Application of a Molecular Dynamics Simulation Method with a Generalized Effective Potential to the Flexible Molecular Docking Problems

Youngshang Pak and Shaomeng Wang*

Georgetown Institute of Cognitive and Computational Sciences, and Departments of Oncology and Neuroscience, Georgetown University Medical Center, The New Research Building, Room EP07, 3970 Reservoir Road, Washington, D.C. 20007

Received: August 30, 1999; In Final Form: October 21, 1999

We present a new molecular dynamics method using Tsallis effective potential for the flexible docking problems of streptavidin/biotin and protein kinase C/phorbol-13-acetate. With a full flexibility of the ligands and a partial flexibility of the receptor active sites included, the new MD scheme accelerates the docking process significantly, by way of infrequent *q*-jumping and *q*-relaxation procedures between a normal potential energy surface and its transformed one by the Tsallis scheme. In the transformation, only the nonbonding interaction terms of an empirical potential energy function were employed. It has been found that the current method can predict the correctly docked structures in quite an effective way. Current results strongly indicate that this new MD method can be a very promising strategy for flexible ligand/flexible receptor docking.

I. Introduction

Predicting structure of a ligand−receptor complex is one of the essential problems in structural biology and structure-based drug design. Thus molecular docking becomes an important computational strategy for molecular recognition and drug design, and it has been a subject of numerous investigations, using various protocols, such as Monte-Carlo (MC),[1−3] molecular dynamics (MD),[4] and genetic algorithm (GA).[5] Recently Vieth et al.[6,7] have carried out an extensive investigation on selectivity of energy function and docking efficiency and also have assessed several searching strategies for flexible docking, based on simulated annealing (SA) and GA protocols.

Previously we reported that a new MD scheme with Tsallis effective potential derived from an empirical force field demonstrated a remarkable improvement in conformational searching efficiency, folding a 16-residue peptide initially in its extended conformation into a complete α-helix in a reasonably short time (300 ps).[8] The key idea of the method is so-called "infrequent smoothing" of the nonbonding interaction potential of any empirical force fields. The smoothing procedure can be achieved by a simple mathematical transformation (Tsallis transformation)[9,10] of the nonbonding interaction potential into a more delocalized and smoother counterpart, so that conformational searching process can possibly be enhanced by running MD on the transformed potential energy surface. For this purpose, infrequent jumping between the original and transformed surfaces is allowed during MD runs. Considering the case of molecular docking in which a large configurational space with a rugged potential energy surface needs to be searched in order to identify a correct binding mode, the new MD method may be a promising docking strategy to tackle the problem, due to its natural capability of smoothing the nonbonding interactions. The van der Waals repulsive interaction can also be reduced by simply scaling the Lennard-Jones parameters $\sigma_{i,j}$. In this case, however, locations of local minima are also shifted, while they are preserved in the Tsallis transformation. In this paper, motivated by the previous success of the new scheme in folding the 16-residue peptide,[8] we have attempted to extend the scope of the new methodology to a domain of flexible docking problems in which flexibility of both ligand and receptor is considered.

In recent years, a number of new docking methods were developed to include a ligand flexibility.[3,11−13] It has been also recognized in many cases that a flexibility in active sites of receptors plays an important role for binding to their ligands and for their biological functions.[14−16] Including the receptor flexibility in docking algorithms, however, remains quite a challenging problem due to the difficulty in dealing with conformational changes of receptors upon ligand binding. Thus so far only a limited number of docking studies have been carried out by dealing with the issue of the receptor flexibility.[17−19] In this work, we have included a full flexibility of ligands and for a more realistic description of their binding sites, we have also incorporated a side chain flexibility of receptor residues near the binding sites. We have chosen two complex systems in our docking investigations. One is the streptavidin/biotin complex, in which the active site of the receptor (streptavidin) is located in a cavity.[16] Since the active site of streptavidin is half-closed,[16] it is important to include a certain flexibility of the active site, in order for biotin to enter the binding site and to achieve a successful docking. The other is protein kinase C (PKC) in complex with phorbol-13-acetate, where side chains of several hydrophobic residues in PKC are flexible and involved in hydrophobic interactions with phorbol-13-acetate.[20] The 2-dimensional structures of both ligands are shown in Figure 1. Using the new MD scheme, we have carried out several MD runs in vacuum with the choice of different initial configurations where the ligands are several angstroms away from their receptor active sites. On the basis of the current study, we have found that the new method not only accelerates the docking procedure significantly but also locates the correct binding modes of the two cases in excellent agreement with the X-ray crystallographic results.[16,20] The sections of this paper are organized as follows:

* To whom correspondence should be sent. E-mail: wangs@giccs.georgetown. edu.

10.1021/jp993073h CCC: $19.00
Published on Web 12/14/1999

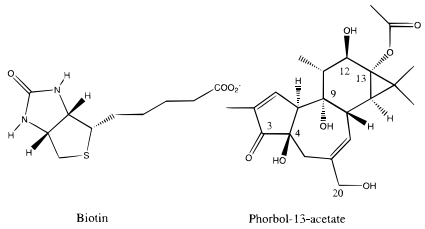

Figure 1. Two-dimensional structures of biotin and phorbol-13-acetate.

section II describes the detailed algorithm for the new MD method that we have implemented. In section III, docking results from the current method are presented and briefly discussed.

II. Method

The generalized statistical mechanics was first formulated by Tsallis[9,10] and later Andricioaei and Straub[21,22] applied the concept to an optimization of a tetraalanine peptide and atomic cluster simulations based on MC protocols. A main idea of the generalized statistical mechanics is to introduce a generalized entropy S_q[9,10] for an N-body system:

$$S_q = \frac{k_B}{q-1}\int P_q(\mathbf{r}^N)(1 - [P_q(\mathbf{r}^N)]^{q-1})\, d\mathbf{r}^N \qquad (1)$$

where q is a real number and $P_q(\mathbf{r}^N)$ is a generalized probability distribution at $\mathbf{r}^N$. In fact, when $q = 1.0$, the standard entropy $S = -k_B\int P(\mathbf{r}^N) \ln P(\mathbf{r}^N)\, d\mathbf{r}^N$ is recovered from eq 1. The generalized probability distribution P_q is determined by extremizing S_q in eq 1 subject to the constraints

$$\int P_q(\mathbf{r}^N)\, d\mathbf{r}^N = 1, \quad \int [P_q(\mathbf{r}^N)]^q U(\mathbf{r}^N)\, d\mathbf{r}^N = U_q \qquad (2)$$

where $U(\mathbf{r}^N)$ is the potential energy. Then the resulting expression for P_q is finally given by

$$P_q(\mathbf{r}^N) = \frac{1}{Z_q}[1 - (1-q)\beta U(\mathbf{r}^N)]^{1/(1-q)} \qquad (3)$$

In the $q = 1.0$ limit, eq 3 reproduces the Boltzmann statistics, and when $q > 1.0$, it tends to exhibit more delocalized distributions.[22] For the purpose of generating P_q with MD, a generalized effective potential was introduced by Andricioaei and Straub:

$$\bar{U}_q(\mathbf{r}^N) = \frac{q}{\beta(q-1)} \ln[1 - (1-q)\beta(U(\mathbf{r}^N) + \epsilon)] \qquad (4)$$

where $\beta = 1/k_BT$ and ϵ is an arbitrary energy shift parameter to make the transformed energy term $U(\mathbf{r}^N)$ positive definite. In order to illustrate a general trend of $\bar{U}_q$ with respect to q, effective potentials derived from a one-dimensional (1D) model potential were displayed at various q values (See Figure 2). As was shown in Figure 2, when $q = 1.0$, $\bar{U}_q$ is the same as U, and then $\bar{U}_q$ and P_q become delocalized, as q increases from 1.0. It is of note that its barrier heights are reduced. It is also important to point out that the locations of potential energy minima remain intact from the transformation in eq 4. Regarding the transformation in eq 4, in principle one can use an empirical potential energy function including both bonding and nonbonding interaction terms. However, it should be noted that including the bonding interaction term in the transformation often results in unphysically large or small bond distances and bond angles of molecules during the simulation. Therefore, in the present study, $\bar{U}_q$ in eq 4 is constructed by considering only the nonbonding

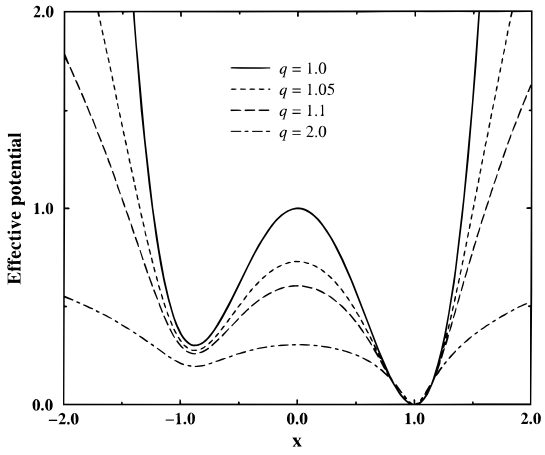

Figure 2. The generalized effective potentials $\bar{U}_q(x)$ obtained from a one-dimensional model potential $U(x)$ using eq 4. The model potential is given as follows: $U(x) = ax^4 + bx^3 + cx^2 + d$, where $a = 1.091048$, $b = -0.182095$, $c = -1.908952$, and $d = 1.0$. A value of $\beta = 20$ was employed for the effective potential.

interaction part (e.g., van der Waals and electrostatic interactions), so that the bonding interaction remains intact. Then the corresponding equation of motion is given by

$$m_i\frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r}_i}U_b(\mathbf{r}^N) - \alpha_q(\beta, \mathbf{r}^N)\cdot\nabla_{\mathbf{r}_i}U_{nb}(\mathbf{r}^N) \qquad (5)$$

where U_b and U_{nb} are the bonding and the non-bonding energy terms, respectively, and $\alpha_q(\beta, \mathbf{r}^N)$ is a scaling function defined as

$$\begin{aligned} \alpha_q(\beta, \mathbf{r}^N) &= 1; && q = 1.0 \\ &= q/[1 - (1-q)\beta(U_{nb}(\mathbf{r}^N) + \epsilon)]; && q > 1.0 \end{aligned} \qquad (6)$$

As was already mentioned in ref 22, in barrier regions where potential energy is large, $\alpha_q(\beta,\mathbf{r}^N)$ reduces the mangnitide of the forces effectively and thereby facilitates barrier crossing. In addition, during the entire MD simulations, the q value infrequently jumps from $q = 1.0$ to $q > 1.0$ for a period of τ (i.e., q jumping) and then gradually decreases to $q = 1.0$ for another period of τ' (i.e., q relaxation). The latter procedure was introduced to make a smooth transition from the transformed potential energy surface ($q > 1.0$) to the original one ($q = 1.0$). Therefore correct canonical distributions are not reproduced by the current scheme, due to the nonphysical jumping/relaxation procedure. As was briefly noted in ref 8, however, it is still possible that correct ensemble averages can be obtained by way of the fluctuating potential method.[23]

We have implemented the Tsallis transformation in eqs 4 and 5 in the CHARMM program,[24] and the CHARMM22 empirical force field[25] was used to describe the receptor part. In a similar manner to ref 7, all the necessary parameters for the ligand are generated by using QUANTA.[26] For the energy evaluation, a distance dependent dielectric constant model has been employed with the nonbonding interactions truncated at 8 Å. The X-ray crystal coordinates of streptavidin/biotin[16] and PKC/phorbol-13-acetate[20] were employed for initial setup of the complex structures, and then the hydrogen atoms were added and minimized. In this work, the ligand flexibility of both biotin

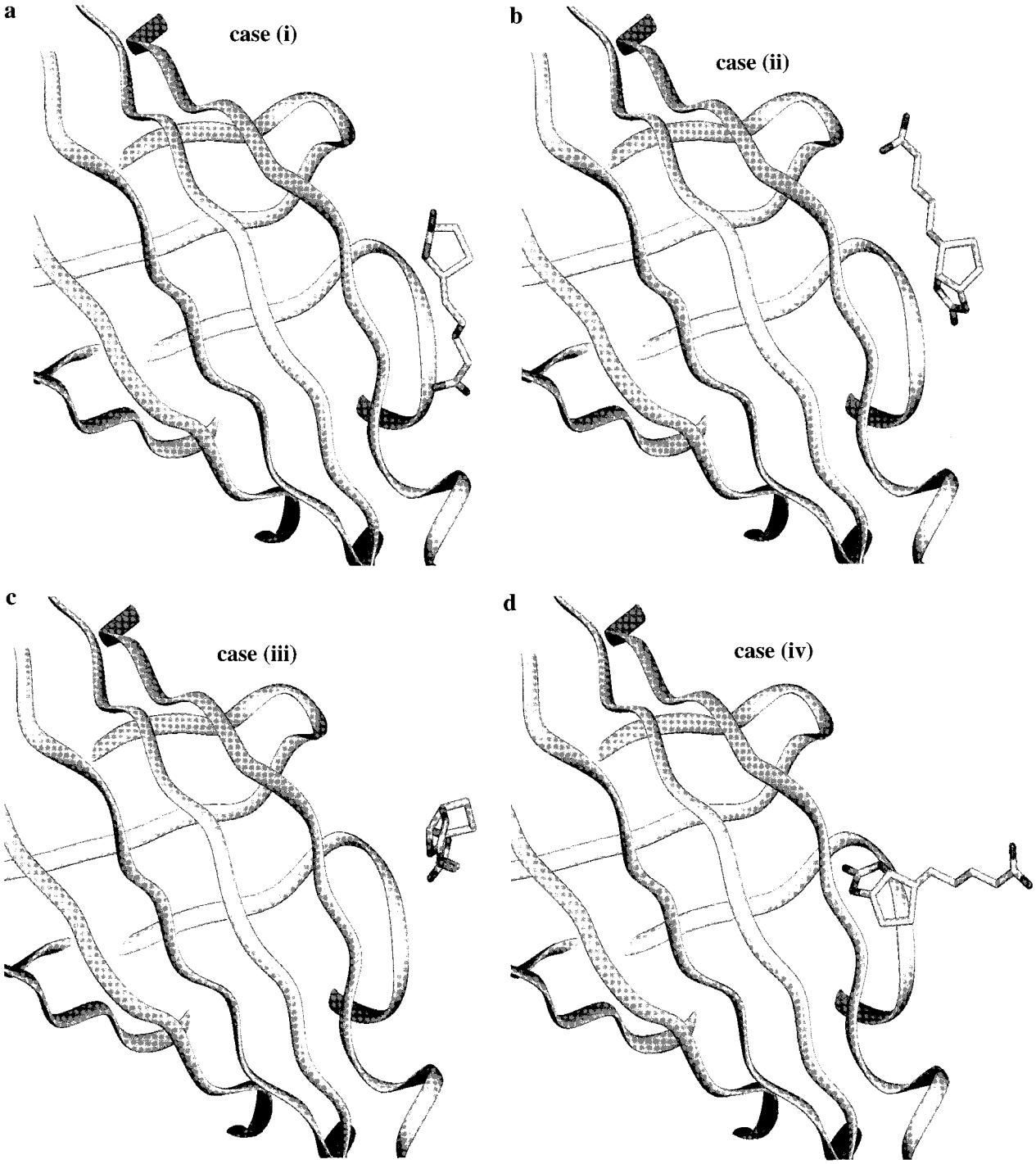

Figure 3. Initial configurations of streptavidin/biotin leading to the correct binding mode in the current docking study. A total of four initial configurations are shown in a, b, c, and d, respectively. A portion of the protein backbone in the receptor is shown by the ribbon, and the hydrogen atoms in the ligand are not shown for a clarity. (a) **case i**, (b) **case ii**, (c) **case iii**, (d) **case iv**.

and phorbol-13-acetate was fully considered. The flexibility of streptavidin was incorporated, such that all the side chains of the 22 residues located within 5 Å of biotin in the crystal structure were allowed to move and everything else in the receptor was fixed. In PKC, its binding site consists of residues 8−13 and 20−27 and the side chains of these residues were allowed to be flexible. (In this work the PKC residues corresponding to 231−281 are renumbered as 1−50.[20]) Using this constraint option, each crystal structure of the complexes with its optimized hydrogen atoms was further minimized to provide a starting structure for generating various initial configurations for the MD runs. Using QUANTA, several initial positions of each ligand were generated from the starting structures by pulling each ligand off the corresponding active site and placing it outside the binding region in several distinctive orientations. Each of the resulting configurations was then minimized with 500 steps of the steepest decent method (SD) to eliminate possible bad overlaps. In this work, the MD simulation described above were carried out for 1 − 2 ns with $\tau = \tau' = 1$ ps, using $q = 1.01$ and $\epsilon = 1500$ kcal/mol for streptavidin/biotin and $q = 1.02$ and $\epsilon = 800$ kcal/mol for PKC/phorbol-13-acetate. In case of PKC/phorbol-13-acetate, two NOE constraints consisting of the carbon atom at position 9 in phorbol-13-acetate and each α-carbon atom of residues 10 and 24 in PKC were introduced to prevent the ligands from escaping from the binding site completely. The NOE constraints,

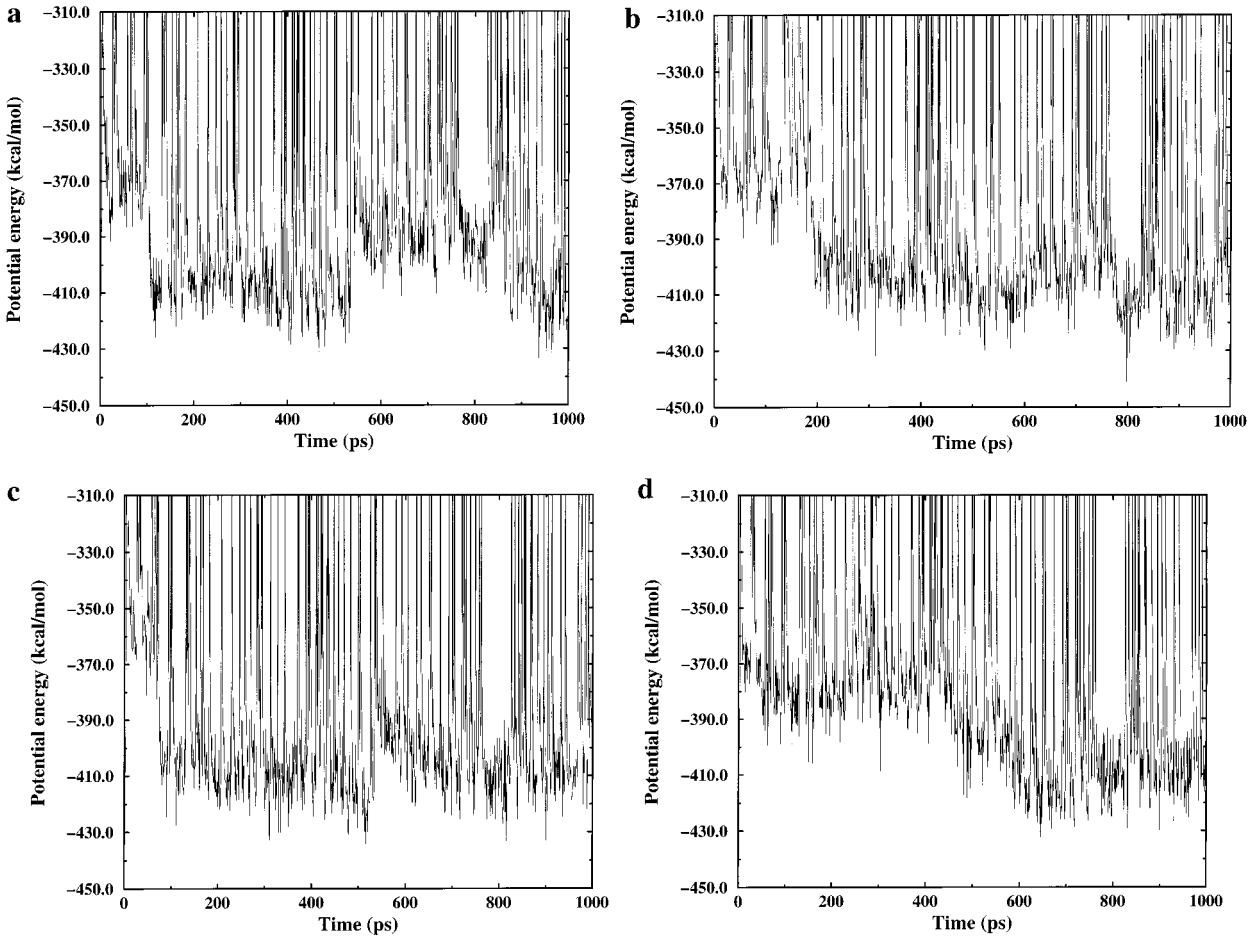

Figure 4. Potential energy values (in kcal/mol) of streptavidin/biotin during the simulation. Each value is collected in every 1 ps for the 1 ns run. The energy values beyond −310 kcal/mol in the high energy profiles are not shown here, and the maximum value in the profile is observed to reach about −75 kcal/mol. (a) **case i**, (b) **case ii**, (c) **case iii**, (d) **case iv**.

however, were applied only when either of two distances is greater than 15 Å, such that the predicted binding mode is not affected by these constraints. Currently the q-jumping frequency is employed, such that the MD run on the surface of $\bar{U}_q$ ($q > 1.0$) approximately consists of 20% of the total MD steps. In this paper, the MD simulations have been carried out in vacuum at $T = 300$ K, using the constant temperature algorithm of Berendsen et al.,[27] and the SHAKE algorithm[28] was used to fix bonds containing hydrogens with the time step of 1 fs.

III. Results and Discussion

With the new MD scheme described in section II, four docking simulations of streptavidin/biotin were carried out with different initial configurations to avoid biases. It was found that in each of the four cases (denoted as **cases i**−**iv**, respectively, in Figure 3), the correct binding mode was located in several hundreds picoseconds (ps). For each of the MD runs with the initial configurations displayed in Figure 3, the potential energy values are plotted with respect to time and given in Figure 4. As can be seen in Figure 4, two distinctive profiles of high and low energy values are observed. The high energy profile is a characteristic pattern of the q-jumping/relaxation processes and found to reach up to −75 kcal/mol during the MD runs. A typical potential energy value in the region of the correct docking mode turned out to be around −410 kcal/mol. The initial ligand position in Figure 3a (**case i**) is chosen to be in the same orientation as one in the correct binding mode, except that the ligand is outside the cavity. With this initial configuration, the correctly docked structure is quickly identified in 100 ps by the new MD scheme. After the docking occurred, the ligand tends to stay in the correct binding mode except for the period of 550−850 ps, during which it exhibits a major conformational change. It is of interest to note that in that time period, the ligand finds itself in a local minimum region by flipping its five-membered ring group from the correct binding position. The energy value of the local minimum is higher than that of the correct binding state by as much as ∼20 kcal/mol. The starting configuration in Figure 3b (**case ii**) seems challenging, since the ligand lies in an opposite orientation to the correct binding mode. As was expected, it requires somewhat longer simulation time to reach the binding mode (in 200 ps) than the previous case, since the ligand needs to make an overall rotation from the head (the ring group) to the tail (the carboxyl group) in order to correctly dock into the binding site of streptavidin. (See Figure 1 for details of the biotin structure.) The ligand orientation shown in Figure 3c (**case iii**) is an intermediate one between **case i** and **case ii**. In this case the correct complex structure is located in 100 ps, and it is found that, in the period of 550−700 ps, the tail part of the ligand moves toward the outside of the cavity region, which seems to be another local minimum (∼10 kcal/mol above the correctly docked state). In the initial configuration of Figure 3d (**case iv**), the head group of the biotin is directly pointed toward the cavity region. This choice of the initial condition leads to the correct binding mode in a relatively longer time step (600 ps). For comparison, the lowest energy docked structure in each case

TABLE 1: Root Mean-Square Deviations (RMSD) of the Initial and the Docked Structures in the Four Cases of the Initial Conditions for Streptavidin/Biotin with Respect to the Crystal Structure[a]

initial condition	RMSD of the initial structure (angstroms)		RMSD of the docked structure (angstroms)	
	ligand	receptor[a]	ligand	receptor
case i	8.57	0.30	0.62	0.47
case ii	10.73	0.30	0.51	0.45
case iii	9.84	0.30	0.58	0.44
case iv	8.04	0.31	0.55	0.41

[a] The docked structures are selected from the lowest energy conformation in each MD run of 1 ns. [b] These values result from the initial minimization procedures (See text).

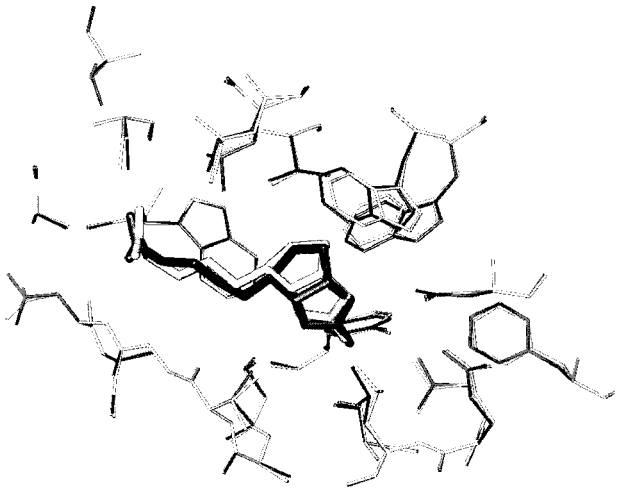

Figure 5. The predicted binding mode of streptavidin/biotin chosen from the lowest energy conformation in the MD trajectory of **case i**. The ligand molecule along with the 22 residue side chains in the correct binding mode is superimposed with the X-ray crystal one (ref 16). The crystal and the predicted structures are represented by the black and the gray colors, respectively. The ligand part is shown in boldface. No hydrogen atoms are shown.

is compared with the crystal one and the resulting values of the root mean square deviation (RMSD) are given in Table 1. All the predicted ligand positions in the correct docking state have small RMSD values of 0.5−0.6 Å, showing excellent agreement with the crystal structure. Also note that in the docked structures, the side chains of the active site exhibit some minor deviations from the positions of the crystal structure (i.e., RMSD values of 0.4−0.5 Å). As can be seen from the RMSD data in Table 1, all the predicted complex conformations are quite similar to one another. Thereby only one of the docked structures (**case i**) superimposed with the crystallographic position is shown in Figure 5.

Using the same simulation scheme mentioned above, the unique binding mode of PKC with phorbol-13-acetate was identified within 2 ns and also reproduced by several independent runs. The predicted binding mode is further refined by a minimization and is then directly compared to the X-ray crystal structure[20] (Figure 6). The resulting RMSDs of the predicted binding mode are 0.2 Å for the ligand and 0.9 Å for the receptor. As was the case of streptavidin/biotin, it is encouraging to see that in the predicted binding mode of PKC/phorbol-13-acetate, the positions of the receptor side chains are in reasonably good agreement with the experimental results.[20] It is of note that the most noticeable deviation occurs with Trp 22 mainly due to dynamic fluctuation of the side chain.

In order to investigate the effect of the q value, we increased the q value, with the other parameters unchanged. The MD test

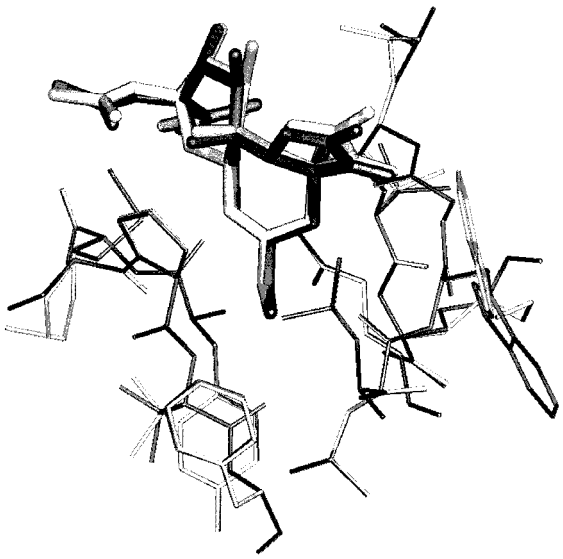

Figure 6. The predicted binding mode of PKC/phorbol-13-acetate superimposed with the X-ray crystal structure (ref 20). The ligand molecule along with the side chains corresponding to residues 8−12 and 20−27 are included here. The crystal and the predicted structures are shown by black and gray colors, respectively. The ligand part is shown in boldface. No hydrogen atoms are shown.

runs with several choices of larger q values showed that the docking event was further accelerated. If the q value is too large, however, the possibility of the ligand escaping from the binding mode increases. In general, we suggest that a reasonable starting point in choosing the q value be $(q - 1)\epsilon = 10-50$ kcal/mol. Obviously the choice of the optimum q value depends on systems and should be considered in conjunction with ϵ, and a systematic way to optimize those parameters is an interesting and practical issue for further improvement of the method.

The present results show that this new MD scheme enhances the docking process significantly, and even after the docking is completed, it still allows conformational changes of ligands and their receptors in search of nearby local minima. Thus the new method clearly demonstrates a superb capability as a new MD-based docking strategy. In particular, it should be mentioned that the same MD protocol with a purely rigid receptor model appears to be less efficient than the current case in which the partial receptor flexibility is included, even though the former requires less cpu time than the latter does for one MD time step. This implies that including the receptor flexibility at least around the active site seems to play a role in a better performance of the MD method for docking problems, probably because the receptor flexibility helps to further reduce the energy barrier when a ligand enters the active site of its receptor. One of the main advantages of the present MD scheme is that the flexibility of both receptor and ligand can be incorporated in quite a straightforward manner. The other advantage of the method is that it can easily be applied to docking problems in presence of explicit water. For this reason, this method may provide an effective conformational searching strategy for flexible ligand/flexible receptor docking problems in vacuous or water environment. Extensive applications of this method to more complicated and challenging problems, such as protein/protein docking and protein folding have now been undergoing in this group or elsewhere.

Acknowledgment. The financial support from the Department of Defense (DOD DAMD17-93-V-3018) to the Georgetown Institute for Cognitive and Computational Sciences and from the Lombardi Cancer Center is greatly appreciated.

References and Notes

(1) Goodsell, D. S.; Olson, A. J. *Proteins* **1990**, *8*, 195.
(2) Caflisch, A.; Fischer, S.; Karplus, M. *J. Comp. Chem.* **1997**, *18*, 723.
(3) Liu, M.; Wang, S. *J. Comput.-Aided Mol. Design* **1999**, *13*, 435.
(4) Di Nola, N.; Roccatano, D.; Berendsen, H. J. C. *Proteins* **1994**, *19*, 174.
(5) Oshiro, C. M.; Kuntz, I. D.; Dixon, J. S. *J. Comput.-Aided Mol. Design* **1995**, *9*, 113.
(6) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III *J. Comp. Chem.* **1998**, *19*, 1612.
(7) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L., III *J. Comp. Chem.* **1998**, *19*, 1623.
(8) Pak, Y.; Wang, S. *J. Chem. Phys.* **1999**, *111*, 4359.
(9) Tsallis, C. *J. Stat. Phys.* **1988**, *52*, 479.
(10) Curado, E. M. F.; Tsallis, C. *J. Phys. A* **1991**, *24*, L69.
(11) Ewing, T. J. A.; Kuntz, I. D. *J. Comp. Chem.* **1997**, *18*, 1175.
(12) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *207*, 727.
(13) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.* **1996**, *261*, 470.
(14) Ren, J.; Esnouf, R.; Hopkins, A.; Willcox, B.; Jones, Y.; Ross, C.; Stammer, D.; Stuart, D. *J. Med. Chem.* **1996**, *39*, 1589.
(15) Harel, M.; Schalk, L.; Ehret-Sabatier, F.; Bouet, M.; Goeldner, M.; Hirth, C.; Axelsen, P. H.; Silman, I.; Sussman, J. L. *Proc. NatI. Acad. Sci. U.S.A.* **1993**, *90*, 9031.
(16) Weber, P. C.; Ohlendorf, D. H.; Wendoloski, J. J.; Salemme, F. R. *Science* **1989**, *243*, 85.
(17) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. *Chem. Phys. Lett.* **1997**, *278*, 297.
(18) Apostolakis, J.; Pluckthun, A.; Caflish, A. *J. Comput. Chem.* **1998**, *19*, 21.
(19) Sandak, B.; Wolfson, H. J.; Nussinov, R. *Prot. Struct. Funct. Genet.* **1998**, *32*, 159.
(20) Zhang, G.; Kanzanietz, M. G.; Blumberg, P. M.; Hurley, J. M. *Cell* **1995**, *81*, 917.
(21) Andricioaei, I.; Straub, J. E. *Phys. Rev. E* **1996**, *53*, R3055.
(22) Andricioaei, I.; Straub, J. E. *J. Chem. Phys.* **1997**, *107*, 9117.
(23) Liu, Z.; Berne, B. J. *J. Chem. Phys.* **1993**, *99*, 6071.
(24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(25) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
(26) QUANTA; Molecular Simulations Inc.: San Diego, CA, 1997.
(27) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(28) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.