]. The structure is preprocessed using the topology generation tool *pdb2gmx* from GROMACS [17] with default setting for residue protonation states and NH^{3+}/COO^{-} terminal capping. The total charge of the protein is +8.0 e.

B. Pose sampling using PSO algorithm

The protein-surface docking problem is an optimization problem where the goal is to look for the favorable position and orientation of the protein molecule near a given surface structure. Taking the protein as a rigid body, then a complete search includes three translational and three rotational degrees of freedom of the protein. In a homogeneous surface where the surface molecule has a certain packing arrangement, the positional search in parallel to the surface can be limited to the lengths of the unit cell. Even so, a rough brute-force search would involve the sampling of $\approx 10^{10}$ protein poses which is inefficient and prohibitively expensive. Alternatively, by applying metaheuristics algorithm, generation of each sample becomes an intelligent procedure based on the "knowledge "of the previous samples. Among the metaheuristics algorithms, particle swarm optimization is a popular technique which has not been exploited, to the best of our knowledge, in the protein-surface docking problem.

Particle Swarm Optimization (PSO) is a population-based stochastic algorithm simulating the food-searching behavior of a flock of birds [18], [19]. It is an extremely simple algorithm that seems to be effective for optimizing a wide range of functions. It requires only basic mathematical operators and is computationally inexpensive in terms of both memory and time complexity.

In PSO, a group of birds is called *swam* and each bird in the swarm, also named *particle*, encodes a feasible solution of a function to be optimized. During the search, each particle adjusts its moves according to its own best solution and the swarm's best solution with some randomness. So the optimal solution for the function is found by cooperation and competition among the swam themselves through iteratively updating their moves. As a consequence of the knowledge sharing in swarm, when one of the particles finds a promising region in the search space, the rest of the swarm will follow quickly. The high speed of convergence and the relative simplicity of the algorithm make PSO an attractive method for our protein-surface docking problem.

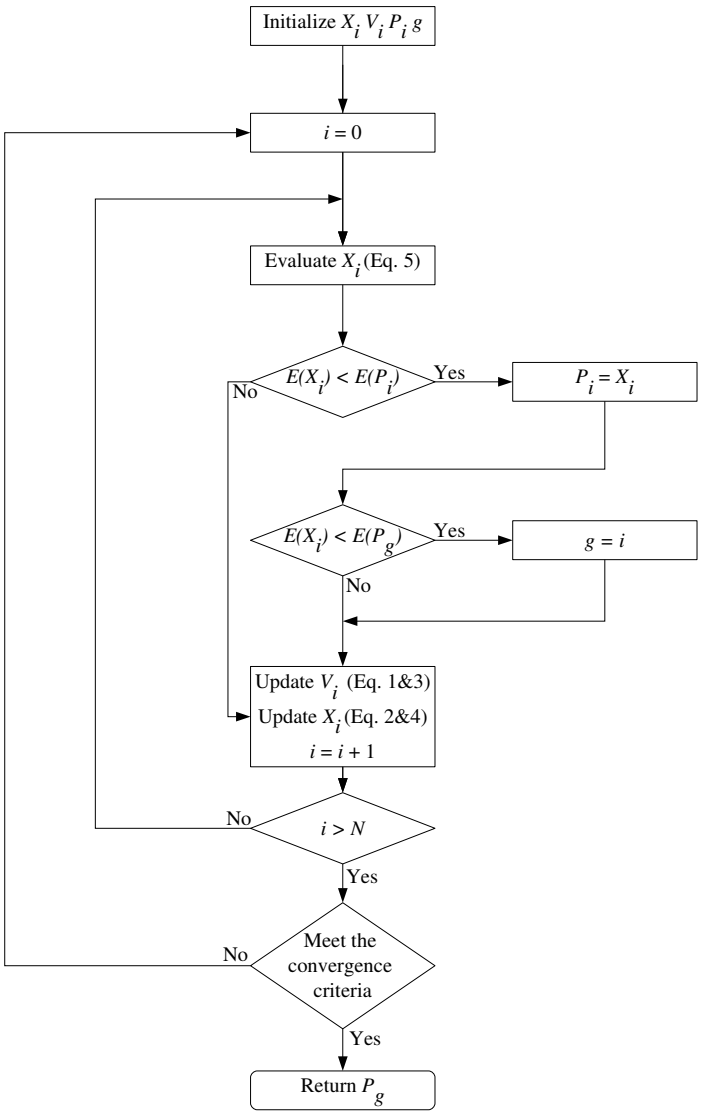


Fig. 1. Flowchart of the PSO algorithm

In a multidimensional search space, the position of a particle can be encoded in a D-dimensional vector where the *i*-th particle is denoted as $X_i = (x_{i1}, x_{i2}, ..., x_{iD})$. Similarly, the velocity of the particle, represents the speed and the direction of the particle's movement, can be denoted as $V_i = (v_{i1}, v_{i2}, ..., v_{iD})$. During the move, the best historical position of the particle, or the *personal best*, is recorded in $P_i = (p_{i1}, p_{i2}, ..., p_{iD})$ where the best historical position among all particles, or the *global best*, in the swarm is recorded in P_g.

At the start, all particles in the swarm of size N are initialized with random positions and random velocities. In each iteration, the swarm are manipulated according to the *velocity update* equation (Eq. 1) and the *position update* equation (Eq. 2) as follows:

$$V_i(t+1) = w \cdot V_i(t) + c_1 \cdot rand() \cdot (P_i(t) - X_i(t)) + c_2 \cdot rand() \cdot (P_g(t) - X_i(t)) \quad (1)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (2)$$

where w denotes the inertia weight, c_1 the cognitive weight, and c_2 the social weight. All three parameters are positive constants and are usually defined empirically. The parameter w plays the role of balancing the global search and local search. If w is too small, all particles will fly quickly to the historical best and lose the chance to perform global search. On the other hand, if w is too large, the swarm knowledge will have minimal effect on the search direction of the particles which leads to slow convergence. Finally, *rand()* is a random function generating random numbers in the range of [0, 1] on every entry.

The flowchart of the PSO algorithm is shown in Figure 1. In each iteration the fitness of each particle is calculated using a scoring function of the particle's position. Then the *personal best* and the *global best* are updated if a better score is obtained to reflect the improvement in the swarm knowledge. Afterwards, the search direction and the search speed are updated according to the equations of velocity-update and position-update. If the particle's velocity is over the predefined limit V^{max}, it is clamped to the maximum value, see Eq. 3. Similarly, if the particle's position is beyond the positional limits, it is re-positioned to the valid range $[a, b)$ with a modulo function (see Eq. 4). The swarm is considered converged until the number of runs exceeds or there is no update of P_g in a number of iterations. The PSO parameters used in our prediction method are listed in Table I. Clerc [20] provided a reference setting of $w = 0.721$ and c_1 = c_2 = 1.193.

The positional limits for the protein lateral translation (along X- and Y-axis) are surface-dependent. In our protein-surface model, the lateral distance between any pairs of surface molecules is about 5.8 Å, hence a positional limit of ± 3 Å relative to the protein geometric center is applied. For the vertical translation (along Z-axis), the translation limit is defined as the minimum distance between the protein and the surface. Our tests showed that a limit of 1.0 Å and 5.5 Å is reasonable. For the maximum velocities, we used 10% of the range of the positional limit as the maximum velocity of the respective dimension. In our preliminary tests, these settings allow the swarm to explore sufficiently while converge rather quickly. The positional limits and the maximum velocities are listed in Table II.

$$\forall j \in [1, D], \quad V_{ij} = \frac{V_{ij}}{|V_{ij}|} \times V_{max,j} \quad \text{if} \quad |V_{ij}| < V_{max,j} \quad (3)$$

$$\forall j \in [1, D], \quad X_{ij} = \Big((X_{ij} - a_{ij}) \mod (b_{ij} - a_{ij}) \Big) + a_{ij} \quad (4)$$

C. The energy function and energy minimization

Goodness of the PSO-generated protein pose on the surface is assessed by a force-field based energy function. Two non-bonded interactions are considered (see Eq. 5): the van der Waals interaction (vdW) describing the attractive and repulsive forces between molecules due to induced dipoles, and the electrostatic interaction describing the electric forces between two charged objects. The former can be modeled using the

TABLE I. PSO PARAMETER SETTINGS FOR PROTEIN-SURFACE DOCKING

Parameter	Description	Value
N	Number of particles	200
w	Inertia weight	0.721
c_1	Cognitive weight	1.193
c_2	Social weight	1.193
C	convergence criteria	10 steps

TABLE II. POSITIONAL LIMITS AND MAXIMUM VELOCITIES FOR DOCKING THE LYSOZYME ON THE PTFE SURFACE

Parameter	Description	Value
$X^{max,r}$	Positional limit for rotation about X, Y, Z	$[0°, 360°)$
X^{max,t_x}	Positional limit for translation along X	$[-3.0, +3.0)$ Å
X^{max,t_y}	Positional limit for translation along Y	$[-3.0, +3.0)$ Å
X^{max,t_z}	Positional limit for translation along Z	$[1.0, 5.5)$ Å
$V^{max,r}$	Maximum velocity of rotation about X, Y, Z	[36, 36, 36]°
$V^{max,t}$	Maximum velocity of translation along X, Y, Z	[0.6, 0.6, 0.4] Å

TABLE III. THE OPLS-AA FORCE FIELD PARAMETERS FOR THE PTFE MOLECULE.

	q (e)	σ (nm)	ϵ (kJ mol^{-1})	OPLS-AA atom type
C of CF_3	0.360	0.35	0.276144	opls_961
C of CF_2	0.240	0.35	0.276144	opls_962
F	-0.120	0.295	0.221752	opls_965

Lennard-Jones potential as shown in Eq. 6–8 and the latter using the Coulomb's law as shown in Eq. 9. The total protein-surface interaction is obtained by summing the vdW and the coulombic interactions between neighboring protein-surface atom pairs where the neighborhood is determined by a distance threshold.

$$E = E_{vdW} + E_{Coulomb} \tag{5}$$

$$E_{vdW} = \sum_{i=1}^{N_p}\sum_{j=1}^{N_s} 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{6}$$

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \tag{7}$$

$$\epsilon_{ij} = \frac{1}{2}(\epsilon_{ii}\epsilon_{jj})^{1/2} \tag{8}$$

$$E_{Coulomb} = \sum_{i=1}^{N_p}\sum_{j=1}^{N_s} \frac{138.9(q_i q_j)}{r_{ij}} \tag{9}$$

where r_{ij} is the Euclidean distance between protein atom i and surface atom j. Atom-type parameters ϵ and σ of the vdW term and q of the Coulomb term are taken from the OPLS-AA force field [21].

Each PSO-generated protein pose is subjected to short period of energy minimization using the steepest descent algorithm with the following settings: $emstep$ = 100 nm (step size), $nsteps$ = 1000 (maximum number of steps), $emtol$ = 100.0 kJ mol^{-1}nm^{-1} (convergence force). Cut-off of 1.0 nm is used for both coulomb and vdW interactions. For the purpose of fast energy calculation, water is modeled implicitly using the GBSA solvation model applying Onufriev-Bashford-Case (OBC) algorithm for Born radii calculation [22]. The dielectric coefficient of the implicit solvent is set to 78.3. The non-polar part of the Born energy is calculated using the ACE type approximation with surface tension set to 0.0054 kcal/mol/Å^2; the cut-off for the calculation of the Born radii is the same as for non-bonded interactions, which is 1.0 nm. Since GBSA parameters are not presented in GROMACS, they are derived from parameters of the closest atom types and the values are presented in Table IV.

TABLE IV. GBSA PARAMETERS FOR THE PTFE MOLECULE

	vdW radius for Born radius calculation (nm)	Scale factor for OBC
C of CF_3	0.1900	0.72
C of CF_2	0.1900	0.72
F	0.1475	0.88

D. Program implementation

The prediction program was implemented using the Python language utilizing the structure manipulation functions from the PyMOL molecular visualization package [23]. The energy minimization and calculation were done using GROMACS version 4.5.5 [17]. GROMACS is an efficient molecular simulation package popularly used for molecular dynamics simulations and energy minimization of structures. A wide-range of atomic force fields are supported such as OPLS, GROMOS, CHARMM and AMBER, etc. This allows us to extend our program to suit different molecular models by applying different force fields. We completed our testing on a single PC which utilize an Intel®i7-4790 processor, 16GB of system memory, and RAID-0 hard disk array. Ubuntu 14.10 was running on the hardware with the Linux 3.16 kernel, Gromacs 4.5.5, PyMOL 1.7.1.3, and Python 2.7.8.

III. RESULTS

Results from 10 independent runs of our proposed PSO-based docking algorithm using the lysozyme-PTFE model are listed in Table V. They are sorted by the energy score and are given an ID from pso1 to pso10 with the final global best position and the energy of that configuration. The position P_g is a vector of three rotational angles about X-, Y-, Z-axis plus two lateral distances from the geometric center of the initial protein position and the protein minimum distance to the surface along Z. The energies of the predicted docking poses range from -1060 kJ mol^{-1} to -811 kJ mol^{-1}, which gives an average of ≈-940 kJ mol^{-1}. On average, a PSO docking run converges in 31.4 iterations and requires 62,800 energy calculations. All runs were conducted using the same set of search parameters in Tables I and II.

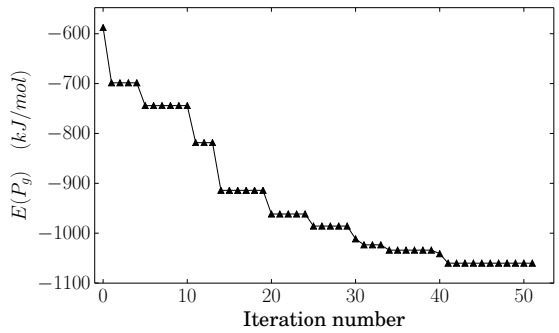
To observe the behavior of the PSO algorithm in finding the optimal protein docked pose, the energy score and the position vector of the global best P_g are plotted against the iteration number. As shown in Figure 2a, the energy of the P_g in the docking run pso1 is improved very fast in the first half of the search process and it slows down considerably at the second half. This drastic change in energy is caused by updates of large movement of the protein molecule relative to

TABLE V. RESULTS FROM 10 INDEPENDENT RUNS OF OUR PROPOSED PSO-BASED DOCKING METHOD USING THE LYSOZYME-PTFE SYSTEM. THE RUNS ARE SORTED BY THE ENERGY SCORE OF THE PREDICTED DOCKING POSE AND LABELED WITH PSO1–10.

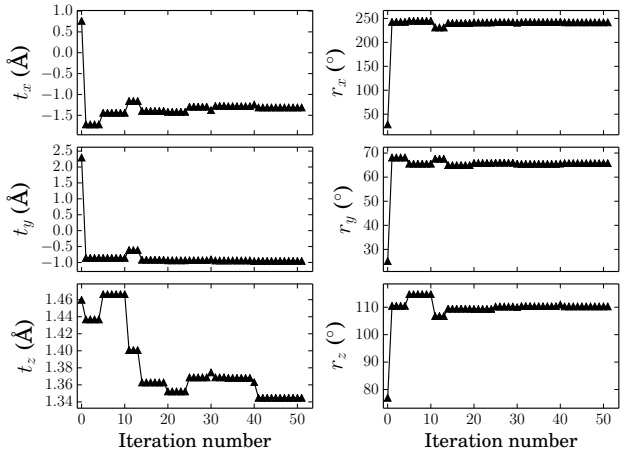
ID	Set[§]	$P_g(r_x, r_y, r_z, t_x, t_y, t_z)$	$E(kJ\ mol^{-1})$	Number of iterations	Total number of $E(X_i)$ evaluations[♯]
pso1	1	241.85, 65.78, 110.28, -1.31, -0.95, 1.34	-1060.16	52	10,400
pso2	2	229.45, 171.49, 257.38, 0.86, 1.66, 1.58	-1013.04	44	8,800
pso3	2	226.79, 176.05, 102.38, -0.10, -0.31, 1.64	-1007.43	35	7,000
pso4	3	199.79, 188.18, 180.08, -0.29, 0.94, 1.62	-959.92	27	5,400
pso5	1	300.51, 64.09, 303.63, -1.42, 1.86, 1.78	-954.11	58	11,600
pso6	2	221.43, 177.86, 190.29, 0.34, -0.42, 1.47	-918.48	28	5,600
pso7	2	223.86, 171.40, 230.22, -1.13, -1.90, 1.17	-907.24	16	3,200
pso8	1	219.96, 48.99, 130.83, -1.12, 0.87, 2.04	-890.88	26	5,200
pso9	1	248.40, 67.35, 113.82, -0.78, -1.26, 1.55	-878.61	18	3,600
pso10	-	252.17, 211.52, 319.75, 2.08, 0.47, 1.47	-811.26	10	2,000
		Average	-940.11±73.94	31.4±15.77	6,280±3,154.12

[§]Poses are clustered into sets of best docking poses based on the residue's minimum distance to surface profiles. Due to the worst energy and dissimilarity to any existing clusters, pso10 is not classified.

[♯]The total number of $E(X_i)$ evaluations is the number of iterations multiplied by the number of birds N in the PSO search.



(a)



(b)

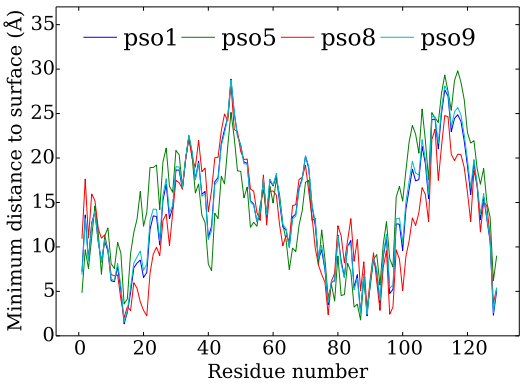
Fig. 2. Evolution of (a) the energy score and (b) the protein position and orientation of P_g from the docking run pso1.

the surface. Interestingly, large lateral (t_x, t_y) and rotational moves of the protein occurred only in the first two iterations of the PSO search (see Figure 2b), meaning that a potentially surface-contacting "protein surface" can be determined rather quickly as long as its vertical distance to the surface is reasonably close. The remaining search process fine-tunes the vertical separation (t_z) between the protein and the surface and performs some small orientational adjustments of the protein.

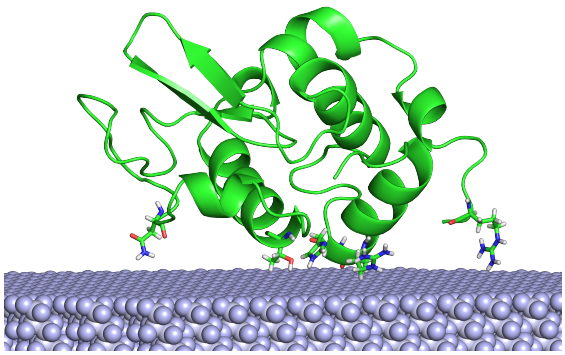
The current setting considers the search as converged if no better position than P_g is found in 10 iterations. Our tests of more iterations to convergence (e.g. 30 and 50) did not show improvements. Likewise, different number of particles N were also tested. A large number of particles makes each iteration more computationally expensive due to increased number of fitness evaluations, which is the most time consuming process in the PSO loop. However, a sufficiently large number will allow the exploration to be done more thoroughly and preventing the swarm to be trapped at local minima in high energy regions. Based on our observations, the value of $N = 200$ turns out to be an optimal value for this case study.

Due to limited computational power, it is not possible to scan the complete configurational space of the protein on the surface, in which the sampling size is in the order of at least 10^{10}. However, to get an idea of how good the PSO-found docking poses are, we run a limited brute-force search on the same protein-surface system starting from the configuration of pso1. By systematically rotating the protein about the X- and Y-axis with a step size of 6°, and about the Z-axis in -30°, 0°, and 30° while keeping the same lateral and vertical position of the protein, a lower energy configuration than pso1 cannot be found. The second lowest energy in this systematic search (the lowest one is the starting configuration, which is pso1) turns out to be -855.15 kJ mol^{-1} with the configuration of [139.86, 71.79, 110.20, -1.31, -0.95, 1.34]. This pose is energetically less favorable than 9 of the 10 PSO-found poses and almost 200 kJ mol^{-1} worser in energy value than pso1.

Previous computational studies of protein-surface adsorption revealed that protein could be adsorbed in different orientations facilitated by different protein residues [9] [24] [3]. In order to identify these important protein residues of the lysozyme for PTFE adsorption, we clustered the docking poses based on their pose similarity, i.e. poses which have the highest similarity in their residue's minimum distance to surface profiles are grouped together. In this way, three sets of similar docking poses are obtained: the first set includes pso1, pso5, pso8, and pso9; the second set includes pso2, pso3, pso6, pso7; the third set has only one predicted pose which is pso4. Pso10 is not classified because it has the worst energy score presumably due to premature convergence (completed in only 10 iterations). Docking poses in the same set are similar in orientation (r_x, r_y, r_z) but different in position (t_x, t_y, t_z) at the surface. Because the model surface is homogeneous, the

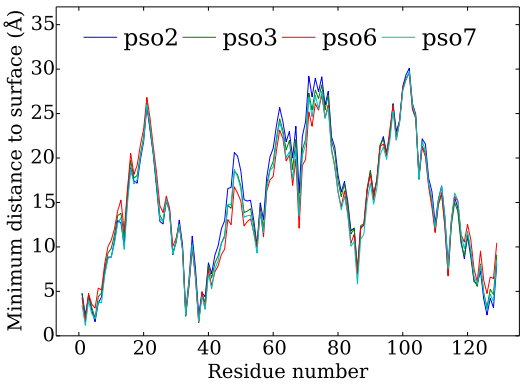


(a)

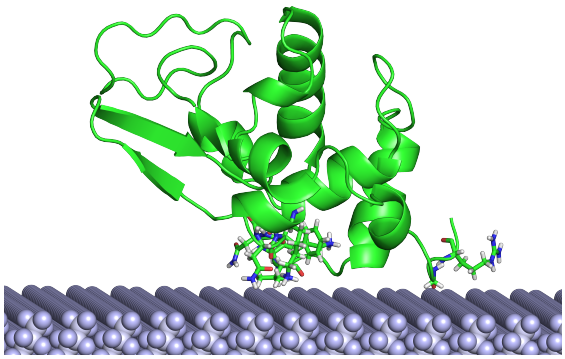


(b)

Fig. 3. Predicted docking poses from set 1: (a) Residue's minimum distance to the surface. (b) Snapshot of the lowest energy docking pose in the set, which is pso1. Identified adsorption sites are Arg14, His15, Asn77, Thr89, and Arg128. The protein is drawn with cartoon representation and the surface atoms with spheres. The contacting residues identified as the major adsorption sites in the set are shown as sticks. Water is modeled implicitly and thus it is not shown.



(a)



(b)

Fig. 4. Predicted docking poses from set 2: (a) Residue's minimum distance to the surface. (b) Snapshot of the lowest energy docking pose in the set, which is pso2. Identified adsorption sites are Lys1, Val2, Lys33, Asn37, Phe38, Asn39, Gly126, and Arg128.

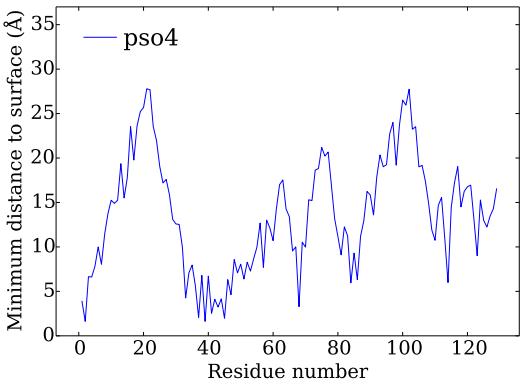
same or similar protein orientation at different surface position might achieve very similar energy scores.

For each set, we identify the *major* adsorption site of lysozyme as the contacting residue found in the majority of the docking poses in the set. The definition of "contacting residue" is taken from [24] that is the protein residue with a minimum distance of < 5 Å to the surface. Profiles of the residue's minimum distance to the surface in set 1–3 and snapshots of the best predicted docking pose in the set are shown in Figure 3–5 respectively.

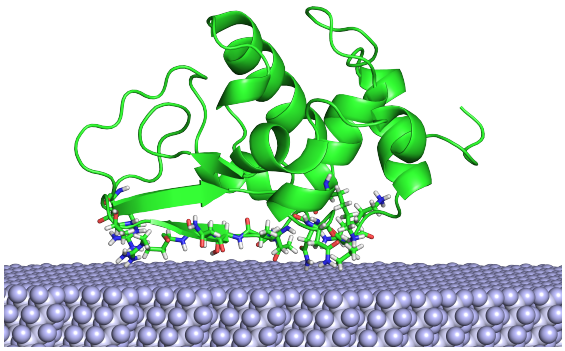
In set 1, five major adsorption sites are identified; they are Arg14, His15, Asn77, Thr89, and Arg128. Except Arg128 which is the contacting residue of 3 docking poses (out of 4), all others are contacting residues in all docking poses in the set by definition. It is noteworthy that Arg128 was also reported as a major adsorption site in a recent long molecular dynamics simulation of lysozyme on polyethylene surfaces (PE) by Wei et al. [3]. It was observed in their study that Arg128 and Leu129, the two C-terminal residues, reached the PE surface only after 100 ns in the simulation, whereas this contact is successfully identified by our prediction method in relatively much shorter computing time. Interestingly, adsorption sites Arg14, His15, Asn77, and Thr89 have so far not been reported in previous computational studies of lysozyme on hydrophobic surfaces. Instead, neighboring residues, Pro79 and Ile88, were found by Zheng et al. [9] as the closest residues to the CH_3-SAM surface in a monte carlo simulation for protein orientation prediction.

In set 2, eight major adsorption sites are identified; they are Lys1, Val2, Lys33, Asn37, Phe38, Asn39, Gly126, and Arg128. All are contacting residues from all docking poses in the set except Arg128, which can only be identified in 3 docking poses. Except Gly126, all major adsorption sites were determined to be the contacting residues in the MD simulation of Wei et al. [3].

Finally, in set 3, twelves major adsorption sites are identified; they are Lys1, Val2, Lys33, Asn37, Asn39, Gln41, Ala42, Thr43, Asn44, Arg45, Thr47, and Arg68. This time, all contacting residues are determined as the adsorption sites in Wei's study.



(a)



(b)

Fig. 5. Predicted docking poses from set 3: (a) Residue's minimum distance to the surface. (b) Snapshot of the lowest energy docking pose in the set, which is pso4. Identified adsorption sites are Lys1, Val2, Lys33, Asn37, Asn39, Gln41, Ala42, Thr43, Asn44, Arg45, Thr47, and Arg68.

IV. Conclusion

Particle swarm optimization is a fast global optimization method which is well-known to solve complex optimization problems. Here, we presented the first application of PSO on the prediction of protein docking poses on a solid surface. Using the lysozyme-PTFE model, our method successfully located the low energy docking poses of the protein, among which the predicted lowest energy pose is about 200 kJ mol^{-1} lower than the lowest structure found by a brute-force search. In addition, we analyzed the predicted adsorption sites of the protein and obtained good agreement to previous computational studies. Our current implementation employs an external program for energy evaluations. In the future, we plan to internalize the energy evaluation routines and explore efficient ways in computing interaction energies. A faster prediction tool would allow us to investigate different surface models and proteins for the fundamental understanding of the protein adsorption process and the rational design of surface.

Acknowledgment

This work was supported by the University of Macau under the grant no. of MYRG2014-00104-FST. Tested MD simulations were done at the High Performance Computing Cluster (HPCC) which is supported by Information and Communication Technology Office (ICTO) of the University of Macau.

References

[1] D. Falconnet, G. Csucs, H. M. Grandin, and M. Textor, "Surface engineering approaches to micropattern surfaces for cell-based assays," *Biomaterials*, vol. 27, no. 16, pp. 3044 – 3063, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142961206000081

[2] M. A. Dobrovolskaia and S. E. McNeil, "Immunological properties of engineered nanomaterials," *Nature Nanotechnology*, vol. 2, no. 8, pp. 469–478, 2007.

[3] T. Wei, M. A. Carignano, and I. Szleifer, "Lysozyme adsorption on polyethylene surfaces: Why are long simulations needed?" *Langmuir*, vol. 27, no. 19, pp. 12 074–12 081, 2011, pMID: 21846132. [Online]. Available: http://dx.doi.org/10.1021/la202622s

[4] K. Kubiak-Ossowska and P. A. Mulheran, "Multiprotein interactions during surface adsorption: a molecular dynamics study of lysozyme aggregation at a charged solid surface," *The Journal of Physical Chemistry B*, vol. 115, no. 28, pp. 8891–8900, 2011, pMID: 21671567. [Online]. Available: http://dx.doi.org/10.1021/jp1121239

[5] G. Yu, J. Liu, and J. Zhou, "Mesoscopic coarse-grained simulations of lysozyme adsorption," *The Journal of Physical Chemistry B*, vol. 118, no. 17, pp. 4451–4460, 2014, pMID: 24785197. [Online]. Available: http://dx.doi.org/10.1021/jp409326f

[6] M. Agashe, V. Raut, S. J. Stuart, and R. A. Latour, "Molecular simulation to characterize the adsorption behavior of a fibrinogen γ-chain fragment," *Langmuir*, vol. 21, no. 3, pp. 1103–1117, 2005, pMID: 15667197. [Online]. Available: http://dx.doi.org/10.1021/la0478346

[7] V. P. Raut, M. A. Agashe, S. J. Stuart, and R. A. Latour, "Molecular dynamics simulations of peptide-surface interactions," *Langmuir*, vol. 21, no. 4, pp. 1629–1639, 2005, pMID: 15697318. [Online]. Available: http://dx.doi.org/10.1021/la047807f

[8] M. Kleijn and W. Norde, "The adsorption of proteins from aqueous solution on solid surfaces," *Heterogeneous Chemistry Reviews*, vol. 2, no. 3, pp. 157–172, 1995.

[9] J. Zheng, L. Li, S. Chen, and S. Jiang, "Molecular simulation study of water interactions with oligo (ethylene glycol)-terminated alkanethiol self-assembled monolayers," *Langmuir*, vol. 20, pp. 8931–8938, 2004.

[10] K. Makrodimitris, D. L. Masica, E. T. Kim, and J. J. Gray, "Structure prediction of protein-solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite," *Journal of the American Chemical Society*, vol. 129, no. 44, pp. 13 713–13 722, 2007, pMID: 17929924. [Online]. Available: http://dx.doi.org/10.1021/ja074602v

[11] H.-J. Hsu, S.-Y. Sheu, and R.-Y. Tsay, "Preferred orientation of albumin adsorption on a hydrophilic surface from molecular simulation," *Colloids and Surfaces B: Biointerfaces*, vol. 67, no. 2, pp. 183 – 191, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0927776508002981

[12] T. Wei, S. Mu, A. Nakano, and K. Shing, "A hybrid multi-loop genetic-algorithm/simplex/spatial-grid method for locating the optimum orientation of an adsorbed protein on a solid surface," *Computer Physics Communications*, vol. 180, no. 5, pp. 669 – 674, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010465508003809

[13] T. Wei, M. A. Carignano, and I. Szleifer, "Molecular dynamics simulation of lysozyme adsorption/desorption on hydrophobic surfaces," *The Journal of Physical Chemistry B*, vol. 116, no. 34, pp. 10 189–10 194, 2012, pMID: 22882159. [Online]. Available: http://dx.doi.org/10.1021/jp304057e

[14] H. K. Chan, S. W. I. Siu, and P. I. Mak, "Electrowetting of water nanodroplet on coplanar electrodes studied by molecular dynamics simulations," *Manuscript in preparation*.

[15] J. Gao, X. Liu, T. Chen, P.-I. Mak, Y. Du, M.-I. Vai, B. Lin, and R. P. Martins, "An intelligent digital microfluidic system with fuzzy-enhanced feedback for multi-droplet manipulation," *Lab Chip*, vol. 13, pp. 443–451, 2013. [Online]. Available: http://dx.doi.org/10.1039/C2LC41156C

[16] R. Diamond, "Real-space refinement of the structure of hen egg-white lysozyme," *Journal of Molecular Biology*, vol. 82, no. 3, pp. 371 – 391, 1974. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0022283674905981

[17] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation," *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008. [Online]. Available: http://dx.doi.org/10.1021/ct700301q

[18] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 4, Nov 1995, pp. 1942–1948 vol.4.

[19] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, May 1998, pp. 69–73.

[20] M. Clerc, "From theory to practice in particle swarm optimization," in *Handbook of Swarm Intelligence*, ser. Adaptation, Learning, and Optimization, B. Panigrahi, Y. Shi, and M.-H. Lim, Eds. Springer Berlin Heidelberg, 2011, vol. 8, pp. 3–36. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-17390-5_1

[21] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11 225–11 236, 1996. [Online]. Available: http://dx.doi.org/10.1021/ja9621760

[22] A. Onufriev, D. Bashford, and D. A. Case, "Exploring protein native states and large-scale con- formational changes with a modified generalized born model," *Proteins: Struct. Funct. Gen.*, vol. 55, no. 2, pp. 383–394, 2004.

[23] Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.3r1," August 2010.

[24] G. Raffaini and F. Ganazzoli, "Protein adsorption on a hydrophobic surface: A molecular dynamics study of lysozyme on graphite," *Langmuir*, vol. 26, no. 8, pp. 5679–5689, 2010, pMID: 20041676. [Online]. Available: http://dx.doi.org/10.1021/la903769c