Journal of Computer-Aided Molecular Design, **15:** 979–996, 2001.
KLUWER/ESCOM
© 2002 *Kluwer Academic Publishers. Printed in the Netherlands.*

Structure-based ligand design for flexible proteins: Application of new F-DycoBlock

Jiang Zhu, Hao Fan, Haiyan Liu & Yunyu Shi*
School of Life Science, University of Science and Technology of China (USTC), Hefei, Anhui, 230026, P. R. China; Laboratory of Structural Biology, Chinese Academy of Science (CAS), USTC, Hefei, Anhui, 230026, P. R. China

Received 3 April 2001; accepted 11 September 2001

Key words: cyclooxygenase2, *De novo* structure-based ligand design, DycoBlock, flexible drug design, HIV protease, multiple copy

Summary

A method of structure-based ligand design – DycoBlock – has been proposed and tested by Liu et al.[1]. It was further improved by Zhu et al. and applied to design new selective inhibitors of cyclooxygenase 2 [2]. In the current work, we present a new methodology – F-DycoBlock that allows for the incorporation of receptor flexibility. During the designing procedure, both the receptor and molecular building blocks are subjected to the multiple-copy stochastic molecular dynamics (MCSMD) simulation [1], while the protein moves in the mean field of all copies. It is tested for two enzymes studied previously – cyclooxygenase 2 (COX-2) and human immunodeficiency type 1 (HIV-1) protease. To identify the applicability of F-DycoBlock, the binding protein structure was used as starting point to explore the conformational space around the bound state. This method can be easily extended to accommodate the flexibility in different degree. Four types of treatment of the receptor flexibility – all-atom restrained, backbone restrained, intramolecular hydrogen-bond restrained and active-site flexible – were tested with or without the grid approximation. Two inhibitors, SC-558 for COX-2 and L700417 for HIV-1 protease, are used in this testing study for comparison with previous results. The accuracy of recovery, binding energy, solvent accessible surface area (SASA) and positional root-mean-square (RMS) deviation are used as criteria. The results indicate that F-DycoBlock is a robust methodology for flexible drug design. It is particularly notable that the protein flexibility has been perfectly associated with each stage of drug design – search for the binding sites, dynamic assembly and optimization of candidate compounds. When all protein atoms were restrained, F-DycoBlock yielded higher accuracy of recovery than DycoBlock (100%). If backbone atoms were restrained, the same ratio of accuracy was achieved. Moreover, with the intramolecular hydrogen bonds restrained, reasonable conformational changes were observed for HIV-1 protease during the long-time MCSMD simulation and L700417 was reassembled at the active site. It makes it possible to study the receptor motion in the binding process.

Introduction

The goal of drug-design strategy is to discover and optimize novel lead compounds for therapeutic needs. The class of 'structure-based drug design' methods have been developed and applied extensively in the past decade, such as LUDY [3], HOOK [4] and CONCEPT [5]. A single, fixed receptor conformation, often from high-quality X-ray crystal complex structure, was used as the target to dock or design ligands in the majority of existing methods. This rigid conformation was believed to be energetically most favorable at the native state.

However, recent studies on the kinetics of protein folding revealed some essential concepts in terms of 'funnel' and 'energy landscape' [6–19]. In the current view, protein folding is a parallel, funnel-like process, in which an ensemble of molecules goes downhill along multiple routines on the energy land-

*To whom correspondence should be addressed. E-mail: yyshi@ustc.edu.cn

scape. Around the bottom of folding funnel, there exists a population of low-energy conformational isomers. The 'new views' of protein folding cast new light on some long-held concepts in the binding mechanism, which are often described as two models. The 'lock-and-key' model [20] corresponds to a single receptor conformation. According to it, whether in the unbound or bound state, the protein can only adopt a well-defined rigid conformation and have just one optimal complementary ligand. However, it is always observed that protein will shift from the most populated conformation to others when introducing a ligand. This is the case called 'induced fit' model [21], in which both the protein and the ligand cannot keep rigid to prevent such a 'deforming' binding process. These two frequently used terms can be explained by the current views of protein folding [22, 23]. In the 'lock-and-key' binding case, the bound conformation is just overlap with the most populated unbound conformation. Whereas in the 'induced fit' case, the ligand promotes the protein to move from one conformation – a minimum at the bottom – to another minimum separated by energy barriers.

The dynamic behavior of a receptor has been recognized as a complicated factor in both the binding mechanism and computer-aided drug design. The receptor flexibility is particularly important for drug discovery if one expects to obtain novel compounds with different binding modes or to properly predict the activity of ligands. The earliest attempts to accommodate small conformational changes were through 'soft' scoring functions [24, 25], which allows some overlap between the ligand and protein. Larger changes were allowed in docking by introducing partial degrees of freedom in the protein, such as hydrogen atoms, hydrogen-bond donor and acceptor atoms, or the whole side chain. Among these early studies, Leach first made use of a rotamer library to incorporate the side chain mobility [26]. An alternative way to allow even larger conformational changes is to generate a subensemble of receptor conformations using MC or MD simulation [27]. It is notable that the treatment of flexibility above is not related to the *de novo* structure-based drug design but only the docking study. One reason is that it is too difficult to cope with the protein flexibility and algorithms involved in drug design simultaneously. Recently, Stultz and Karplus used an extended multiple-copy simultaneous search (MCSS) methodology to allow for the protein flexibility in searching binding sites of functional groups [28, 29]. Then, they used the dynamic ligand design (DLD) algorithm to design inhibitors from the found binding sites [30]. However, the protein structure was fixed in DLD method.

DycoBlock was first proposed by Liu et al. as an approach for the *de novo* structure-based drug design [1]. The essential ideas of combinatorial chemistry were incorporated into the structure-based drug design by the combined use of MCSMD simulation and a special connecting algorithm. Molecular building blocks are dynamically assembled to compounds in the active sites of receptor. DycoBlock has been proved effective at reproducing the known ligands by breaking two inhibitors of HIV-1 protease (L700417 and pepstatin) into fragments and reconstructing them. DycoBlock was improved by Zhu et al. [2] in a recent work. Simulated annealing (SA) was used to enhance the efficiency in the search of binding sites. High-energy copies were automatically replaced. The multi-type atomic probe was used to calculate the grid potential and force. COX-2 was used in both reproducing a selective inhibitor – SC-558 – and in design of new selective inhibitors. Moreover, an analytical algorithm for solvent accessible surface area (SASA) calculation and a clustering algorithm based on the molecular similarity were used to analyze the resulting compounds. The accuracy of recovery grew to 58.8% and a high diversity was achieved for the designed ligands.

It is a big challenge to properly combine the advantages in DycoBlock – MCSMD simulation and dynamic assembly – with the receptor flexibility. We developed a new methodology – F-DycoBlock in the current work. It can be used to design new ligands with different binding modes or predict complex conformation for a flexible receptor. It can also aid us in understanding the conformational mobility of the receptor in the binding process. Compared with the advances in the past few years, there are some important differences. The essence of F-DycoBlock is stochastic dynamics (SD) simulation and GROMOS molecular mechanics force field. When dynamic assembly was combined with the receptor flexibility, it proved even more effective. In other methods, searching for binding sites and designing are often accomplished separately and have different theoretical basis. For instance, MCSS is used to search binding sites and HOOK connects them using bridge groups from a specific database while all building blocks fixed. This treatment lacks consistence in methodology and causes difficulties in the evaluation of results.

Materials and computational methods

In this section, the essential ideas in our work for flexible drug design are outlined. First, the theoretical basis is briefly described, then the general settings in F-DycoBlock are given. Different protocols in the testing study for two enzymes are described for comparison.

Theoretical Basis

In F-DycoBlock, locally enhanced sampling (LES) [31–34] is combined with the multiple-copy stochastic molecular dynamics (MCSMD) simulation and dynamic connecting algorithm [1, 2]. The LES technique, developed from the mean field theory (MFT, see an excellent review) [35], is efficient for simultaneously sampling multiple conformations of a ligand with a large receptor. The receptor moves under the average force of all copies. The copies are energetically transparent to one another and each one responds to the full force from the protein. The efficiency is achieved by computing the protein intramolecular interaction once for all the copies.

In the LES protocol, the probability density ρ of the whole system can be expressed by a product of two independent probability densities of different subsystems, namely, of the molecular building blocks (or constructed compounds) and of the receptor. It can be written as:

$$\rho(P, Q, t) \approx \rho_s(P_s, Q_s, t)\rho_{N-s}(P_{N-s}, Q_{N-s}, t), \quad (1)$$

where s designates the subsystem of our interest, and N-s denotes the rest of system – the receptor in our case. Furthermore, ρ_s can be expanded in terms of δ-functions, as

$$\rho_s(P_s, Q_s, t) = \sum_w w\delta_w[P_s - P^0_{s,w}(t), Q_s - Q^0_{s,w}(t)], \quad (2)$$

where w is determined by the initial distribution and kept constant in the LES protocol. It presents the normalization factor verifying the following relationship.

$$\sum_w w = 1. \quad (3)$$

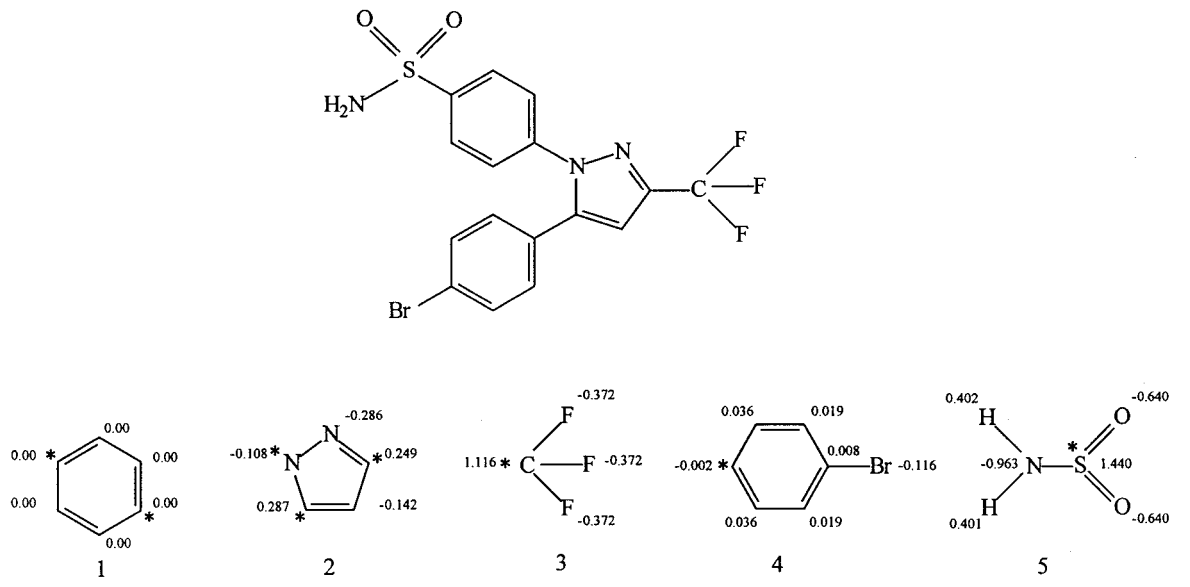
The normalization factor w is usually taken as $1/J$, where J is the number of copies for subsystem.

There are some important issues to consider prior to including the protein flexibility into the multiple copy simulation [31–36]. First, the number of copies must be suitable to make the sampling efficient. Here it was set to the same value as in previous studies. Second, the maximum conformational dispersions allowed for the copies have to be considered. The complexity stems from an inherent limitation of the multiple copy approach. Since there is only a single set of protein coordinates, the differential equation of motion for each copy is exact, while the protein trajectory is approximate due to its motion in the average field of all copies. The imbalance of the forces between the protein atoms and the copies leads to non-Newtonian dynamics. LES works best when copies are close; however, it often causes errors for the separate distribution of copies. While a broad conformational dispersion of copies is desirable, the copies must be similar in conformation to have physically meaningful averages and to prevent the protein from distortion. At the beginning of F-DycoBlock run, building block copies are put into a rectangular region with random positions and orientations, which often conflict with protein atoms. A series of strategies were used to tackle this and to obtain the protein flexibility. Third, a major assumption of LES approach is that the overall conformation of protein won't be significantly affected by the interactions with copies. Therefore, it seems still problematic to start from an unbound conformation to obtain a bounded conformation, especially when there exists a significant conformational barrier.

General setting in the F-DycoBlock simulation

The strategy of F-DycoBlock is similar to DycoBlock, including four continuous steps [1, 2].

In the step zero, five building blocks from SC-558 and four from L700417, as in previous studies, were used for two enzymes (Figure 1) [1, 2]. Single point (SP) energy calculation is performed for each constructed molecular building block with 6-31g(d) basis set using the Gaussian98 package [37]. Calculated atomic charges are assigned to every non-hydrogen atom. Polar hydrogen atoms are explicitly treated, whereas non-polar hydrogen atoms are united to the covalently bound heavy atoms. The sum of atomic charges is assigned to the united atom. Other force field parameters, such as bond lengths, bond angles and (improper) dihedral angles, are chosen from the GROMOS96 force field [38] and augmented by the *ab initio* calculation. They are specified in the molecular topology file. Then, 1000 steps conjugated gradient (CG) minimization will be performed for the building block copies. The optimized coordinates



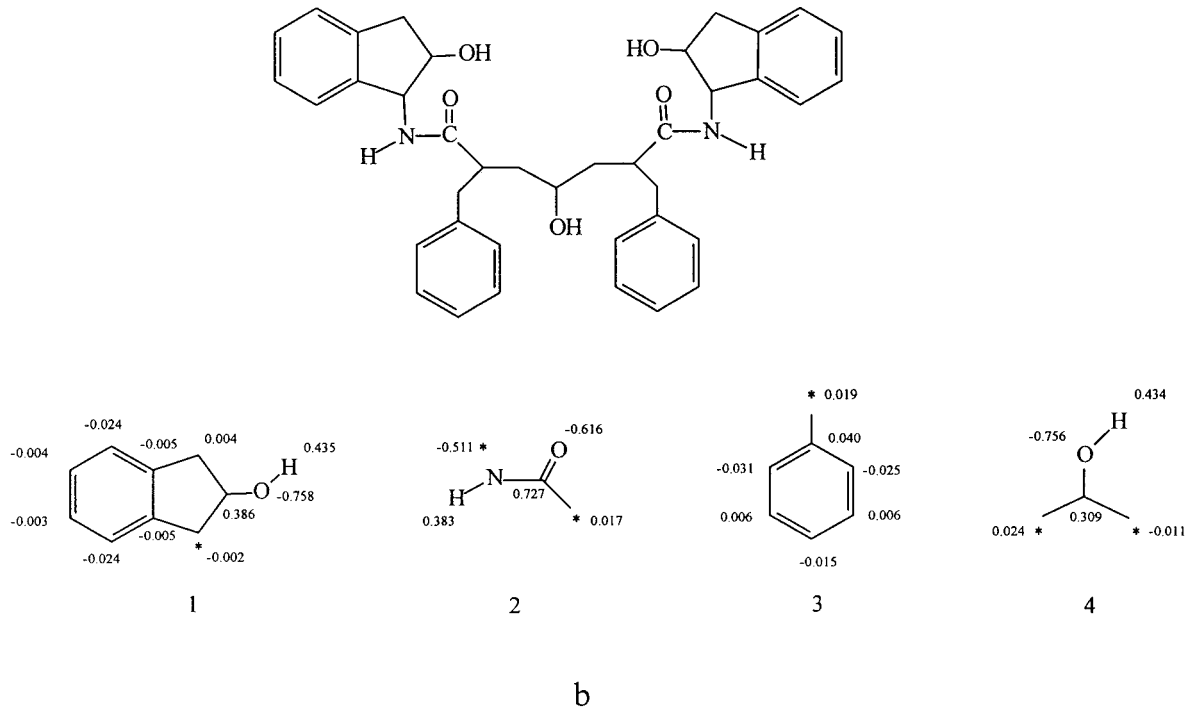


Figure 1. (a) Structure of the selective inhibitor of cyclooxygenase 2-SC-558 and building blocks from SC-558 (1–5), the link site atoms are marked with asterisks (*). Only those linkages appeared in SC-558 are allowed in connecting procedure. (b) The building blocks from L700417 (1–4), the link site atoms are marked with asterisks (*). Only those linkages appeared in L700417 are allowed in connecting procedure.

and intramolecular potential energies will be recorded when the energy threshold is reached.

In step one, a total of 10 000 steps MCSMD simulation is performed to search for the binding sites of building block copies while the protein moved in the mean field of all copies. The temperature of copies was kept 300K using Berendsen's weak coupling algorithm [39] , with $\tau_{copy} = 0.01$ ps^{-1}. The mobility of protein atoms was limited by coupling the temperature to 10 K with $\tau_{pro} = 0.02$ ps^{-1}. The temperatures of the heat bath would not change in the following two steps. Since the initial positions and orientations of all building block copies are randomly chosen, the excess kinetic energy of copies induced by bad interactions can flow to protein. To drain off this energy, the protein can be strongly coupled to a heat bath at low temperature [29]. The coupling constants in this work were chosen to ensure that each subsystem remained at the desired temperature. The results at low temperature can be compared with the recent work of Stultz and Karplus [29], in which the MCSS functionality maps of two types of representative groups – methanol (polar) and methyl (apolar) – were constructed using a similar strategy as the step one in our case. A thermal bath at 300 K was only tested for subprotocol one (see below) of COX-2, however it was not adopted in the following studies.

The time step is 1 fs and SHAKE algorithm [40] is used to constrain the bond lengths of both the protein and building blocks with a tolerance of 10^{-4}. The friction coefficient is 80 ps^{-1} for all atoms of the building blocks but it was proportional to the atomic exposure for the protein atoms and was measured by a simple contact algorithm [41]. The number of copies is 200 for each type of building block. All building block copies are put into the region containing active site with random positions and orientations. After 8000 steps, the copies with low interaction energy are clustered and recorded as seeds every 100 steps. The coordinate trajectories are recorded every 50 steps.

In step two, the MCSMD simulation is continued, in which the connecting algorithm is used to dynamically assemble the building blocks to compounds while the protein still moves in the mean field of all copies. The assembly stage consists of a number of constructing and optimizing units (CO unit). The possible linkages between building blocks will be explored and constructed in a single MCSMD step and then optimized during the remaining steps of CO unit.

In the last stage, the candidate compounds together with the protein were optimized during a long-time MCSMD simulation. In the optimization, protein atoms were restrained to their initial positions in the first 1,000 MCSMD steps. It is based on the consideration that if the atomic velocities for each compound are chosen via a random number from Boltzmann distribution, the compounds possibly conflict with the protein atoms and lead to the distortion of protein structure. An argument is from the backbone positional RMS deviation of the protein with or without this treatment (see Results). The compounds are ordered according to their binding energies at the end of simulation, while the free and binding SASAs of each compound are also calculated. The snapshots of the protein and building block copies are stored to disk every 50 steps.

Cyclooxygenase 2 (COX-2)

Cyclooxygenase 2 (COX-2), induced by cytokines, mitogens and endotoxins in inflammatory cells, was responsible for the elevated production of prostaglandins (PGs). The side effects of many anti-inflammatory drugs arise from their inhibition of both COX-2 and COX-1 (another isoform of COX) [42, 43, 44]. Therefore, design of selective inhibitors of COX-2 is important for discovering safer anti-inflammatory drugs. An extensive study on designing selective inhibitors for COX-2 has been carried out and was reported recently [2].

The structure of COX-2 complexed with SC-558 was taken from the Protein Data Bank (PDB id 1cx2). This enzyme consists of 552 amino acid residues, is definitely a large protein for any design strategy. Five types of molecular building block were used in this study (Figure 1). First, all the hydrogen atoms in the PDB structure were deleted. And then, the hydrogen atoms explicitly treated were regenerated by PROGCH in the GROMOS96 package [38]. The topology was constructed by PROGMT and the GROMOS96 vacuum simulation force field 43b1 [38] was used. The termini and charged side chains were neutralized to mimic the electrostatic screening effect of the solvent. A total of 200 steps steepest descent minimization was performed to relax the positions of receptor atoms.

Considering the size of COX-2, it is a heavy burden to include all atomic degrees of freedom in the calculation. We thereby used an improved grid approximation to deal with such cases. In DycoBlock, the effect of protein atoms beyond a simulation box was simply assigned to the grid points covering the box [1, 2]. In F-DycoBlock, the protein is decomposed into

Table 1. A total of 63 residues of COX-2 set flexible in subprotocol two and three

Residues in the active region box					
No.	Type	Choosing rules	No.	Type	Choosing rules
58	HIS	backbone and side chain	353	LEU	side chain
59	TYR	backbone	354	TYR	C_α and side chain
61	LEU	side chain	356	TRP	backbone and side chain
62	THR	side chain	358	PRO	backbone and C_β of the side chain
82	MET	side chain	359	LEU	backbone and side chain
85	VAL	backbone and side chain	403	VAL	backbone and side chain
88	SER	O_γ and H_γ of the side chain	404	ALA	backbone and side chain
89	ARG	backbone and side chain	439	PHE	side chain
92	LEU	side chain	444	TYR	OH, HH of the side chain
159	ASP	side chain	476	LEU	O of the backbone and
161	GLN	backbone and side chain			$C_\delta 2$ of the side chain
162	GLY	C of the backbone	479	GLU	side chain
163	SER	backbone and side chain	482	ARG	N, H, C_α of the backbone and
167	PHE	side chain			side chain
174	PHE	side chain	485	ALA	backbone and side chain
175	THR	C_β, $C_\gamma 2$ of the side chain	486	ILE	backbone and side chain
178	PHE	side chain	487	PHE	backbone and side chain
179	PHE	side chain	488	GLY	backbone and side chain
311	LYS	C, O of the backbone	489	GLU	backbone and side chain
313	VAL	backbone and side chain	490	THR	backbone and side chain
314	ILE	backbone and side chain	491	MET	backbone and side chain
315	GLU	backbone and C_β of the side chain	492	VAL	backbone and side chain
316	ASP	backbone	493	GLU	backbone and side chain
317	TYR	backbone and side chain	494	LEU	backbone and side chain
318	VAL	backbone and side chain	495	GLY	backbone and side chain
319	GLN	backbone and side chain	496	ALA	backbone and side chain
320	HIS	backbone and side chain	497	PRO	backbone and side chain
321	LEU	backbone and side chain	498	PHE	backbone and side chain
322	SER	backbone and side chain	499	SER	backbone and side chain
323	GLY	backbone and side chain	500	LEU	N, H of the backbone and
324	TYR	backbone and side chain			side chain
328	LEU	side chain	503	LEU	$C_\delta 2$ of the side chain
350	PHE	side chain	550	ASN	side chain

three regions – active-site region, buffer region and outside region. The active-site region was presented by a rectangular box, which contains all residues of the active site. In our case, there were a total of 63 residues inside this region (Table 1). A buffer region, which often spans at least two or three bond-lengths along each coordinate axis, was constructed. Protein atoms in this region will be harmonically restrained during the whole design procedure. A total of 1788 protein atoms were explicitly considered within these two regions. The atoms beyond the buffer region are treated as in DycoBlock, namely, the electrostatic and van der Waals interactions were preliminarily evaluated by multi-type atomic probe and assigned to the grid points covering the active-site region. The revised PROCMT in the GROMOS96 package [38] was used to generate the topology of the protein in the active-site box. A $2.2 \times 2.0 \times 2.5$ (nm^3) box was specified and a thickness of 0.5 nm was used for the buffer region.

The protocol consists of three subprotocols to present different strategies for introducing the protein flexibility. In subprotocol 1, protein atoms in the active-site and buffer regions are positionally re-

strained. In subprotocol 2, protein atoms in the active-site region move in the mean field of all copies during the connecting stage, while atoms in the buffer region are restrained. In subprotocol 3, the mobility of protein atoms in the active-site region was introduced in the first step to search binding sites, while protein atoms in the buffer region are restrained.

Subprotocol 1

The step zero has been described. In step one, the building block copies were put into a rectangular region ($2.0 \times 2.0 \times 2.0\ nm^3$). It centered at the geometric center of the C_α atoms of twelve residues at active site [2]. The energy cut-off was set to 2000 $kJ \cdot mol^{-1}$ for replacing high-energy copies every 100 steps. All explicitly treated protein atoms (1788 atoms) were restrained to the energy-minimized positions with a constant of 2.5×10^4 $kJ \cdot mol^{-1}\ nm^{-2}$. This subprotocol corresponds to a limited conformational mobility and can be compared with the work of Zheng et al. [36]. In step two, the MCSMD simulation was continued for 2500 steps. The maximal number of building blocks in a compound was set to 6. During the whole connecting procedure, the 1788 atoms were harmonically restrained to the last conformation of step one, which led to the better results than restrained to the crystal structure. In step three, a 10 ps MCSMD simulation was carried out to optimize the candidate compounds and protein. In this stage, the explicitly treated protein atoms were still restrained to the last conformation of step two.

Subprotocol 2

The protein conformation and seeds used to assemble compounds were taken from the subprotocol 1, while the protein flexibility was incorporated to the connecting stage in this subprotocol. A total of 7500 steps MCSMD simulation was performed in step two. The restraints on the residues involved in the active site decreased linearly to zero in the first 2000 steps. The 63 active-site residues (461 atoms) were set flexible as listed in the Table 1. During the following 5.5 ps MCSMD simulation, these residues responded and moved in the mean field of 1000 molecular building block copies. During the 10 ps optimization, both the 63 residues and candidate compounds were subjected to the dynamics simulation. Due to that all copies in this step have the same structure as SC-558, this procedure may be considered as 'multiple-copy flexible docking'.

Subprotocol 3

The protein flexibility was included from step one. A total of 15 000 steps MCSMD simulation was performed to search binding sites of building block copies. The 63 residues in the active site were set free gradually from 10 000 to 12 000 steps. The building blocks with low interaction energy during different stages – rigid, relaxing and flexible stage – were clustered and recorded as seeds for the assembly. The difference between subprotocol 2 and 3 is the time when the protein flexibility was introduced into the design. Recorded seeds were significantly increased by the protein flexibility, which led to a higher diversity of resulting compounds. In step two, 7500 steps MCSMD simulation was performed with the active site flexible during the whole connecting procedure. It should be noted that both recorded seeds and other copies were used in the assembly previously but only seeds in this subprotocol.

Human immunodeficiency type 1 (HIV-1) protease

HIV-1 protease has been identified as a successful target for the treatment of HIV-1 infection. When the enzyme is inhibited or genetically altered, non-infectious virions are produced. The enzyme has been used in the early test of DycoBlock [1]. The structures of this enzyme complexed with different inhibitors were known clearly and the conformational transfer between bound and unbound states can be observed for it is mainly composed of β-sheets and loops. Therefore, it is a good and critical test system for the F-DycoBlock. The structure of HIV-1 protease complexed with L700417 was downloaded from Protein Data Bank (PDB id 4phv). This enzyme, a dimmer with 99 residues for each one, is a medium-size protein for drug design. Four kinds of molecular building block as in Liu's [1] work were used here. The topology of each monomer was generated by PROGMT and the dimmer topology was constructed by PROMMT [38]. GROMOS96 43b1 force field [38] was used with the termini and charged side chains neutralized. Then a total of 200 steps steepest descent minimization followed.

There are three subprotocols as following. In subprotocol 1, all protein atoms are positionally restrained during the whole procedure. In subprotocol 2, the rotameric conformational space was explored during the connecting and optimizing stages, while heavy atoms of the backbone were restrained. In subprotocol 3, the whole protein was flexible during the design procedure, with hydrogen bonds in secondary structures

weakly restrained. The major difference between the strategies of two enzymes is that the grid approximation was not used for HIV-1 protease due to its appropriate size. As expected, the exact calculation of all protein atoms led to the same accuracy in both the all-atom and backbone restrained cases. It is surprising that when the hydrogen bonds in secondary structures were restrained during the design procedure, the accuracy of recovery is 100%.

Subprotocol 1

In step one, the building block copies were put into a rectangular region, which was consistent with Liu's work and centered at the geometric center of the four carboxyl oxygen atoms of the residues ASP25 and ASP25' [1]. The energy cut-off was set to 1000 kJ·mol^{-1} for replacing high-energy blocks every 100 steps. All atoms of the protein, a total of 1902 atoms, specified in the control file, were restrained to the energy-minimized positions with a constant of 2.5×10^4 kJ·mol^{-1} nm^{-2}. In step two, 16 000 steps MCSMD simulation was performed, which consisted of 80 constructing and optimizing units (CO unit, 200 steps per unit). Different possible combinations of building blocks were explored and building blocks were dynamically assembled to compounds by the connecting algorithm. The maximum number of building blocks in a compound was set to 8. During the whole connecting procedure, the 1902 atoms were harmonically restrained to the last conformation of step one. In step three, 10 ps MCSMD simulation was carried out to optimize the resulting compounds. Binding energy, free and binding SASAs of each compound were calculated for comparison. In this stage, the protein atoms were restrained to the last conformation of the previous step. The trajectories of the protein and building blocks were stored every 50 steps.

Subprotocol 2

The only difference between this subprotocol and the previous one is in step two. To introduce more flexibility of the protein into the design, step two were decomposed into several stages. During the first 4 ps MCSMD simulation, the protein was still restrained to the last conformation of step one. Then the restraints on all side-chain atoms were gradually relaxed in the following 4 ps. In the last 8 ps, the side chains were fully flexible and could adopt any reasonable rotameric conformation under the mean effect of copies. According to the connecting algorithm, the operation of breaking all linkages among building blocks was repeated every 200 steps, only those energetically favorable linkages remained. It is interesting to compare the results from these two subprotocols, which indicates whether the side chain mobility will introduce additional errors into design except those introduced by mean field approximation.

Subprotocol 3

It is assumed that the overall protein conformation will be perturbed by only ligand-protein interactions in the multiple copy simulation. On the other hand, it is expected to observe the conformational change in the binding process. A trade-off is to restrain the fundamental structure elements to prevent the protein from distortion while the relative motions between them are allowed. These motions play key roles in the binding mechanism known as the 'induced fit' model [22]. The program PROCHB in GROMOS96 package [38] was used to find the hydrogen bonds from the crystal structure. There are a total of 143 hydrogen bonds. Since the strength of a single hydrogen bond is approximately 5 kcal mol^{-1} while that of a covalent bond involving hydrogen atom is 50–100 kcal mol^{-1}, we used a force constant of 50 000 kJ·mol^{-1} nm^{-1} to present the weak restraints on the distance between hydrogen-bonding donors and receptors. In the step two, only the recorded seeds were used to assemble compounds during a total of 16 000 steps simulation. In step three, 10 000 steps simulation was performed to optimize the resulting compounds with the hydrogen bonds restrained.

Results

The major difference between F-DycoBlock and DycoBlock is the flexibility. DycoBlock has been proved effective in both recovery and design of ligands for two proteins. Therefore, we must first ensure that the flexibility will not lower the accuracy prior to a practical design using F-DycoBlock. Based on this consideration, the accuracy of recovery was chosen as a major criterion in the estimation. Furthermore, the advantages, drawbacks and possible reasons are described and discussed in this section.

Cyclooxygenase 2 (COX-2)

In subprotocol 1, if the temperature of protein was kept at 10 K, there were six recovered SC-558 among ten

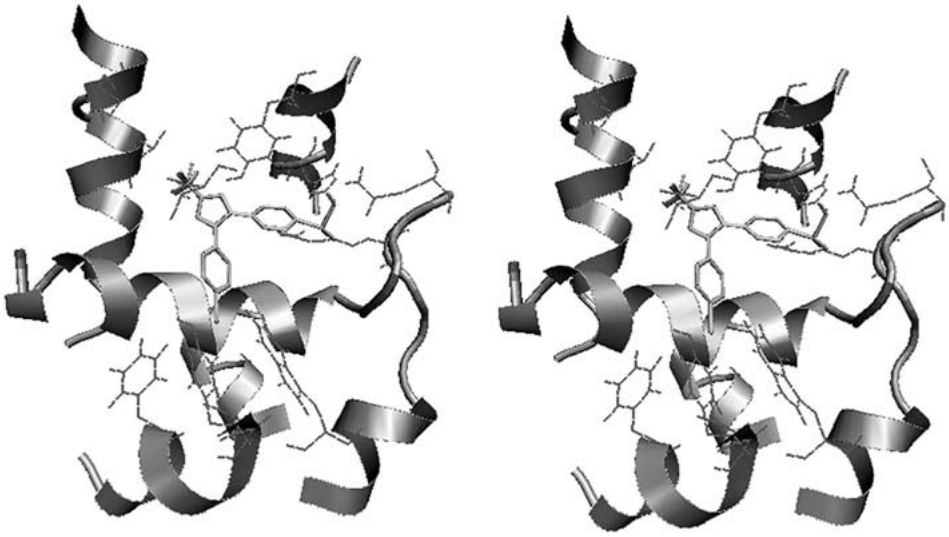


Figure 2. (a) The superimposition of six recovered ligands from a testing run of subprotocol 1, in this simulation the COX-2 was coupled to a heat bath at 300 K and all protein atoms were restrained. (b) Obtained from (a) by a five-degree rotation.

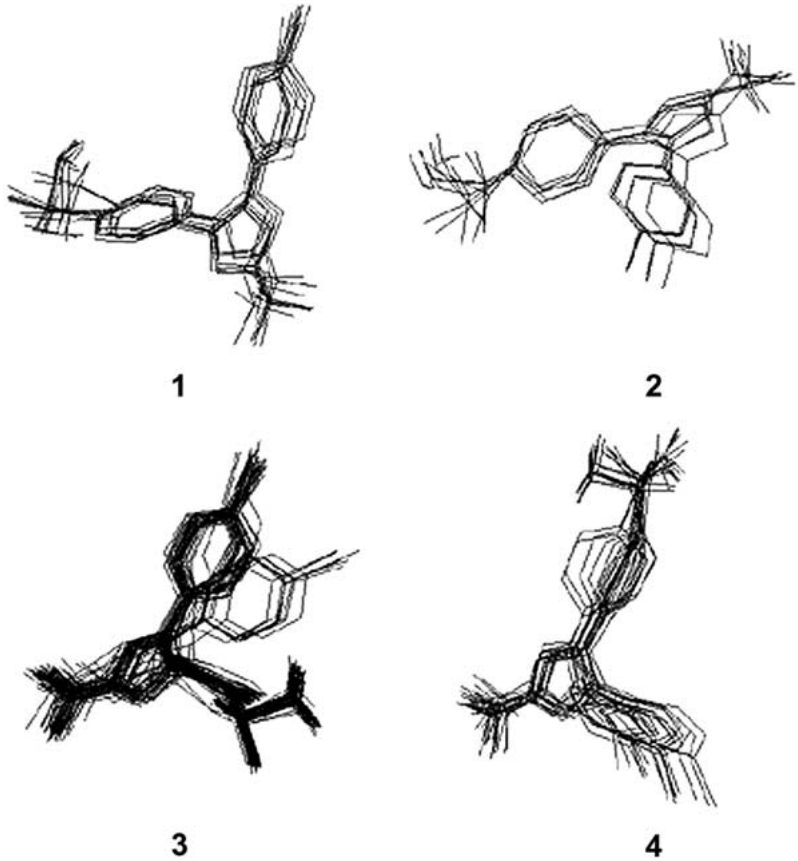


Figure 3. Four possible binding modes presented by four clusters from subprotocol 3 of COX-2. 63 residues in the active region were flexible from the search of binding sites. A clustering analysis was performed for the resultant compounds using a cutoff of 0.2. It shows enhanced diversity.

Table 2. Generated compounds ordered with respect to binding energies

		Explicit+grid (kJ/mol)			SASA (Å^2)	
	E_{tot}	E_{bind}	E_{ele}	E_{lj}	Bind	Free
(1) Results from a testing run of subprotocol 1, in which COX-2 was kept at 300 K and positionally restrained.						
1[a]	482.63	−211.32	−91.29	−120.03	9.99	315.08
2[a]	460.13	−208.81	−86.22	−122.60	10.13	315.12
3[a]	456.31	−208.16	−84.68	−123.48	10.14	315.12
4[a]	469.37	−206.88	−92.38	−114.50	10.64	315.24
5[a]	540.03	−204.51	−86.37	−118.14	9.08	313.77
6[a]	466.94	−195.22	−74.59	−120.63	10.13	313.88
(2) Results from a testing run of subprotocol 1, in which COX-2 was kept at 10K and positionally restrained.						
1[a]	560.40	−243.75	−75.13	−168.61	9.01	313.41
2[a]	528.86	−238.08	−78.96	−159.12	9.73	314.76
3[a]	472.39	−237.41	−79.45	−157.96	9.50	314.48
4[a]	495.44	−218.45	−46.23	−172.22	8.94	312.11
5[a]	552.01	−192.71	−31.44	−161.27	9.53	313.06
6[a]	682.04	−191.77	−25.38	−166.38	8.76	313.05
7	361.36	−172.70	−84.15	− 88.56	14.67	313.78
8	583.96	−141.20	−15.10	−126.10	10.54	314.40
9	610.03	−139.16	− 9.21	−129.95	9.94	313.24
10	541.38	−115.42	−19.86	− 95.56	12.01	311.97

[a]Presents recovered inhibitors – SC-558, other resultant compounds are also listed.

resulting compounds. The accuracy of recovery was 60% and they rank top six when ordered according to the binding energy. If the protein was coupled to the thermal bath at 300 K, all six resulting compounds all are recovered SC-558. The accuracy of recovery even reached 100% while it was 12.5% and 58.8% for the earliest and revised DycoBlock, respectively [1, 2]. Therefore, the results from 300 K are of our great interest. The superimposition of six recovered ligands is shown in Figure 2. The positional RMS deviations between them ranged from 0.14 to 0.24 Å while ranged from 0.76 to 0.85 Å with respect to the SC-558 in crystal structure. The results from different temperatures are shown in Table 2. F-DycoBlock shows two advantages at both 10 K and 300 K. First, it often yields higher accuracy of recovery, especially when the protein was kept at high temperature. Small RMS deviations suggest F-DycoBlock has high accuracy in assembling known ligands in active site. Second, all recovered inhibitors have lower binding energies than other resulting compounds. A similar phenomenon was also observed in the work of Zheng et al. [36]. The possible explanation is that the unfavorable interactions between the protein and randomly located building blocks were removed by the fluctuation of protein atoms. It led to a sampling on a smoothed effective energy surface and resulted in fewer but correct seeds for assembly. The temperature of 300 K induced larger fluctuation of protein atoms, which is equal to using a weak restraining constant. Clearly, F-DycoBlock provides a better way to design or dock ligands for a specific receptor conformation instead of DycoBlock.

In subprotocol 2 and 3, the combined use of grid approximation and protein flexibility was carefully investigated. They correspond to taking the flexibility into account at different stage. In subprotocol 2, there are 47 resulting compounds and 5 of them were recovered SC-558. The recovered ligands were not observed in the first 5 compounds when ranking according to the binding energies. In subprotocol 3, great care was devoted to keeping the protein structure. A total of 325 seeds were used in the assembly. There are 7 recovered SC-558 among 76 resulting compounds. The active-site pockets were correctly occupied by the recovered ligands. The resulting compounds fell into

Table 3. Clustering results of constructed compounds in recovery study

Sequence number of clusters	Sequence number of compounds in each cluster
1	0, 15, 31, 57, 52, 9, 70, 45, 64
2[a]	1, 53, 61, 60, 19, 5, 43[b], 25[b], 12
	3 2, 16, 76, 10, 28, 24, 56, 3, 21, 59
	63, 18, 23, 26, 47, 8, 71, 49, 75, 29
	32, 35, 50, 58, 62, 13, 37, 22, 46, 14
	65, 30, 17, 41, 69, 36, 74, 44, 20, 72
4[c]	4, 42, 38, 67, 48, 40, 55, 54, 34
5[c]	6, 39, 33, 11, 73, 77

Recovered ligands from the second cluster ordered with respect to binding energies

		Explicit+grid(kJ/mol)			SASA(Å^2)	
	E_{tot}	E_{bind}	E_{ele}	E_{lj}	Bind	Free
1	276.30	−193.42	−69.40	−124.02	104.90	313.35
2	359.46	−161.07	−56.44	−104.63	113.05	313.35
3	475.49	−155.76	−52.12	−103.64	108.18	312.87
4	315.53	−150.09	−45.69	−104.42	106.18	313.62
5	448.25	−149.63	−43.18	−106.44	109.22	313.39
6	470.08	−142.47	−52.68	−89.79	112.56	313.47
7	450.80	−138.04	−48.34	−89.69	115.60	313.58

[a]This cluster represents the set of recovered ligands from subprotocol 3 of COX-2. 63 residues in the active region were set flexible from the search of binding sites.
[b]Generated compounds which are not counted for high binding energy larger than 10^3.
[c]Compounds in the 4 and 5 clusters have similar binding modes and structural motif. In the corresponding Figure 3, they are shown together.
The clustering cutoff is set at 0.2.

4 clusters (see Table 3) when using a cutoff value of 0.2. They represent possible binding modes (see Figure 3), showing enhanced diversity. The protein structure is kept well during the MCSMD simulation (shown in Figure 4). It is confirmed by the backbone positional RMS deviation plots (shown in Figure 5). The maximal value of RMS deviation is not larger than 0.11 nm and a gradual decrease was observed in the optimization. However, the structure of active site was still partially lost after optimization. There were a total of 9 residues in two helices disappeared from the visualization of secondary structures. It indicated that additional errors together with the protein flexibility were introduced into drug design by the grid approximation. The backbone RMS deviation of protein fluctuated between 0.08 nm and 0.11 nm in the optimization, while it grew to 0.12 nm without the positional restraints in the first 1000 steps simulation.

There are two major sources of errors in calculation – the grid and LES approximations. To speed up the calculation, the grid approximation was still adopted for COX-2. Therefore, the possible explanation for the structural loss was that the protein could not keep stable around the initial conformation under the associate effect of two approximations. It came mainly from two aspects. First, the effect of protein atoms beyond the buffer region was represented by the grid interactions. Due to the linear extrapolation used in the grid approximation instead of the exponential form in the Coulomb and Lennard-Jones formula [1, 2], the mean potential around a protein atom was significantly smoothed. Therefore, the explicitly treated protein atoms could not be trapped in the neighborhood of initial positions. Furthermore, a comparison between the results from three subprotocols indicates that the grid approximation only affects the protein stability but contributes little to the interaction between building blocks and protein. Second, a number of building block copies were randomly put into the active-site region [1, 2]. In the LES approximation, the weight factor w is determined by the initial distribution of copies, namely, the number of copies, their positions and orientations. As pointed out by Stultz and Karplus in a recent work [29], the error is a consequence of the renormalized potential function employed in the LES approximation. Certainly,

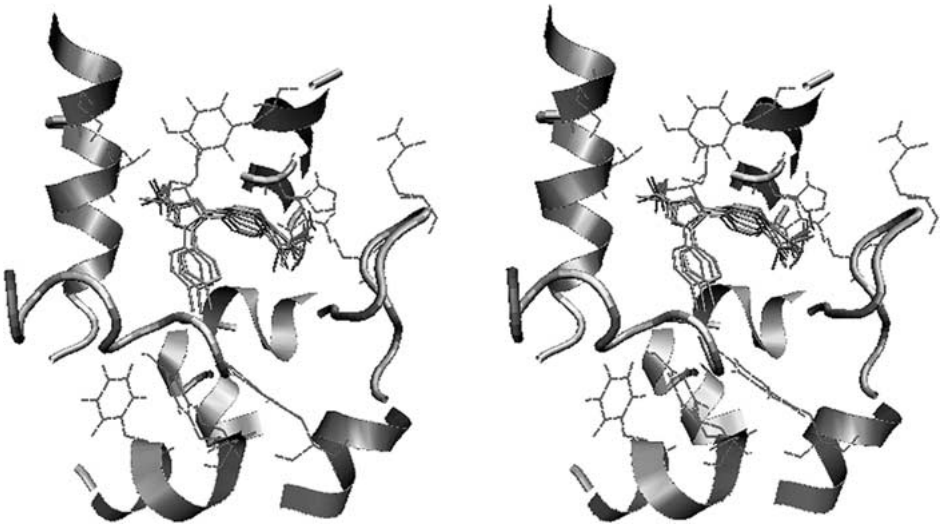


Figure 4. (a) The superimposition of seven recovered ligands is shown as the result of the subprotocol 3 of COX-2. (b) Obtained from (a) by a five-degree rotation.

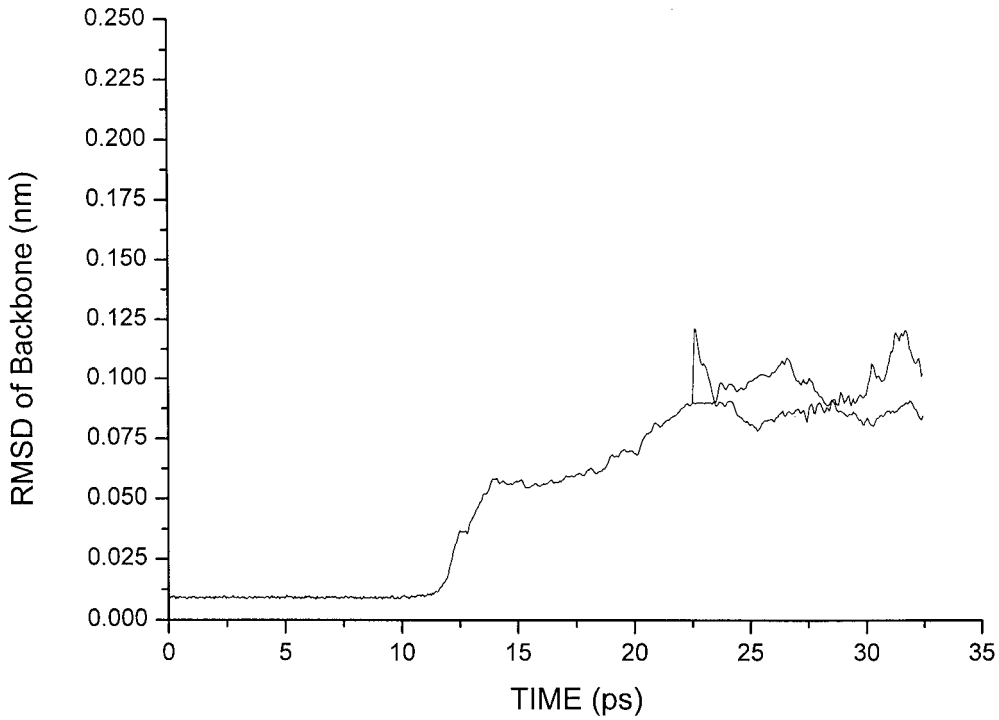


Figure 5. Backbone positional RMS deviation plots of the protein in the active site and buffer region are shown. The coordinates trajectory in the whole designing procedure of subprotocol 3 was compared with respect to the initial conformation. The black line presents the simulation with 1000 steps positional restraints at the beginning of optimization. The gray line presents the simulation without this treatment.

the high temperature of protein atoms will amplify the errors induced by these approximations. This is the reason why the protein was coupled to 10 K with a strong coupling constant in our strategy.

Human immunodeficiency type 1 (HIV-1) protease

The protocol for HIV-1 protease consists of three subprotocols. The grid approximation that induced the artificial instability of COX-2 was not adopted here. The ligand – L700417 – is a complicated compound made up of seven building blocks and is characterized by central symmetry. It is a strict test to reassemble L700417 at the active site of the flexible enzyme using F-DycoBlock.

Subprotocol 1 is a reference simulation with all protein atoms positionally restrained, which presents a limited distribution around the bounded conformation. There are 6 resulting compounds with the same building blocks as L700417 and two of them are recovered ligands. It is notable that they can superimpose well with L700417 in the crystal structure (shown in Figure 6). In Table 4, the compounds were ordered according to their binding energies, other energy terms and the binding and free SASAs also were listed.

In subprotocol 2, the side chain flexibility was tested. The sequence number of backbone atoms and the restraining constant were specified in the control file. As mentioned by Stultz and Karplus in the recent work, local optimization of existent MCSS minima – binding sites in our case – with a flexible protein results in lower energy minima in regions of particular interest [29]. We adopted a similar strategy here. In step one, the protein was restrained to the crystal structure, but in step two the restraints on the side-chain atoms decreased to zero from 4000 to 8000 steps during the 16 000 steps simulation. During this procedure, the building blocks were further optimized for assembly and possible linkages would be constructed while the protein sampled the rotameric conformational space. There was one recovered L700417 (shown in Figure 6) among three candidate compounds. The results were listed in the Table 4. It is not surprising that the recovered ligand has a much lower binding SASA than other two ones. The accuracy of recovery is the same as in the previous subprotocol. This result indicates that with the exact calculation of all atomic degrees of freedom, the flexibility will not introduce additional errors into the designing.

Subprotocol 3 was of our particular interest because its strategy is well consistent with the essence of current view of binding mechanism. The whole protein was set flexible in the assembly and optimization stages, except that the hydrogen bonds in the secondary structures were restrained. The result (shown in Table 4) seems encouraging.

First, under the mean effect of all seeds or candidate compounds, the protein structure was kept stable during the assembly and optimization. Structural distortion was not observed in the designing. Since copies interact with the protein exactly, the correct protein structure will ensure assembled compounds and the last complex structure are reasonable. Furthermore, it is also expected to observe the conformational transfer between the minima at the bottom of folding funnel. This process may take a long simulation time from hundreds of picoseconds to nanoseconds. And generally, the destruction or formation of secondary structures is not involved in the binding process. Therefore, the stability of protein structure under the effect of LES and other approximations must be first identified.

Second, it yielded obviously better results than all-atom and backbone restrained cases, the accuracy of recovery reached 100% for three assembled compounds (Figure 6). They all have similar binding modes to that in the crystal structure. This indicates that in our methodology, with more degrees of freedom of the protein relaxed, the accuracy of design will not be lowered but enhanced. Third, a reliable conformational change was observed for the last protein conformation. PROCOX in GROMOS96 package [38] was used to calculate the positional RMS deviation of the resulting conformation and also other five bounded conformations with respect to a well-defined reference conformation (Table 5). The protein was extracted from the complex structure (with L700417) and subjected to the 150 ps equilibration and 100 ps molecular dynamics simulation with explicit SPC [45] waters. The last recorded structure was treated as the reference point. It was observed that the last conformation from F-DycoBlock shares the similar conformational mobility with other structures, especially for residues surrounding the active site (Figure 7). All residues with larger C_α deviation than the average were sought out and are shown in the Figure. It is interesting that one sheet in the MCSMD-relaxed structure changed more than the corresponding sheet in another monomer. The possible explanation is the secondary structures adjusted their arrangement to adopt the energetically more favorable complex conformations. In this sense, it is a feasible way to void the errors induced by MFT

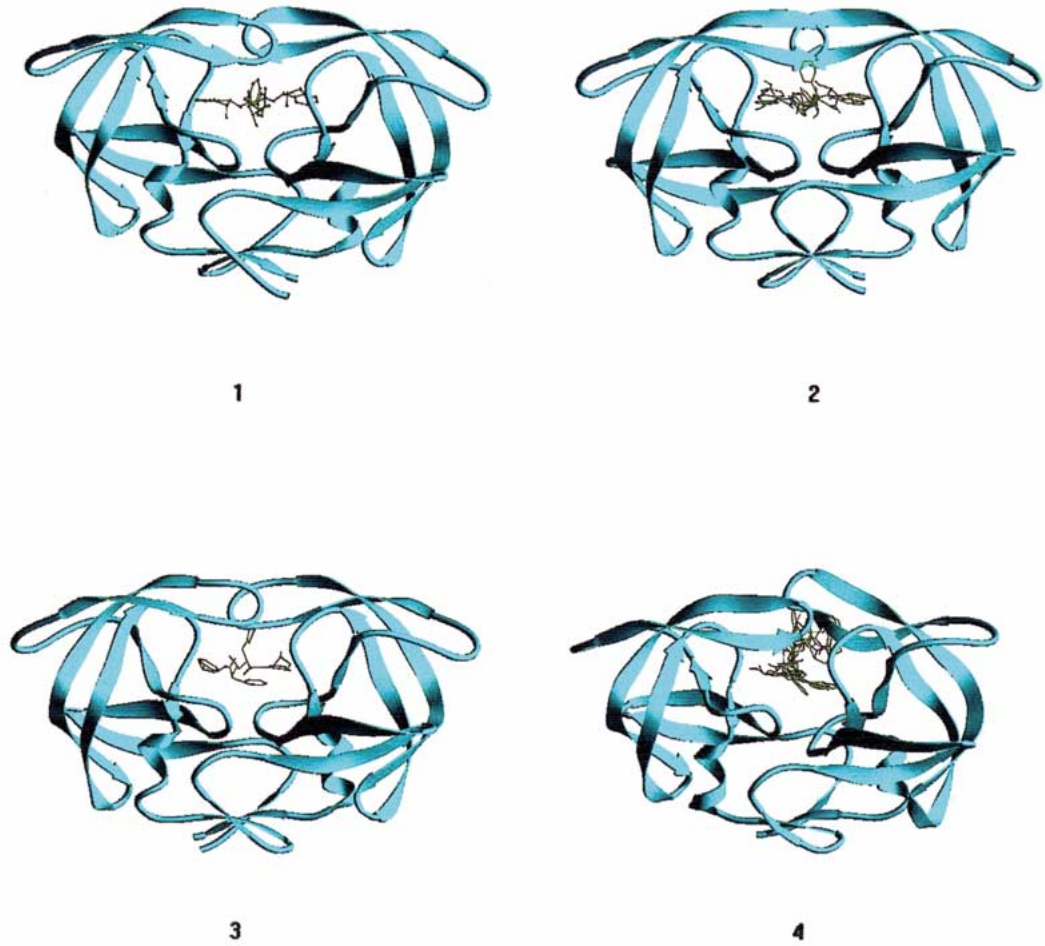


Figure 6. Complex conformations of HIV-1 protease from crystal structure and simulation results. (1) HIV-1 protease in the crystal structure complexed with L700417. (2) The final complex conformation of protein and two reassembled L700417s, all the protein atoms were positionally restrained in the simulation. The accuracy of recovery is ~33.3% (3) The final complex conformation of protein and one reassembled L700417, from the connecting stage, all side chains were set flexible with the backbone restrained. The accuracy of recovery is also ~33.3% (4) The final complex conformation of protein and three reassembled L700417s, only the hydrogen bonds in the secondary structures were restrained from the connecting stage. The accuracy of recovery reaches 100%.

methods while retain the conformational mobility as possible.

The solvent accessible surface areas of the recovered compounds seemed larger than that of subprotocol 1 and 2. Without any positional restraints on the protein, the active-site pocket becomes larger to accommodate multiple copies of building blocks or recovered ligands. It is also confirmed by visualization that the reassembled ligands have not docked to the same position as the crystal structure due to the search in a larger complex conformational space for the flexible protein.

Discussion

In this study we carried out a series of careful tests for a new methodology of flexible drug design – F-DycoBlock. Different strategies for including the protein flexibility were tested and compared.

For both enzymes, the all-atom restrained case was first tested for comparison. It is interesting to find that the accuracy of recovery was significantly enhanced with the grid approximation turned on or off, in contrast with DycoBlock [1, 2]. In this sense, it provides an alternative way to dock known ligands into active site or assemble novel ligands from separate building blocks. The docking study using (F-) DycoBlock

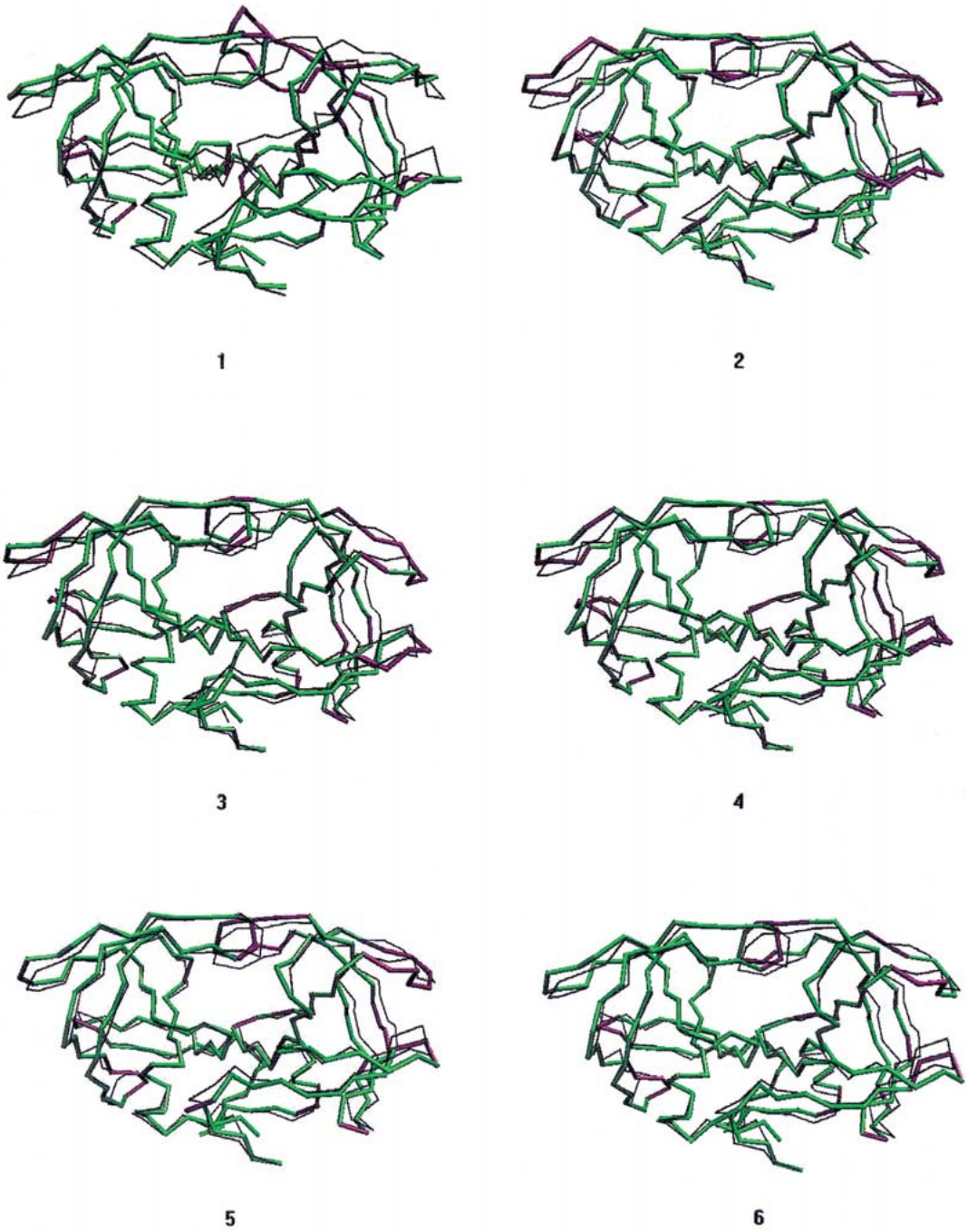


Figure 7. (1) Superimposition of the resultant conformation of HIV-1 protease from subprotocol 3 and the reference conformation from explicit MD simulation. Other five present superimpositions of crystal structures liganded with different inhibitors and the reference conformation, which are downloaded from Protein Data Bank (PDB) with PDB id (2) 1AAQ (3) 1HOS (4) 1HPS (5) 1HTE (6) 5HVP. The black lines present the reference conformation.

Table 4. Generated compounds ordered with respect to binding energies

		Explicit(kJ/mol)			SASA(Å^2)	
	E_{tot}	E_{bind}	E_{ele}	E_{lj}	Bind	Free
Subprotocol 1: all protein atoms were positionally restrained.						
1	−109.43	−380.11	−163.10	−217.01	152.30	558.47
2	−10.20	−376.05	−139.74	−236.31	139.73	557.48
3[a]	−83.35	−338.90	−75.37	−263.54	140.13	534.55
4[a]	−12.99	−313.70	−22.74	−290.96	141.58	550.32
5	13.20	−311.23	−73.78	−237.45	174.96	554.21
6	−37.66	−308.17	−73.21	−234.96	169.81	545.94
Subprotocol 2: backbone atoms were positionally restrained.						
1	212.95	−115.99	−15.99	−99.99	150.46	541.59
2[a]	297.91	−110.32	−11.91	−98.40	113.43	553.89
3	222.59	−99.22	48.92	−148.14	137.23	532.61
Subprotocol 3: hydrogen bonds in the secondary structures were restrained.						
1[a]	37.98	−284.78	−64.35	−220.43	231.76	551.64
2[a]	−5.12	−261.51	−40.14	−221.37	225.46	556.70
3[a]	24.55	−209.86	−28.37	−181.50	236.35	538.87

[a]Presents reassembled inhibitor – L700417, other resultant compounds are also listed.

will be presented elsewhere, in which the applicability of F-DycoBlock as a docking algorithm was also identified.

The side chain flexibility has been incorporated in the docking study for years. Rotamer library was most often used in this type of methods to provide the possible side chain conformations in a discrete way [25, 26, 46, 47]. However, it is clear that these methods strongly depend on the size of rotamer database and cannot present the cooperative behaviors of receptor and ligand in the binding process. The side chain flexibility was naturally achieved in the current study through the MCSMD simulation when the backbone atoms of protein were harmonically restrained. It was notable that the side chain flexibility has been perfectly combined with the dynamic assembly in F-DycoBlock and yielded at least the same accuracy as the restrained case for HIV-1 protease and higher accuracy in a docking study (shown elsewhere). Since no other special treatments but the dynamics simulation is involved in F-DycoBlock, it's definitely a general way to accommodate the side chain flexibility in drug design. The grid approximation led to the artifact that the active-site flexible strategy failed in contrast with other cases. However, this is not true when all protein atoms are explicitly treated (data not shown). Here we give the details of the strategy for COX-2 to show the effect of grid approximation (also mean field approximation) on the protein stability and accuracy of design.

The flexibility of whole protein can be easily included by removing the restraints on atoms. However, the results from several testing runs indicate that it must be carefully investigated in the theory and practice. If the whole protein was set flexible from the first stage (e.g., HIV-1 protease), the secondary structures were often partially lost. This problem is mainly caused by the LES approximation. First, when the copies are close enough, the renormalized force of all copies is meaningful [29]. However, the building block copies were located randomly when searching binding sites, the renormalization becomes ambiguous and leads to the incorrect motion of protein atoms. Second, as mentioned in the theoretical study, LES-MCSS methods are based on the time-dependent Hartree (TDH) approximation and assume that the protein conformation is only perturbed upon the formation of the ligand-protein complex [31, 33, 34, 48]. Therefore, it is still problematic when a large conformational barrier needs to be spanned between the unbound and complex conformations.

There are two possible approaches to handle this question. First, from the practical point-of-view, the

Table 5. HIV-1 protease complexed with different ligands

PDB ID	Residues mutated[a]	Ligand ID	Content of ligand
4PHV	None	VAC	N,N-BIS(2-HYDROXY-1- INDANYL)-2,6-DIPHENYLMETHYL-4-HYDROXY-1,7-HEPTANDIAMIDE
1AAQ	A,B37: ASN→SER A,B63: LEU→ILE	PSI	2-(2-{5-[2-(2-AMINO- PROPIONYLAMINO)-PROPIONYLAMINO]-4-HYDROXY-6-PHENYL-HEXANOYLAMINO}-3-METHYL-BUTYRYLAMINO)-3-METHYL-BUTYRIC ACID METHYL ESTER
1HOS	A,B37: ASN→SER	PHP	(2-PHENYL-1-CARBOBENZYL-OXYVALYL-AMINO)-ETHYL-PHOSPHINIC ACID
1HPS	A,B37: ASN→SER	RUN	2-[(1R,3S,4S)-1-BENZYL-4-[N-(BENZYLOXYCARBONYL)-L-VALYL]AMINO-3-PHENYLPENTYL]-4(5)-(2-METHYLPROPIONYL)IMIDAZOLE
1HTE	A,B37: ASN→SER	G23	(2R,4S)-2-[(R)-BENZYLCARBAMOYL-PHENYLACETYL-METHYL]-5,5-DIMETHYL-THIAZOLIDINE-4-CARBOXYLIC ACID
5HVP	None	STA	Acetyl-Pepstatin (NY5 Strain)

[a]1AAQ, 1HOS, 1HPS, 1HTE, 5HVP are compared with 4PHV to find mutated residues.

secondary structure elements can be restrained in the simulation. The protein structure will be protected from the errors introduced by the LES approximation, while the relative motions of local structures are allowed. The reasonable movement in this case may provide an approach to the challenging target – a *de novo* drug design without any information of complex structure. Second, the optimization methods using a Boltzmann-weighted mean field approach have recently been applied in sampling side-chain conformations [49, 50], which can be seen as the modification of TDH approximation. In the standard LES protocol, protein moves in the algebraic average field of a swarm of copies and each copy within the system is of equal probability, independent of time. However, the probabilities of conformations are treated as additional variables and must satisfy the Boltzmann relation in the new methods. Incorporating the Boltzmann-weighted methods into F-DycoBlock seems a possible theoretical solution.

There is much room to improve for F-DycoBlock except the limitation of LES approximation. The solvent effect is another factor that affects conformation of a flexible protein and is presented by the frictional and random force now. An implicit solvent model based on the stochastic dynamics and accessible-surface model has been parameterized in this group, whose applicability in study of protein folding and structure prediction has been identified. It provides an approach to deal with the solvent effect in F-DycoBlock. Furthermore, the explicit dynamics simulation of protein and building blocks or ligands makes it possible to incorporate more accurate methods of screening the candidate compounds into F-DycoBlock.

A new methodology, which contains the Boltzmann-weighted optimization method, implicit solvent model and fast evaluation of binding free energy, is under development and testing.

Acknowledgements

Chinese National Fundamental Research Project (Grant number: 19998051115), Chinese National Fundamental Research Project (Grant number: G1999075605) and Chinese National Natural Science Foundation (Key Project. Grant number: 39990600). We gratefully thank van Gunsteren W.F. (Department of Chemistry, ETH) for GROMOS96 force field.

References

1. Liu, H.-Y., Duan, Z.-H., Luo, Q.-M. and Shi, Y.-Y., Proteins, 36 (1999) 462.
2. Zhu, J., Yu, H.B., Fan, H., Liu, H.-Y. and Shi, Y.-Y., J. Comput. Aid. Mol. Des., 15 (2001) 447.
3. Böhm, H.J., J. Comput. Aid. Mol. Des., 6 (1992) 61.
4. Eisen, M.B., Wiley, D.C., Karplus, M. and Hubbard, R.E., Proteins, 19 (1994) 199.
5. Pearlman, D.A. and Murcko, M.A., J. Med. Chem., 39 (1996) 1651.
6. Bryngelson, J.D. and Wolynes, P.G., J. Phys. Chem., 93 (1989) 6902.
7. Shakhnovitch, E.T. and Gutin, A.M., Protein Eng., 6 (1993) 793.
8. Boczko, E.M. and Brooks, C.L. III, Science, 269 (1995) 393.
9. Dill, K.A. and Chan, H.S., Protein Sci., 4 (1995) 561.
10. Karplus, M., Sali, A. and Shakhnovitch, E.T., Nature (London), 373 (1995) 664.
11. Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z. and Socci, N.D., Proc. Natl. Acad. Sci. USA, 92 (1995) 3626.
12. Wolynes, P.G., Onuchic, J.N. and Thirumalai, D., Science, 267 (1995) 1619.
13. Dill, K.A. and Chan, H.S., Nature Struct. Biol., 4 (1997) 10.
14. Karplus, M., Folding Des., 2 (1997) S69.
15. Lazaridis, T. and Karplus, M., Science, 278 (1997) 1928.
16. Gruebele, M. and Wolynes, P.G., Nature Struct. Biol., 5 (1998) 662.
17. Martinez, J.C., Pisabarro, M.T. and Serrano, L., Nature Struct. Biol., 5 (1998) 721.
18. Dill, K.A., Protein Sci., 6 (1999) 1166.
19. Tsai, C.-J., Kumar, S., Ma, B. and Nussinov, R., Protein Sci., 6 (1999) 1181.
20. Fischer, E., Ber. Dtsch. Chem. Ges., 27 (1894) 2985.
21. Koshland, D.E., Proc. Natl. Acad. Sci. USA, 44 (1958) 98.
22. Ma, B., Kumar, S., Tsai, C.-J. and Nussinov, R., Protein Eng., 12 (1999) 713.
23. Carlson, H.A. and McCammon, J.A., Mol. Pharmacol., 57 (2000) 213.
24. Gschwend, D.A., Good, A.C. and Kuntz, I.D., J. Mol. Rec., 9 (1996) 175.
25. Schnecke, V., Swanson, C.A., Getzoff, E.D., Tainer, J.A. and Kuhn, L.A., Proteins, 33 (1998) 74.
26. Leach, A.R., J. Mol. Biol., 235 (1994) 345.
27. Lamb, M.L. and Jorgensen, W.L., Curr. Opin. Chem. Biol., 1 (1997) 449.
28. Miranker, A. and Karplus, M., Proteins, 11 (1991) 29.
29. Stultz, C.M. and Karplus, M., Proteins, 37 (1999) 512.
30. Stultz, C.M. and Karplus, M., Proteins, 40 (2000) 258.
31. Elber, R. and Karplus, M., J. Am. Chem. Soc., 112 (1990) 9161.
32. Roitberg, A. and Elber, R., J. Chem. Phys., 95 (1991) 9277.
33. Zheng, Q., Rosenfeld, R. and Kyle. D.J., J. Chem. Phys., 99 (1993) 8892.
34. Zheng, W.-M. and Zheng, Q., J. Chem. Phys., 106 (1997) 1191.
35. Koehl, P. and Delarue, M., Curr. Opin. Struct. Biol., 6 (1996) 222.
36. Zheng, Q. and Kyle, D.J., Proteins, 19 (1994) 324.
37. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery, J.A., Jr., Stratmann, R.E., Burant, J.C., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K.N., Strain, M.C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G.A., Ayala, P.Y., Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Ortiz, J.V., Baboul, A.G., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Andres, J.L., Gonzalez, C., Head-Gordon, M., Replogle, E.S. and Pople, J.A., Gaussian 98, Revision A.7, Gaussian, Inc., Pittsburgh PA, 1998.
38. van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hunenberger, P.H., Kruger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G., Biomolecular simulation: the GROMOS96 manual and user guide. University of Groningen, the Netherlands, ETH Zurich, Switzerland: Biomos, 1996.
39. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Dinola, A. and Haak, J.R., J. Chem. Phys., 81 (1984) 3684.
40. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327.
41. Stouten, P.F.W., Frommel, C., NaKamura, H. and Sander, C., Mol. Simul., 10 (1993) 97.
42. DeWitt, D.L., El-Harith, E.A., Kraemer, S.A., Andrews, M.J., Yao, E.F., Armstrong, R.L. and Smith, W.L., J. Biol. Chem., 265 (1990) 5192.
43. Seibert, K., Zhang, Y., Leahy, K., Hauser, S., Masferrer, J., Perkins, W., Lee, L. and Isakson, P., Proc. Natl. Acad. Sci. USA, 91 (1994) 12013.
44. Masferrer, J.L., Zweifel, B.S., Manning, P.T., Hauser, S.D., Leahy, K.M., Smith, W.G.,
45. Isakson, P.C. and Seibert, K., Proc. Natl. Acad. Sci. USA, 91 (1994) 3228.
46. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J., In: Intermolecular Forces, D. Reidel Publ. Co., Dordrecht, pp. 331–342.
47. Totrov, M. and Abagyan, R., Proteins, Suppl. 1 (1997) 215.
48. Schaffer, L. and Verkhivker, G.M., Proteins, 33 (1998) 295.
49. Straub, J.E. and Karplus. M., J. Chem. Phys., 94 (1991) 6737.
50. Zacharias, M., Luty, B.A., Davis, M.E. and McCammon, J.A., J. Mol. Biol., 238 (1994) 455.
51. Huber, T., Torda, A.E. and van Gunsteren, W.F., Biopolymers, 39 (1996) 103.