Computer Physics Communications 180 (2009) 669–674

Contents lists available at ScienceDirect

Computer Physics Communications

www.elsevier.com/locate/cpc

A hybrid multi-loop genetic-algorithm/simplex/spatial-grid method for locating the optimum orientation of an adsorbed protein on a solid surface

Tao Wei [a], Shengjing Mu [b], Aiichiro Nakano [c], Katherine Shing [a,*]

[a] Mork Family Department of Chemical Engineering and Material Science, University of Southern California, Los Angeles, CA, USA
[b] Yokogawa Engineering Asia Pte Ltd, Singapore
[c] Department of Computer Science, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 3 September 2008
Received in revised form 1 November 2008
Accepted 5 November 2008
Available online 18 November 2008

PACS:
02.70.-c
02.70.Ns
87.15.-v
68.35.Md

Keywords:
Hybrid genetic algorithm
Spatial grid method
Lysozyme
Adsorption
Optimum orientations and positions

ABSTRACT

Atomistic simulation of protein adsorption on a solid surface in aqueous environment is computationally demanding, therefore the determination of preferred protein orientations on the solid surface usually serves as an initial step in simulation studies. We have developed a hybrid multi-loop genetic-algorithm/simplex/spatial-grid method to search for low adsorption-energy orientations of a protein molecule on a solid surface. In this method, the surface and the protein molecule are treated as rigid bodies, whereas the bulk fluid is represented by spatial grids. For each grid point, an effective interaction region in the surface is defined by a cutoff distance, and the possible interaction energy between an atom at the grid point and the surface is calculated and recorded in a database. In searching for the optimum position and orientation, the protein molecule is translated and rotated as a rigid body with the configuration obtained from a previous Molecular Dynamic simulation. The orientation-dependent protein–surface interaction energy is obtained using the generated database of grid energies. The hybrid search procedure consists of two interlinked loops. In the first loop A, a genetic algorithm (GA) is applied to identify promising regions for the global energy minimum and a local optimizer with the derivative-free Nelder–Mead simplex method is used to search for the lowest-energy orientation within the identified regions. In the second loop B, a new population for GA is generated and competitive solution from loop A is improved. Switching between the two loops is adaptively controlled by the use of similarity analysis. We test the method for lysozyme adsorption on a hydrophobic hydrogen-terminated silicon (110) surface in implicit water (i.e., a continuum distance-dependent dielectric constant). The results show that the hybrid search method has faster convergence and better solution accuracy compared with the conventional genetic algorithm.

1. Introduction

Protein adsorption on solid surfaces has been a focus of intense research for more than three decades. In order to control and manipulate protein adsorption, the mechanisms which govern the adsorption process need to be well understood. Atomistic Molecular Dynamic (MD) simulation is an important computational technique used to model the interatomic interactions realistically and to elucidate the orientation alignments and conformation evolution of an adsorbed protein molecule. Due to its large size, protein molecular rotation is slow if compared with other dynamic events such as segment relaxation and local conformational changes. Therefore, the determination of the preferred protein orientations on the solid surface usually serves as an initial step in simulation studies. Typically, brute-force search is applied to predict lysozyme retention behavior on a poly(vinylimidazole) polymer using detailed molecular models of protein and the surface and an implicit water model [1]. Orientation-dependent free energy contours were developed by searching the configurational space with the use of a quaternion rotation and grid-based calculation method with an empirical energy function and implicit solvent models [2]. Alternatively, Monte Carlo (MC) simulations can be employed to identify the lowest energy protein orientations in a continuum solvent medium prior to conduction of actual MD simulation in explicit water medium [3]. Brownian dynamics simulation is also commonly used to study the initial stages of protein adsorption behaviors [4].

Genetic algorithms (GAs) are intelligent stochastic methods, inspired by the Darwinian natural evolution principle of "survival of the fittest". GAs have wide applications as global optimization methods with distinguishing characteristics such as being highly parallel, robust and derivative-free. GAs exhibit great efficiency and accuracy in studies involving molecular geometry optimization for systems with a large number of atoms and a large number of local energy minima, whereas conventional MD simulation and simu-

* Corresponding author.
E-mail address: shing@usc.edu (K. Shing).

0010-4655/$ – see front matter
doi:10.1016/j.cpc.2008.11.008

lated annealing techniques can be trapped in local minima [5,6]. GA defines a genetic representation of the solution domain characterized by a fitness function. A standard evolution proceeds in cycles of initialization, selection and reproduction until termination conditions have been reached. Standard GAs are known to be weak in the exploration of optimum solutions and to suffer from "premature" convergence. A number of hybrid GAs have been proposed to improve the performance of standard GA [7]. Intensive research has been carried out to address the problem of premature convergence in the application of GA in automatic protein/ligand docking and structure-based design, with focuses on algorithm design [8], the fitness function design [9,10], and the choice of operator combination [11].

Locating the optimum orientations of an adsorbed protein molecule on a solid surface is a complicated global optimum problem and needs to sample enormous space, similar to the protein/ligand docking problem. In this paper, we present an effective method for searching the low energy protein location and orientations on a solid surface. The method combines an adaptive, hybrid GA/NM-simplex search and spatial grid-based atom–surface interaction energy computations. Adsorption of lysozyme on a H–Si (110) surface is studied as a model system. Chicken egg lysozyme has been extensively studied [12] as a model protein in order to understand the mechanism of protein structure, folding, function and adsorption behavior. Hydrogen-terminated Si (H–Si) surfaces are frequently used in adsorption studies because they can effectively suppress the formation of the oxidation layer and can also be modified to vary the surface morphological, hydrophobic/hydrophilic and chemical properties [13–15]. We compare the performances of the hybrid-GA loop search and a standard GA in locating the optimum lysozyme orientations and distances from the H–Si surface. The accuracy of the searches is verified by comparison to the results obtained by a brute-force contour search.

2. Method

In this paper, the surface and the protein molecule are treated as rigid bodies. The surface is modeled as a hydrophobic (110) silicon cleavage plane with dimensions 65.3 Å × 70.5 Å × 18 Å, terminated by hydrogen atoms at the top. The lysozyme configuration was previously deduced by a bulk all-atom MD simulation in explicit water at pH = 7.4. The simulated lysozyme structure is in good agreement with solution NMR data (PDB code: 1E8LA) [12], which was downloaded from the protein data bank (http://www.rcsb.org/pdb/home/home.do). An implicit water model (i.e., a continuum with distance-dependent dielectric constant) is chosen to simulate the aqueous environment. The rigid body and implicit water assumption are justified here because our interest in this study is to establish the initial orientation and position of a protein molecule on the surface as the first step to a full MD simulation, in which water is explicitly modeled and protein relaxation is permitted.

2.1. Empirical energy function and grid-based computation

The interatomic interaction energy between a lysozyme molecule and the Si–H surface includes van der Waals (vdW) energy and electrostatic energy represented by Consistent Valence Force Field (CVFF) parameters [16]. The protein–surface interaction energy is obtained by summing over all protein atoms (N_p) and all surface atoms (N_s), as shown in Eqs. (1)–(3):

$$E = \sum_{i=1}^{N_p}\sum_{j=1}^{N_s}\left(\frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^{6}}\right) + \sum_{i=1}^{N_p}\sum_{j=1}^{N_s}\frac{332(q_i\cdot q_j)}{D_r(r_{i,j})r_{i,j}}, \tag{1}$$

$$A_{i,j} = \sqrt{A_i\cdot A_j}, \tag{2}$$

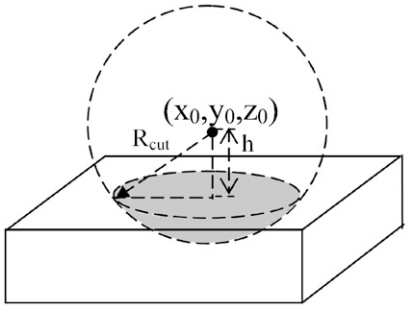


Fig. 1. Effective interaction region defined by a cutoff distance (R_{cut} = 18 Å) for evaluating interaction energies between an atom at the grid position (x_0, y_0, z_0) in the bulk space and the surface.

$$B_{i,j} = \sqrt{B_i\cdot B_j}, \tag{3}$$

where E is interaction energy between a lysozyme molecule and the surface in kcal/mol; r_{ij} is the interatomic distance (Å) between atom i in the lysozyme molecule and atom j on the surface; A_i and B_i are vdW energy parameters for atom i; D_r is the distance dependent dielectric constant to approximate the effect of water environment on the protein adsorption [2], given as follows:

$$D_r(r_{i,j}) = D_A + \frac{D_B}{1 + k\cdot e^{-\lambda\cdot D_B\cdot r_{i,j}}}, \tag{4}$$

where $D_A = -8.5525$, $D_B = \varepsilon_0 - D_A$, $\varepsilon_0 = 78.4$, $k = 7.7839$ and $\lambda = 0.03624$ Å^{-1}.

Because the Si–H surface is modeled as a rigid body, interaction energy computational efficiency can be improved by gridding the bulk space above the surface and creating a database of the absolute values of the three energy terms given by Eq. (5)–(7): vdW repulsive potential ($E_{\mathrm{vdW1}}(\vec{r}_i)$), vdW attractive potential ($E_{\mathrm{vdW2}}(\vec{r}_i)$) and electrostatic potential ($E_{\mathrm{elec}}(\vec{r}_i)$), associated with each grid point:

$$E_{\mathrm{vdW1}}(\vec{r}_i) = \sum_{j=1}^{N_s}\frac{\sqrt{A_j}}{r_{i,j}^{12}}, \tag{5}$$

$$E_{\mathrm{vdW2}}(\vec{r}_i) = \sum_{j=1}^{N_s}\frac{\sqrt{B_j}}{r_{i,j}^{6}}, \tag{6}$$

$$E_{\mathrm{elec}}(\vec{r}_i) = \sum_{j=1}^{N_s}\frac{332q_j}{D_r\cdot r_{i,j}}. \tag{7}$$

For each bulk-space grid point, an effective interaction region in the surface is defined by a cutoff distance as shown in Fig. 1, while the three interaction energy terms (Eqs. (5)–(7)) between an atom at the grid point and the N_s surface atoms in the effective interaction region are calculated and recorded.

The interaction energy between each of the N_p atoms in the protein molecule and the N_s surface atoms in the interaction region is obtained by interpolation using the stored energy values corresponding to the 8 nearest grid points [17]. The interaction energy between the whole protein molecule and the surface is calculated by summation over all the N_p atoms in the protein molecule as given in Eq. (8):

$$E = \sum_{i=1}^{N_p}\sqrt{A_i}E_{\mathrm{vdW1}}(\vec{r}_i) - \sum_{i=1}^{N_p}\sqrt{B_i}E_{\mathrm{vdW2}}(\vec{r}_i) + \sum_{i=1}^{N_p}q_iE_{\mathrm{elec}}(\vec{r}_i). \tag{8}$$

Interpolation of the electrostatic energy from Eq. (7) can be calculated with simple trilinear interpolation. However vdW potential terms (Eqs. (5) and (6)) cannot be approximated by simple trilinear interpolation due to the possible occurrence of large errors [17]. In

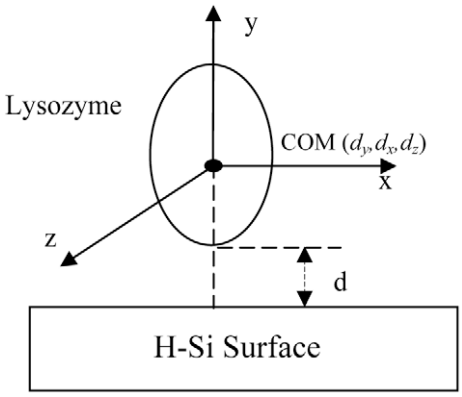


Fig. 2. Protein's three translational degrees of freedom (d_y, d_x, d_z) on a Si–H surface.

this work, we calculate exactly the vdW potential terms for protein atoms located in the close surface region (⩽2 Å) (see Fig. 1). For protein atoms farther away from the surface (>2 Å), we employ a modified trilinear interpolation scheme [17].

2.2. Protein translation and rotation on the surface

A complete search of the protein/surface minimum interaction energy requires six degrees of freedom (3 rotations and 3 translations). As shown in Fig. 2, the three translational degrees of freedoms are defined as the distance of the protein center of mass (COM), d_y, to the top surface layer atoms and the displacements, d_x and d_z, of the protein COM on the plane parallel to the surface layer. The distance to the top of surface, d_y, is adjusted to maintain the minimum protein/surface atomic distance (d) in the range of 1.2 Å $\leqslant d \leqslant$ 5 Å. Because Si surface is periodic and homogenously terminated with hydrogen atoms, the ranges of d_x and d_z are defined as the unit cell dimensions of the crystal, i.e., 0 Å $\leqslant d_x \leqslant$ 3.8396 Å, 0 Å $\leqslant d_z \leqslant$ 5.4307 Å. The protein orientation is defined by a unique set of 3 consecutive rotations around its center of mass represented by a set of Euler angles (Φ, θ, ψ) [18] in the ranges: $0 \leqslant \Phi < 2\pi$; $-0.5\pi \leqslant \theta \leqslant 0.5\pi$; $0 \leqslant \psi < 2\pi$.

2.3. Standard genetic algorithm

Standard GA is preformed by modifying the GA optimization toolbox (Gaot) [19] developed by Houck et al. We first briefly describe the real-coded standard GA which will be compared with the hybrid GA loop search scheme discussed in the next section. The fitness in the GA and the object function for the local NM-simplex search in the hybrid GA loop search are represented by the negative total interaction energy $(-E)$ in Eq. (1). For the standard GA, the initial solutions with population size (p) are randomly generated. For the system studied, in evolution generation of i, the solution is coded as a vector $[a_{i1}, a_{i2} \ldots a_{i6}]$, where a_{ik} (k = 1–6) represents the COM displacement (d_y, d_x, d_z) and orientation (Φ, θ, ψ) of the lysozyme molecule as defined in Section 2.2. The usual normalized geometric selection operator [20], arithmetic crossover operator [21] and non-uniform mutation operator [21] are implemented. Normalized geometric selection method is a ranking selection method based on normalized geometric distribution. The probability of selecting the fittest individual at an early stage of evolution is limited, and the least fit members of population are gradually driven out of the population during the evolution. The probability P_i of selecting a current individual from the old population is defined as

$$P_i = \frac{q}{1-(1-q)^p} \cdot (1-q)^{r-1}, \tag{9}$$

where q is the possibility of selecting the best individual (set $q = 0.08$), r is the rank of the individual, and p is the population size, set at 100 in this study. The selected parents undergo the followed reproduction (crossover and mutation). Arithmetic crossover operator produces two offsprings, s_a^{i+1} and s_b^{i+1} in generation $i+1$, which are complimentary linear combinations of parents, s_a^i and s_b^i, in generation i according to:

$$s_a^{i+1} = Rand \cdot s_b^i + (1 - Rand) \cdot s_a^i, \tag{10}$$

$$s_b^{i+1} = Rand \cdot s_a^i + (1 - Rand) \cdot s_b^i, \tag{11}$$

where *Rand* is a uniform random number in the range [0, 1]. The non-uniform mutation operator has the capability of fine tuning. For the element a_{ik} (k = 1–6) in the boundary range $[LB_k, UB_k]$ of a solution $s_a^i = [a_{i1}, a_{i2} \ldots a_{i6}]$, the mutated element a'_{ik} is defined as

$$a'_{ik} = \begin{cases} a_{ik} + \Delta(i, UB_k - a_{ik}), & Rand \geqslant 0.5, \\ a_{ik} - \Delta(i, a_{ik} - LB_k), & Rand < 0.5, \end{cases} \tag{12}$$

$$\Delta(i, y) = y \cdot \left(Rand \cdot \left(1 - \frac{i}{N_{\max}} \right) \right)^m, \tag{13}$$

where i is the generation number; LB_k and UB_k represent lower and upper boundaries of a_{ik} respectively; $N_{\max}$ is the maximum evolution generation number set at 1200 in this study, and m is a shape parameter set at 2. An elitist model [21] is used to maintain the best solution in the evolution process.

2.4. Hybrid genetic loop search scheme

Standard GA is expected to have good global search capability and to be able to overcome potential barriers. However, in some complicated cases, such as automated docking and protein adsorption, GA can become trapped in local potential wells (premature convergence). In this paper, we study the protein introduce a GA loop search coupled with a local search scheme to overcome this problem. The search procedure consist of two loops (loop A and loop B), as illustrated in the flowchart (Fig. 3). The hybrid GA search is executed in loop A, which keeps the improvement of the best solution. An additional loop search (loop B) is added to enlarge the population diversity during the evolution in order to avoid prevent premature convergence. The population number (p) in the hybrid GA loop search is set at 100 as in the standard GA.

The Nelder–Mead (NM) simplex method, which is derivative-free local search, is incorporated into the hybrid GA loop search scheme. GA is used to maximize the fitness value of the population, while NM simplex search is used to fine-tune the solutions in loop A, as well as to prune the search space [22] for the GA loop search in loop B.

In loop A, we combine a standard real-coded GA with NM simplex method to search for the low energy protein orientations in the generated grid energy database. The standard GA search procedures (selection, crossover and mutation) are first performed, and then local NM simplex search is executed. The frequency of the execution of local search method is determined by the parameter *N_L* (= 80 GA generations between consecutive NM simplex searches in this work). For every local search, a 10-step NM simplex search is used. The decision regarding whether to terminate a GA loop search (loop A) and initiate a new one (loop B) is made based on a Similarity Coefficient (SC) analysis [23]. SC analysis is executed every *N_S* (= 20) generations. By using this adaptive control, we expect that the occurrence of GA premature convergence can be reduced. The SC analysis is defined as follows. For a given generation, an individual solution is represented by a real-number vector, e.g., $[a_{i1}, a_{i2} \ldots a_{in}]$. The similarity coefficient SC_{ij} between individuals, i and j, for the same generation is defined as

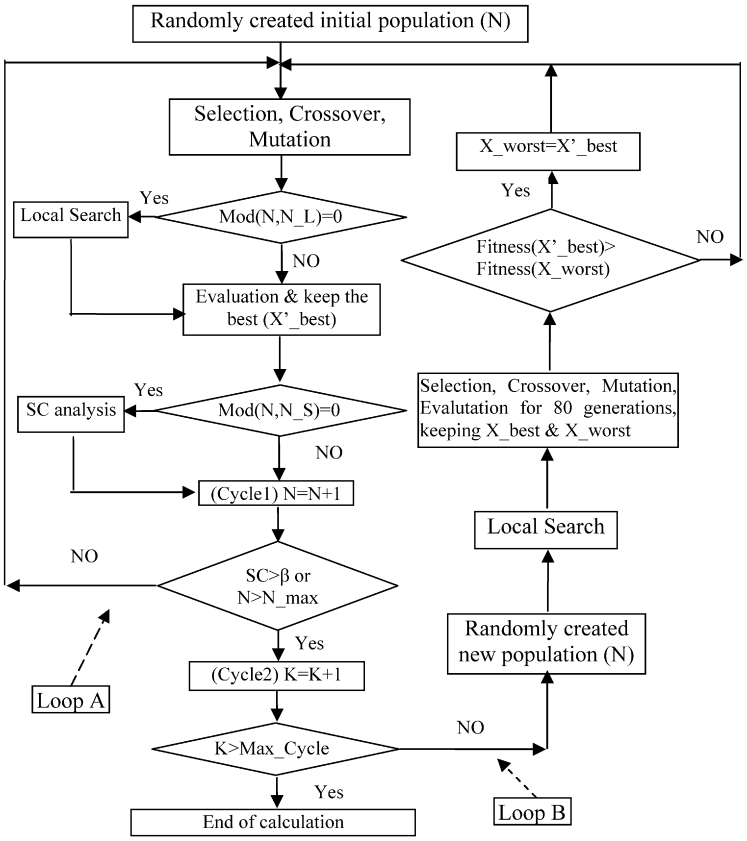


Fig. 3. Flow chart for the hybrid genetic algorithm.

$$SC_{ij} = \frac{\sum_{k=1}^{n} \delta(a_{ik}, a_{jk})}{n}, \tag{14}$$

$$\delta(a_{ik}, a_{jk}) = \begin{cases} 1, & \text{if } \left|\frac{a_{ik}-a_{jk}}{\bar{a}_{(i,j)k}}\right| \leqslant \alpha, \\ 0, & \text{otherwise}, \end{cases} \tag{15}$$

where n is the total number of individuals per generation, $\bar{a}_{(i,j)k}$ is the average of a_{ik} and a_{jk}, and α is the threshold value of similarity between α_{ik} and α_{jk}, which is set as 0.999 in the program. The average similarity, $\overline{SC}$ for this generation, is expressed as

$$\overline{SC} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} SC_{ij}}{n}. \tag{16}$$

The value of similarity coefficient, $\overline{SC}$, for all individuals in the same generation is compared with threshold value β and is used to control the loop iteration as

$$\begin{cases} \text{stop loop A}, & \text{if } \overline{SC} > \beta, \\ \text{repeat loop A}, & \text{otherwise}, \end{cases} \tag{17}$$

where β is 0.99 in the program. If SC is unsatisfactory (the solutions do not converge) and cycle number of loop A is smaller than the threshold value (*N_max*), loop A is repeated; otherwise, loop A is terminated and loop B is initiated. The best solution from loop A is recorded.

When the SC condition is satisfied, loop B is initiated to enlarge the search space. In loop B search, a new population is randomly generated and improved through combination of a global and a local search for 80 generations. The worst solution (*X_worst*) in loop B is replaced by the best solution (*X′_best*) from the previous search of loop A, if the latter proves to be better fit than the former. The solutions serve as the initial population to start a new iteration of loop A, which is then re-executed until the number of loop A passes reaches the maximum value (*Max_Cycle*).

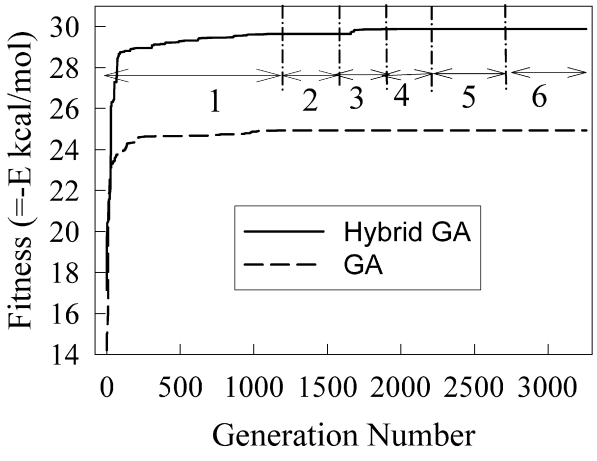


Fig. 4. Comparison of standard GA and hybrid GA. Fitness is plotted as a function of GA generations. The termination of 6 runs in the loop is adaptively controlled by SC analysis. The numbers (1–6) within the plot denote 6 passes of loop A.

3. Results and discussion

We apply the hybrid-GA loop method to the examination of lysozyme adsorption on a hydrophobic H–Si(110) surface in implicit water. The performance of the hybrid GA in a 6 degrees-of-freedom (3 translational plus 3 rotational) search is compared with that of the standard GA in Fig. 4. The cumulative number of generations for loop A of the hybrid-GA loop search is shown in Fig. 4, while the additional 80 generations per every loop B cycle is not included. In the hybrid GA loop search, the SC condition has been satisfied 6 times and loop B has been executed 5 times. The length of each pass of loop A is adaptively con-

Table 1
Comparison of the standard GA, hybrid-GA loop and brute-force searches.

	(d_y, d_x, d_z) (Å)	(Φ, θ, ψ) (radians)	Fitness, $-E$ (kcal/mol)	Runtime[a] (min)
Standard GA search	(1.5144, −1.9198, −2.0234)	(2.6125, 1.5149, 1.9797)	24.929	52
Hybrid GA search	(1.2305, −0.8206, −0.3489)	(0.6720, 1.2204, 6.2562)	29.878	85
Brute-force search	fixed at (1.2305, −0.8206, −0.3489)	(0.6981, 1.222, 6.2831)	29.886	771

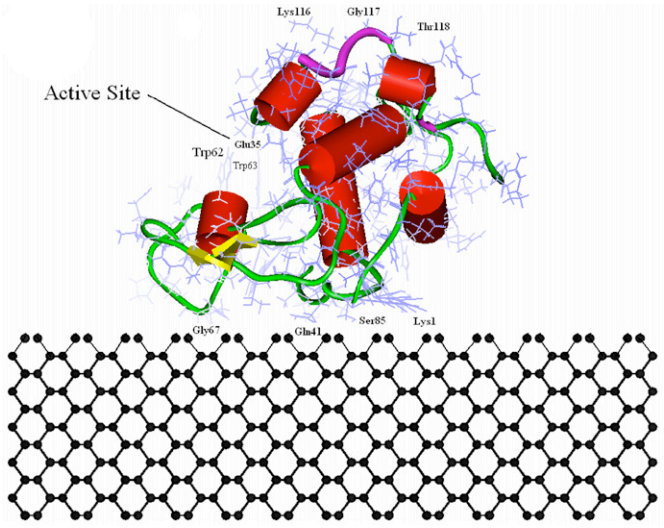
[a] This computation is carried out on single core Pentium IV, 3.0 GHz.

trolled by SC analysis. Compared with the initial pass of loop A, the length of subsequent passes of loop A is shorter. This is understandable because, as the optimum solution is approached, the converge becomes easier as the program fine-tunes the solutions. We see that the local exploration capability is enhanced when NM simplex search is employed. The premature convergence of the GA is effectively prevented without sacrificing the computational efficiency. In contrast, the standard GA clearly suffers from premature convergence.

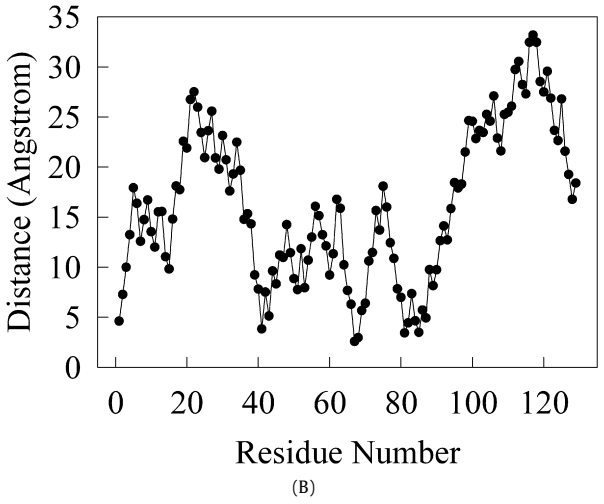
In order to confirm that the optimum orientation is indeed detected by the hybrid GA search, a brute-force exact calculation is carried out. The protein molecule center of mass location is fixed at the optimum (d_x, d_y, d_z), which is obtained from the 6-degrees-of-freedom hybrid GA search, whereas the 3 angles, Φ, θ and ψ, are systematically altered over the complete relevant range in intervals of 4°. The results are sorted according to their fitness. The lowest energy (best fitness) results are compared in Table 1. We see that the 6-degrees-of-freedom hybrid GA yields results very close to the (exact) brute-force results, whereas the standard GA is stranded at a local minimum (at about −25 kcal/mol) far away from the optimum (at about −30 kcal/mol). The more accurate result of the hybrid GA is accompanied by computational overhead (additional calculations in loop B cycles, the periodic execution of NM simplex local searches, and the SC analysis), which is only 85% of the computational time of the standard GA, whereas the hybrid GA is 8 times faster than the brute-force search (in intervals of 4° of the three Euler angles) with nearly identical results. The result in Fig. 4 and Table 1 demonstrate the high accuracy and efficiency of the hybrid-GA loop search.

Fig. 5(A) shows the lysozyme position and orientation on the surface corresponding to the converged result of the hybrid-GA search. The active site associated with lysozyme's catalytic ability is seen pointing away from the surface. Fig. 5(B) shows the center of mass distance of each amino acid residue from the surface. The distribution indicates that residues 40–80, which are predominantly involved in turn structure, are located close to the surface.

It should be noted that, in order to determine the equilibrium conformation and orientation of the protein on the surface, a full MD simulation is needed, using the optimum positions and orientations located in this work as a starting point.



(A)



(B)

Fig. 5. Lowest-energy lysozyme orientation on Si–H surface: (A) graphical representation showing lysozyme activity and secondary structures: α-helix (yellow), β-sheet (red), turn (green), random coil (purple); (B) lysozyme amino acid residue distribution relative to the surface, where each amino acid residue is represented by a dot.

4. Summary

We have developed a hybrid genetic-algorithm method that combines both local and global search capabilities in two interlinked search loops. The switching between the two loops is adaptively controlled by the use of similarity coefficient analysis. The method is applied to the search of the optimum center of mass location and molecular orientation of lysozyme on H-terminated Si surface. Comparisons are made among the hybrid GA, the standard GA and a brute-force search. The results show that the standard GA algorithm suffers from premature convergence (converging to a local minimum), whereas the hybrid GA loop search successfully locates the global minimum without sacrificing rapid convergence. The optimality of the solution has been validated by comparing it with brute-force search. The optimum orientation and position found in this research will serve as the initial values for further all-atom MD simulations.

Acknowledgements

This work was partially supported by NSF. Numerical tests were performed at the University of Southern California using the 52.4 teraflops Linux cluster at the Research Computing Facility.

References

[1] V. Noinville, C. Vidal-Madjar, B. Sebille, J. Phys. Chem. 99 (1995) 1516–1522.
[2] Y. Sun, W.J. Welsh, R.A. Latour, Langmuir 21 (2005) 5616–5626.

[3] J. Zhang, L.Y. Li, S.F. Chen, S.Y. Jiang, Langmuir 20 (2004) 8931–8938.
[4] S. Ravichandran, J.D. Madura, J. Talbot, J. Phys. Chem. B 105 (2001) 3610–3613.
[5] D.M. Deaven, K.M. Ho, Phys. Rev. Lett. 75 (1995) 288–290.
[6] J. Maddox, Nature 376 (1995) 209.
[7] J. Yen, J.C. Liao, B. Lee, D. Randolph, IEEE Trans. Syst. Man Cybern. 28 (1998) 173–191.
[8] M. Cecchini, P. Kolb, N. Majeux, A. Caflisch, J. Comput. Chem. 25 (2003) 412–422.
[9] D. Douguet, H. Munier-Lehmann, G. Labesse, S. Pochet, J. Med. Chem. 48 (2005) 2457–2468.
[10] V.J. Gillet, P. Willett, J. Bradshaw, J. Chem. Inf. Comput. Sci. 38 (1998) 165–179.
[11] R. Thomsen, BioSystems 72 (2003) 57–73.
[12] H. Schwalbe, S.B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C.M. Dobson, C. Redfield, L.J. Smith, Protein Sci. 10 (2001) 677–688.
[13] T. Strother, R.J. Hamers, L.J. Smith, Nucleic Acids Res. 28 (2000) 3535–3541.
[14] T.J. Su, R.J. Green, Y. Wang, E.F. Murphy, J.R. Lu, Langmuir 16 (2000) 4999–5007.
[15] T.L. Clare, B.H. Clare, B.M. Nichols, N.L. Abbott, R.J. Hamers, Langmuir 21 (2005) 6344–6355.
[16] P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, A.T. Hagler, Proteins: Struct., Function Genetics 41 (1988) 31–47.
[17] C.M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, J. Mol. Graphics Modell. 21 (2003) 289–307.
[18] W.F. Phillips, Mechanics of Flight, John Wiley & Sons, Inc., New Jersey, 2004.
[19] C. Houck, J. Joines, M. Kay, The Genetic Algorithm Optimization Toolbox (GAOT) for Matlab 5, Technical Report: NCSU-IETR-95-09, North Carolina State University, Raleigh, NC, 1995.
[20] J.A. Joines, C.R. Houck, On the use of non-stationary penalty functions to solve constrained optimization problems with genetic algorithms, in: IEEE International Symposium Evolutionary Computation, Orlando, Fl, 1994, pp. 579–584.
[21] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Program, 3rd ed., Springer-Verlag, New York, 1996.
[22] J. Yen, B. Lee, D. Randolph, J.C. Liao, IEEE Trans. Neural Networks 4 (1998) 173–191.
[23] Y. Yun, Comp. Ind. Eng. 51 (2006) 128–141.