PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Combining coarse-grained nonbonded and atomistic bonded interactions for protein modeling

Martin Zacharias*

Physik-Department T38, Technische Universität München, James Franck Str. 1, 85748 Garching, Germany

ABSTRACT

A hybrid coarse-grained (CG) and atomistic (AT) model for protein simulations and rapid searching and refinement of peptide–protein complexes has been developed. In contrast to other hybrid models that typically represent spatially separate parts of a protein by either a CG or an AT force field model, the present approach simultaneously represents the protein by an AT (united atom) and a CG model. The interactions of the protein main chain are described based on the united atom force field allowing a realistic representation of protein secondary structures. In addition, the AT description of all other bonded interactions keeps the protein compatible with a realistic bonded geometry. Nonbonded interactions between side chains and side chains and main chain are calculated at the level of a CG model using a knowledge-based potential. Unrestrained molecular dynamics simulations on several test proteins resulted in trajectories in reasonable agreement with the corresponding experimental structures. Applications to the refinement of docked peptide–protein complexes resulted in improved complex structures. Application to the rapid refinement of docked protein–protein complex is also possible but requires further optimization of force field parameters.

Proteins 2013; 81:81–92.

Key words: protein–protein interaction; binding site prediction; biased force field; docking by energy minimization; protein–protein complex formation.

INTRODUCTION

Molecular dynamics (MD) simulations are frequently used to study the dynamics of peptides and proteins. However, despite recent algorithmic progress to improve sampling and the development of specialized hardware to run MD simulations up the microsecond and even millisecond regime,[1] the routine use of such approaches up to biologically relevant time scales is still limited.

To reduce the computational burden during molecular simulations coarse-grained (CG) models have been developed that represent the protein or peptide by a smaller number of sites compared with a full atomistic (AT) description.[2–19] A variety of CG models have been developed that differ in the number of interacting sites and the form of the force field to model effective bonded and nonbonded interactions (reviewed in Refs. 16–19). Several approaches have been designed to fit CG force fields to match both the interactions and sampled conformations of full AT models as closely as possible.[11–14] One major difficulty is to realistically represent effective nonbonded interactions between CG particles. Effective interactions between pairs of particles depend on whether the particles are surrounded by solvent or are located in the interior of a protein. There have been efforts to approximately account for such effects by including multibody potentials.[20,21] Another possibly even more severe difficulty is to accurately represent the bonded geometry of a given protein or peptide molecule at the CG level. Typically, for the CG model a force field representation similar to an AT force field is used consisting of bond length, bond angle, and dihedral torsion terms based on the pseudoatoms coordinates of the CG model.[16,17] The parameters for the force field (e.g., equilibrium bond lengths, angles, etc., and corresponding

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Deutsche Forschungsgemeinschaft (DFG); Grant number: Za-153/5-3; Grant sponsor: Pacific Northwest National Laboratories (PNNL); Grant number: grant st44711.

*Correspondence to: Martin Zacharias, Physik-Department T38, Technische Universität München, James Franck Str. 1, 85748 Garching, Germany.
E-mail: martin.zacharias@ph.tum.de.

Received 19 June 2012; Revised 9 August 2012; Accepted 15 August 2012
Published online 22 August 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24164

force constants) can be derived from AT simulations or from data bases of experimental structures of proteins.

However, due to the coarse-graining of a structure, many of the bonded interactions between pseudoatoms are strongly coupled because a bonded interaction implicitly includes many nonbonded as well as bonded interactions between the "real" atoms it represents. An accurate accounting of all coupling effects between bonded force field terms for a CG model might be difficult to achieve. One way to get around this problem is to keep an AT description of bonded interactions but introduce pseudocenters (CG beads) that represent centers of geometry (or centers of mass) of groups of atoms (virtual sites). An approach related to such model is the Rosetta low-resolution force field.[22] In this force field, the protein backbone is represented at full atomic resolution, but each side chain is represented by a single center just located at the C_β of each amino acid. A drawback of the model is that different rotameric states of a side chain are not represented which might be a poor approximation especially for large side chains. Second, many side chains have a dual physicochemical character (e.g., containing hydrophobic as well as hydrophilic groups) such that in a single bead representation in the worst case, both characters are lost.

In the present approach, a reasonably accurate force field representation of a peptide chain based on heavy atoms (or united atoms) for both protein backbone and side chains and one explicit backbone hydrogen per residue) has been combined with virtual centers that represent centers of mass of side chain chemical groups. As the positions of side chain centers are calculated on the fly during the simulations, the motion of the CG representation is always compatible with the AT model (the CG centers cannot occupy unrealistic positions with respect to the AT representation). Note that this is different from other hybrid models between AT and CG representations that usually represent one part of a simulation system at the AT level (e.g., functional site of a large protein) and the other spatially separate part (rest of the protein) at the CG level, whereas in the present approach, both CG and AT levels are present, one controlling the nonbonded and the other the bonded interactions.

As force field parameters for the CG side chain–side chain (SC–SC) and side chain–backbone interaction, the empirical ATTRACT CG force field[23,24] with parameters previously optimized for protein–protein docking was used. The performance of the model was evaluated in MD simulations of peptides and proteins as well as for protein–peptide docking and refinement including full side chain and limited backbone flexibility.

MATERIALS AND METHODS

The GROMOS united atom force field[25] with modifications (see Supporting Information) was used in the present hybrid CG and AT representation for controlling the bonded geometry. The force field does not include hydrogens attached to nonpolar atoms (only united heavy atoms) and has the typical form given below:

$$V_{\mathrm{tot}} = \sum_{i=1}^{N\mathrm{bonds}} 1/2\; k_{\mathrm{b\_i}}(b_i - b_{\mathrm{oi}})^2 + \sum_{i=1}^{N\mathrm{angles}} 1/2\; \mathrm{k}_{\theta\_i}(\theta_i - \theta_{\mathrm{oi}})^2 + \sum_{i=1}^{N\mathrm{dihedral}} \sum_{n=1}^{N\tau} k_n(1 + \cos(n\tau_i + \delta_{ni})) + \sum_{i\neq j}^{\mathrm{backbone\_and\_1\_4pairs}} \left(A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6} + q_i q_j (\epsilon(r_{ij}) r_{ij})^{-1}\right) + \sum_{i:N-H:j:O}^{H-\mathrm{bonds}} \left(A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-10}\right)$$

In the present application, only polar hydrogen atoms attached to the backbone nitrogen were included. The bonded terms (first three summations in the equation below) contain a sum over all bonds, all bond angles, and dihedral angles of the protein structures (including the side chain heavy atoms). Van der Waals and Coulomb interactions between atoms that are separated by three bonds (1–4 neighbors) also control the local bonded geometry and were explicitly included. In addition, interactions between backbone atoms were considered at AT resolution (this allows accurate representation of hydrogen-bonded secondary structures). For electrostatic interactions, a distance dependent dielectric scheme was used ($\epsilon = 16r$; to be compatible with the CG model). To reproduce secondary structures more accurately an extra term for hydrogen bonds ($r^{-12} - r^{-10}$) of the protein main chain was added (see Supporting Information). However, all side chain heavy atoms were treated as dummy atoms with respect to nonbonded interactions to all other atoms. Instead the nonbonded SC–SC as well as side chain–main chain (SS–MC) interactions were calculated via a CG model taken from the ATTRACT docking program.[23,24] It uses a protein representation with two pseudoatoms per residue representing the main chain (located at the backbone nitrogen and backbone oxygen atoms, respectively). Small amino acid side chains (Ala, Asp, Asn, Cys, Ile, Leu, Pro, Ser, Thr, and Val) are represented by one pseudoatom (centers of mass of side chain heavy atoms). Larger and more flexible side chains are represented by two pseudoatoms to account for the shape and dual chemical character of some side chains (for details see Ref. 23). Note that the ATTRACT docking program allows some degree of flexibility of proteins during docking (e.g., in protein normal mode directions) but not a full mobility of all atoms in each Cartesian degree of freedom (as in the present approach). A schematic representation of the force field description is given in Figure 1. Effective interactions between pseudoatoms are described by soft distance-de-

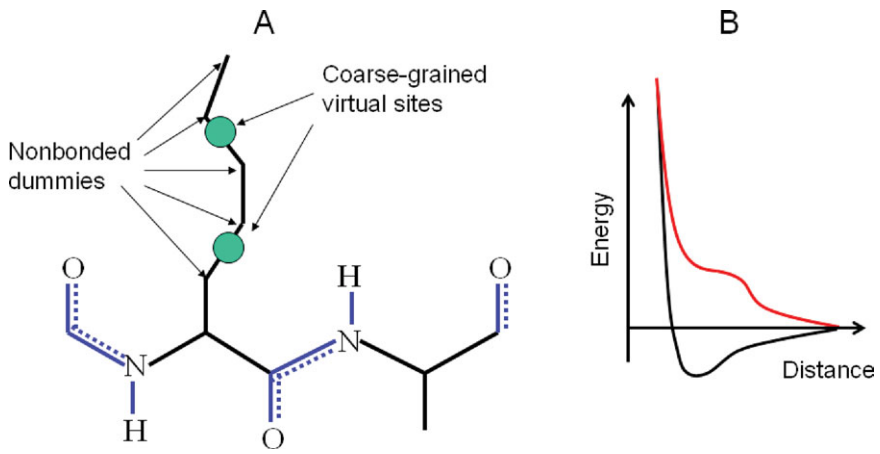

Figure 1

Schematic illustration of the hybrid coarse-grained and atomistic force field for proteins. (**A**) Each coarse-grained center (green circles) corresponds to the center-of-mass of groups of atoms (definition given in reference [23]). (**B**) All nonbonded interactions between side chains and between side chains and main chain are calculated using the CG representation and a Lennard–Jones type interaction function with a saddle point for purely repulsive pairs (red curve in B) of centers or attractive pairs (with an attractive minimum, black line in B) as described in the Methods section (see also Supporting Information of Ref. 24). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

pendent Lennard–Jones (LJ)-type potentials of the following form.[24]

$$V = \epsilon_{AB}\left[\left(\frac{R_{AB}}{r_{ij}}\right)^8 - \left(\frac{R_{AB}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{\epsilon(r_{ij})\, r_{ij}}$$

in case of attractive pair:

$$V = -\epsilon_{AB}\left[\left(\frac{R_{AB}}{r_{ij}}\right)^8 + \left(\frac{R_{AB}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{\epsilon(r_{ij})\, r_{ij}} \quad \text{if } r_{ij} > r_{\min}$$

and in case of purely repulsive pairs:

$$V = 2\, e_{\min} + \epsilon_{AB}\left[\left(\frac{R_{AB}}{r_{ij}}\right)^8 + \left(\frac{R_{AB}}{r_{ij}}\right)^6\right] + \frac{q_i q_j}{\epsilon(r_{ij})\, r_{ij}} \quad \text{if } r_{ij} \le r_{\min}$$

R_{AB} and ϵ_{AB} are effective pairwise radii and attractive or repulsive LJ parameters, respectively. At the distance $r_{\min}$ between two pseudoatoms the standard LJ potential has the energy $E_{\min}$.(attractive pair) or a saddle point (with energy $-E_{\min}$, for a repulsive pair). A Coulomb-type term accounts for electrostatic interactions between real charges (Lys, Arg, Glu, and Asp) damped by a distance-dependent dielectric constant ($\epsilon = 16r$). In contrast to a standard LJ potential, the above form allows for purely repulsive interacting pseudoatom pairs (Fig. 1). As described in the study by Fiorucci and Zacharias,[24] the attractive and repulsive parameters for each pseudoatom pair have been iteratively optimized by minimizing the root-mean-square deviation (Rmsd) of a large number of near-native protein–protein docking minima (compared with the experimental structure) and by comparing the scoring of near-native minima with many decoy complexes of high surface complementarity. Hence, the parameterization represents the effective interactions between pseudocenters at interfaces. The parameters implicitly account for solvation effects, and no explicit solvent molecules were included during simulations. Energy minimization and test MD simulations indicated that a scaling of the ϵ_{AB} parameters by 2.5 in the hybrid CG/AT approach compared with those used during rigid protein docking in ATTRACT gave best agreement of minimized or average structures with experiment. Note that using rigid protein partner structures in the bound conformation, the parameter set ranks near-native docking solutions within the top 10 scoring complexes in 90% of test cases.[24] The Rmsd_{lig} (backbone Rmsd of smaller partner protein from the native protein after superposition of the larger partner onto the corresponding protein in the native complex) of near-native docking minima (rigid partner proteins) was <2 Å for 90% of test complexes.[24]

MD simulations at various temperatures were performed using the leapfrog scheme and either an Andersen thermostat[26] or a Berendsen thermostat[27] (2-ps time constant) and random velocity reassignments every 5000 MD steps. Simulations were performed with an in house code. As in the current study, the focus is on sampling possible conformational states and not on accurately describing kinetics of conformational changes. Therefore, where all atom masses were set to the mass of oxygen. A time step of 2 fs was used in all simulations. For all nonbonded interactions, a cut off radius of 8 Å and a smooth switching function (up to 10 Å) was applied with frequent updates of the pair list every five steps. The small cutoff was used because it corresponds to the

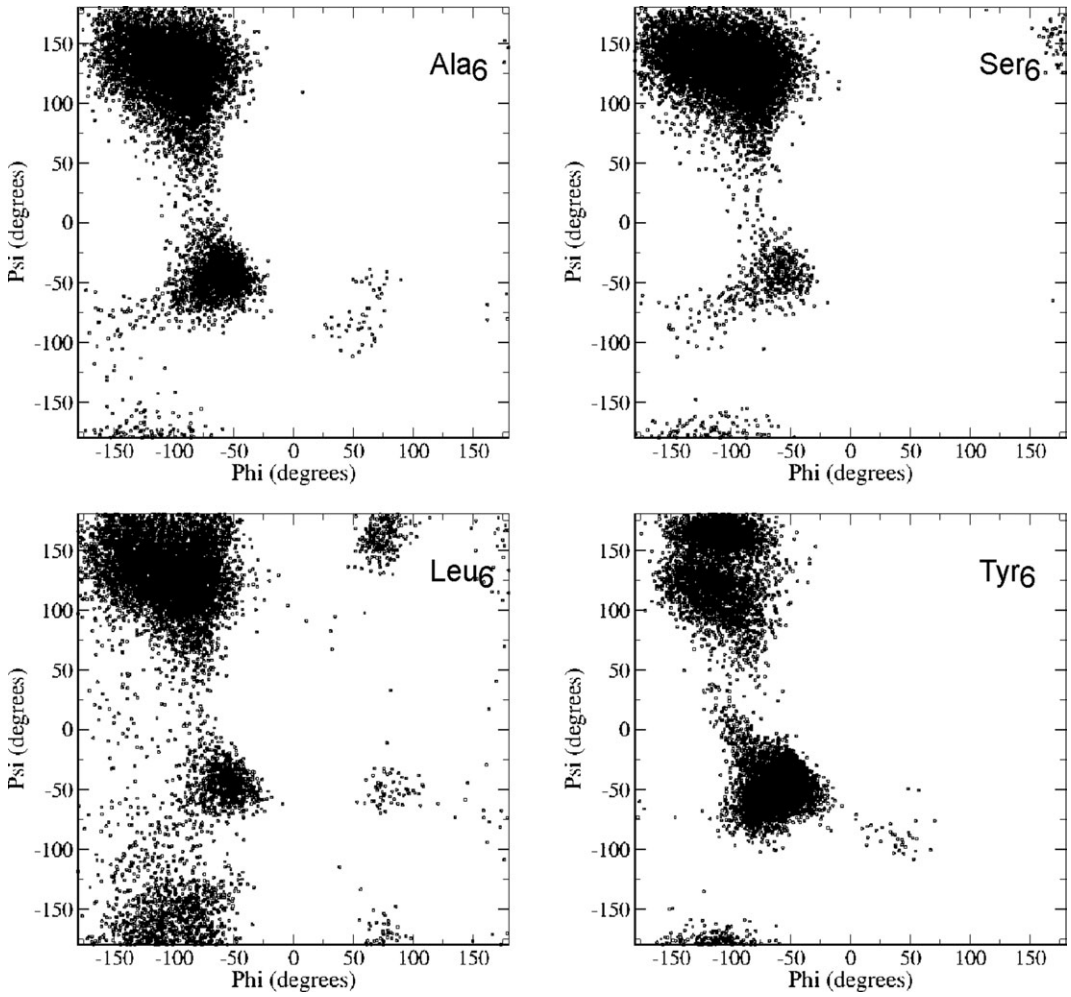

Figure 2

Sampling of peptide main chain dihedral states of short six mer oligopeptides during MD simulations with the hybrid CG/AT force field. Each point represents a sampled pair of ϕ/ψ main chain dihedral angles during 50 ns simulations at 300 K.

cut off radius that was used for optimization of the CG nonbonded parameters.[24]

RESULTS AND DISCUSSION

Molecular dynamics simulations on peptides and proteins

To evaluate the sampling of relevant main chain states, the hybrid (CG/AT) force field description (see Methods and Fig. 1) was used in MD simulations (50 ns) of several homo-oligopeptides (Fig. 2). Note that the interactions between CG centers represent effective short-range interactions derived previously from an analysis of contacts at native protein–protein interfaces compared to decoy interfaces[24] (Fig. 1). Solvation effects are therefore included implicitly in the parameterization (no explicit solvent molecules are included during simulations). As already indicated in the Methods section, the interactions between pairs of pseudocenters can be overall repulsive or attractive (Fig. 1) depending on the type of interacting pseudocenters.

For the test cases, qualitatively similar Ramachandran plots of the peptide main chain were observed (Fig. 2) that show reasonable agreement with the known favorable regions in the Ramachandran plot of peptides (population of the pp_{II}/β-strand regime and the α-helical regime). However, the pp_{II} regime (ψ = 140–160° and ϕ between −150 and −120°) and β-strand regime (ψ = 140–160° and ϕ between −120 and −60°) overlap with no clear separation. The result indicates that the presence of CG pseudocenters for the calculation of effective SC–SC as well as SS–MC interactions does not result in significant sampling of unrealistic main chain dihedral states. However, the distribution of sampled states indicates a preference for the β-strand region compared with the α-helical regime of peptide conformations for the hybrid force field. This might be due to an intrinsic preference of the AT force field (Gromos) used for the

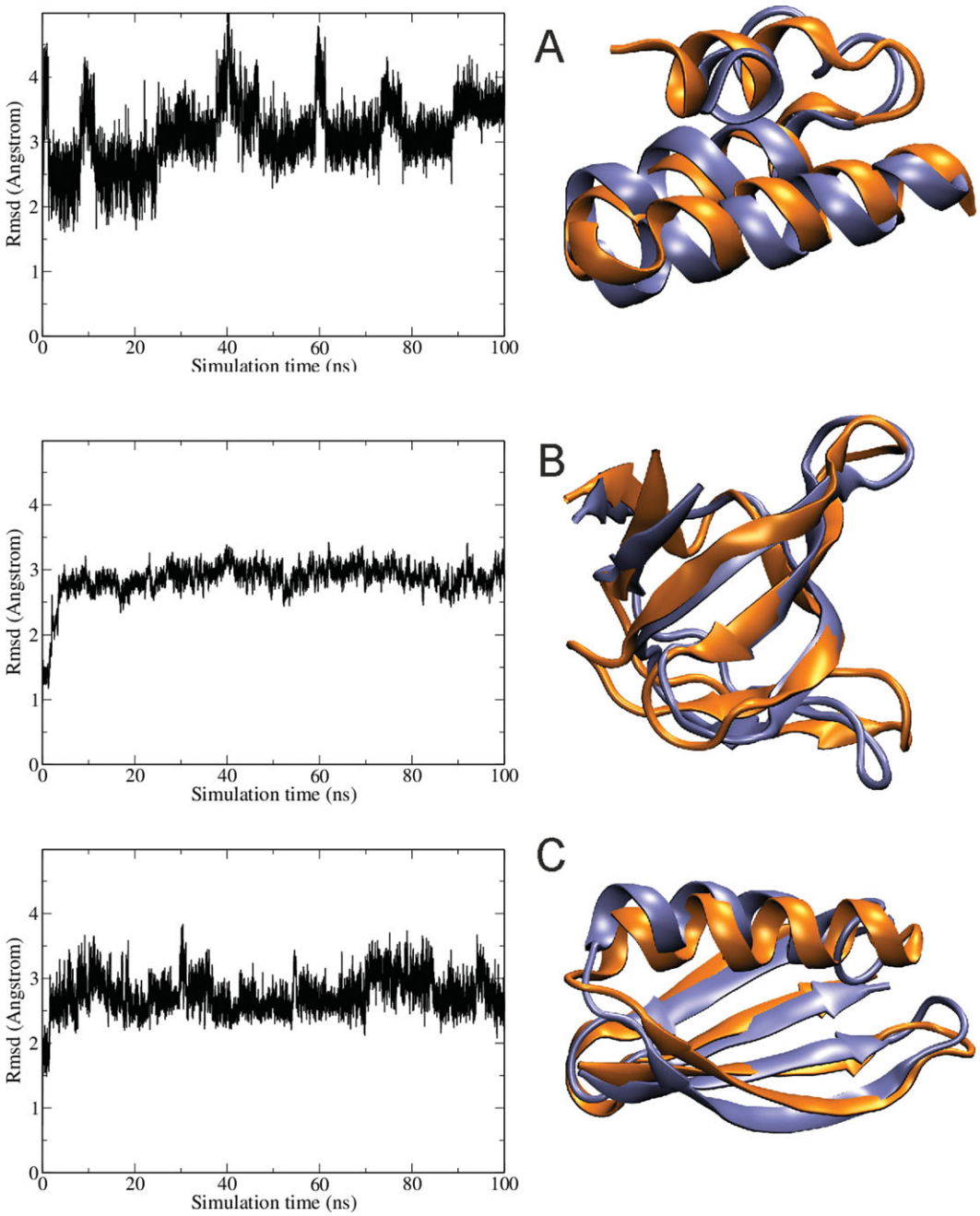

Figure 3

(**A**) (left panel) Backbone root-mean-square deviation ($Rmsd_{C\alpha}$) of the E domain of protein A from *S. aureus* from the X-ray start structure (residues 8–52 of pdb1EDL) during CG/AT-MD simulation versus simulation time. A superposition of the final structure (iceblue cartoon) onto the start structure (yellow cartoon) is shown in the right panel. (**B**) and (**C**) same as (A) but for an SH3 domain (pdb1FYN) and the immunoglobulin binding domain of Streptococcal protein G (pdb1GB1), respectively. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.].

bonded interactions. It is possible to further improve the description of the bonded interactions by a better parameterization of the bonded terms in future studies.

Unrestrained MD simulations using the hybrid CG/AT model on three different proteins (corresponding to the main types of folded proteins: α-helical, β-strand-containing, and mixed-α/β fold) were performed starting from the energy-minimized experimental crystal structures (Fig. 3). Simulations at 300 K were extended to 100 ns and used for comparison of the sampled states with respect to the native start structure. In all cases, the overall folding topology and secondary structure remained reasonably close to the start structure (Fig. 3). The $Rmsd_{C\alpha}$ of the protein backbone from the start structure

reached ~3–4 Å after a few nanoseconds and stayed at this level for the rest of the simulations. For the α-helical protein (E domain of protein A from *Staphylococcus aureus*, pdb 1EDL), fluctuations in the Rmsd-plot were observed that corresponded to partial dissociation and reassociation of the short N-terminal α-helix [Fig. 3(A)]. Interestingly, the overall shift in Rmsd was lowest in case of the β-sheet (SH3 domain, pdb1FYN) fold which might be due to the force field preference for the β-strand regime. The observed deviations are larger than what is typically observed in MD simulations including explicit solvent molecules. However, all 100-ns atom simulations performed with an implicit solvent model based on a distance-dependent dielectric constant resulted in similar or even larger deviations from the starting structure (see Supporting Information Fig. S1). Comparison of the atomic fluctuations indicated a similar pattern of atom mobility during both all-atom simulations and the simulations with the hybrid-CG/AT force field (see Supporting Information Fig. S2). The simulations with the hybrid-CG/AT force field were ~8–10 times faster when compared with all-atom simulations (using the same processor).

It should be emphasized that for the present version of the force field model the experimentally observed structure corresponded probably only to a locally stable folding regime. Attempts to use the approach to fold proteins starting from an extended chain in long simulations were for all three cases unsuccessful (data not shown).

Application to flexible peptide–protein docking and refinement

As indicated by the MD simulations on small proteins the present hybrid CG/AT force field is of only limited accuracy for applications that involve completely unrestraint proteins or peptides. More extensive future parameterization of bonded as well as nonbonded force field parameters is necessary to improve the performance in unrestraint simulations.

Nevertheless, even with the current parameters useful applications are possible if reasonable additional assumptions can be introduced as additional restraints. One possible application is the flexible docking or refinement of peptide–protein complexes with limited allowed flexibility of the binding partners. Although protein–protein and peptide–protein binding can involve substantial conformational changes such as refolding of entire parts of proteins most globularly folded proteins undergo only limited changes upon association of less than 1–2 Å of backbone Rmsd.[28] Hence, at least for these cases it is possible to limit the allowed conformational movements to fully flexible side chains and modest changes of the backbone by including restraints during simulations.

As test cases complexes of a fragment from the p53 tumor suppressor protein and the murine double minute 2 (MDM2) protein[29,30] (protein data bank entry 1T4F) and of an α-helical fragment of the Axin protein and the adenomatous polyposis coli tumor suppressor protein (APC) were used. The former complex consists of a small MDM2 protein domain and a peptide binding segment of the p53 protein in an α-helical conformation. Also in case of the Axin-ADC complex the Axin fragment binds in an α-helical conformation to the ADC receptor structure.[31]

During MD simulations harmonic distance restraints between the backbone Cα atoms at a distance range of 5–12 Å were included for the p53-peptide and the Axin-peptide to keep it in an α-helical conformation. The reference distances corresponded to the distances in the native experimental X-ray structure (pdb1T4F and pdb1EMU, respectively). The distance restraints allowed for full side chain flexibility and deviations of the backbone of ~1 Å from the reference structure. Positional restraints on the Cα atoms of the receptor protein (1 kcal $mol^{-1}A^{-2}$) were used to keep the structure close to the experimental receptor conformation. In case of the MDM2 receptor, the bound structure was used (no unbound form available), and in case of the ADC-receptor, the unbound receptor structure was used (pdb1DK8). Note, however, that during the initial simulation at 500 K, the surface side chains can undergo many transitions to alternative rotameric states such that even in case of the MDM2 system, the receptor start structure corresponded to an "unbound" conformation (of side chains) with small backbone deviation from the bound form. In addition, a weak harmonic restraint between the centers of mass (only Cα atoms) of the peptide and receptor protein was included (force constant 0.005 kcal $mol^{-1}A^{-2}$). The center-of-mass restraint allowed dissociation and motions of the peptide on the receptor surface but prevented complete separation of the partners (penalty energy increases with separation distance).

Using this set-up, it was possible to perform a three-stage simulated annealing (SA) procedure starting at a high simulation temperature of 500 K (10 ns) to rapidly search for binding sites on the surface of partner proteins without unfolding of the partner structures. The SA procedure was completed by 10 ns at 450 K and 10 ns at 300 K (Fig. 4). Simulations were started from eight peptide–protein docking placements (using the ATTRACT program and using rigid partners) of the α-helical peptide distributed over the surface of the corresponding (MDM2 or the ADC) receptor protein, respectively, [Fig. 4(a)] including placements on opposite side to the side with the binding interface. One SA run (30 ns) required ~6 h computer time on a single processor core.

In all MDM2/p53 cases except one, a placement close to the experimental binding geometry was found after ~30 ns simulation time with a total backbone $Rmsd_{C\alpha}$ of ~1.5 Å, which translates to a deviation of the ligand peptide (after superposition of only the receptor atoms

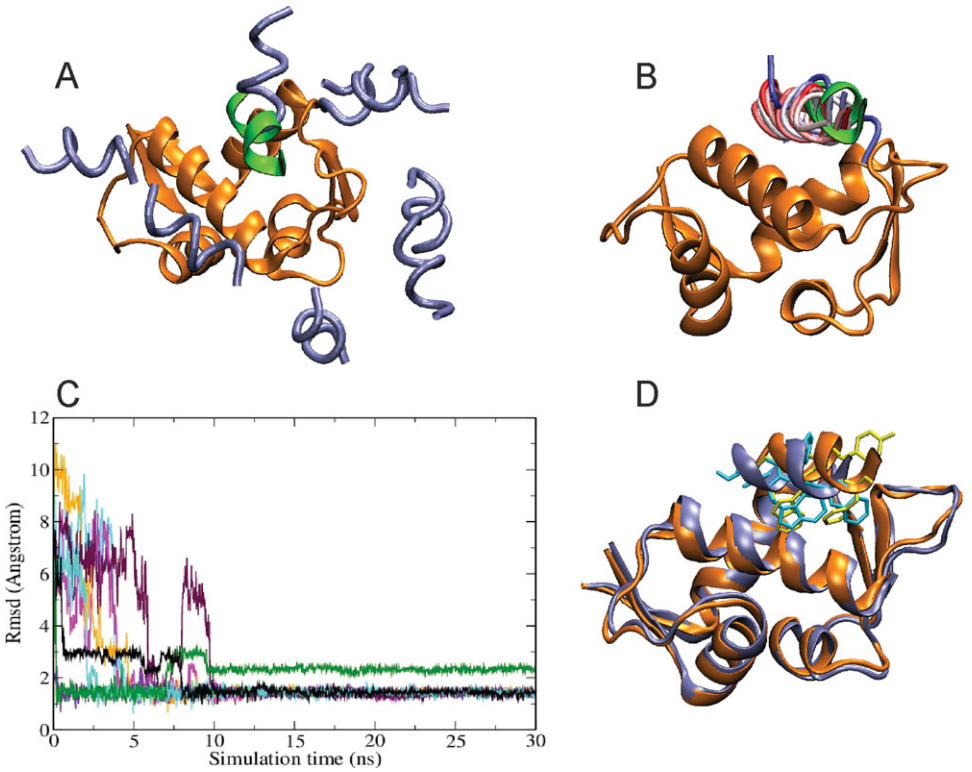

Figure 4

(**A**) Starting configurations for the MD-simulated annealing (MD-SA) simulations with the CG/AT hybrid model of the α-helical p53 peptide binding (iceblue cartoon models) to the MDM2 receptor protein (orange cartoon). The native binding mode of the p53 peptide is shown in green (pdb14TF). (**B**) Superposition of the (eight) final p53 peptide binding placements (tube representation in different colors; native placement as green cartoon). (**C**) $Rmsd_{C\alpha}$ (of the complete complex with respect to the native complex structure) of the sampled conformations during a three-stage restraint MD-SA simulation (10 ns: 500 K; 10 ns: 400 K; 10 ns: 300K). Each color represents a different starting geometry. (**D**) Superposition of one near-native complex (iceblue) onto the native complex (orange) including the side chains (stick model) at the peptide–protein interface.

on the native complex) of $Rmsd_{lig}$ ~3 Å. It should be emphasized that a very similar placement is obtained if one starts the simulation from the native complex indicating that the force field favors this placement over the native complex structure. In one case, the final complex structure had an Rmsd of 2.2 Å ($Rmsd_{lig}$ = 3.6 Å) from the native complex indicating an alternative trapped complex state [green line Fig. 4(C)].

For the second test case, the same MD protocol was used starting from 10 randomly docked Axin peptides onto the surface of the ADC-receptor protein. Of 10 starting geometries, four trajectories resulted in final complex structures in close agreement with the experimental complex structure (Fig. 5; Table I) after ~20–30 ns ($Rmsd_{C\alpha}$ ~1.3–1.9 Å; $Rmsd_{lig}$ ~3.5–5 Å). The other six simulations were trapped at other putative interaction sites that have a concave shape and contain hydrophobic patches, which may interact favorably with the hydrophobic part of the peptide ligand (Fig. 5).

Inspection of the Rmsd versus time plots [Fig. 5(E)] indicates that the final 10 ns (at 300 K) involved only small adjustments or changes of the complex structures, and this part was used to evaluate average total potential and interaction energies. Interestingly, during this phase all four successful trajectories mainly fluctuated between two states with slightly different $Rmsd_{C\alpha}$ of 1.9 or 1.3 Å, respectively, from the native complex structure. To check if the different final complexes can be distinguished energetically, the average total force field energies and interaction energies as well as the corresponding energies after extensive energy minimization (of the final snapshot) were compared (Table I).

The nonbonded average energy or final interaction energy (after EM) of the CG centers was in most but not all near-native complexes more favorable compared with the alternative binding geometries (Table I). However, one non-native binding mode resulted in the best average interaction energy and second best interaction energy after energy minimization (Structure 10, Table I). The total potential energy after energy minimization of the final snapshots also showed for most of the four docking solutions with near-native geometry the lowest energy, but one alternative complex (Structure 3) gave a slightly better score than one of the near-native cases (Structure 4). As

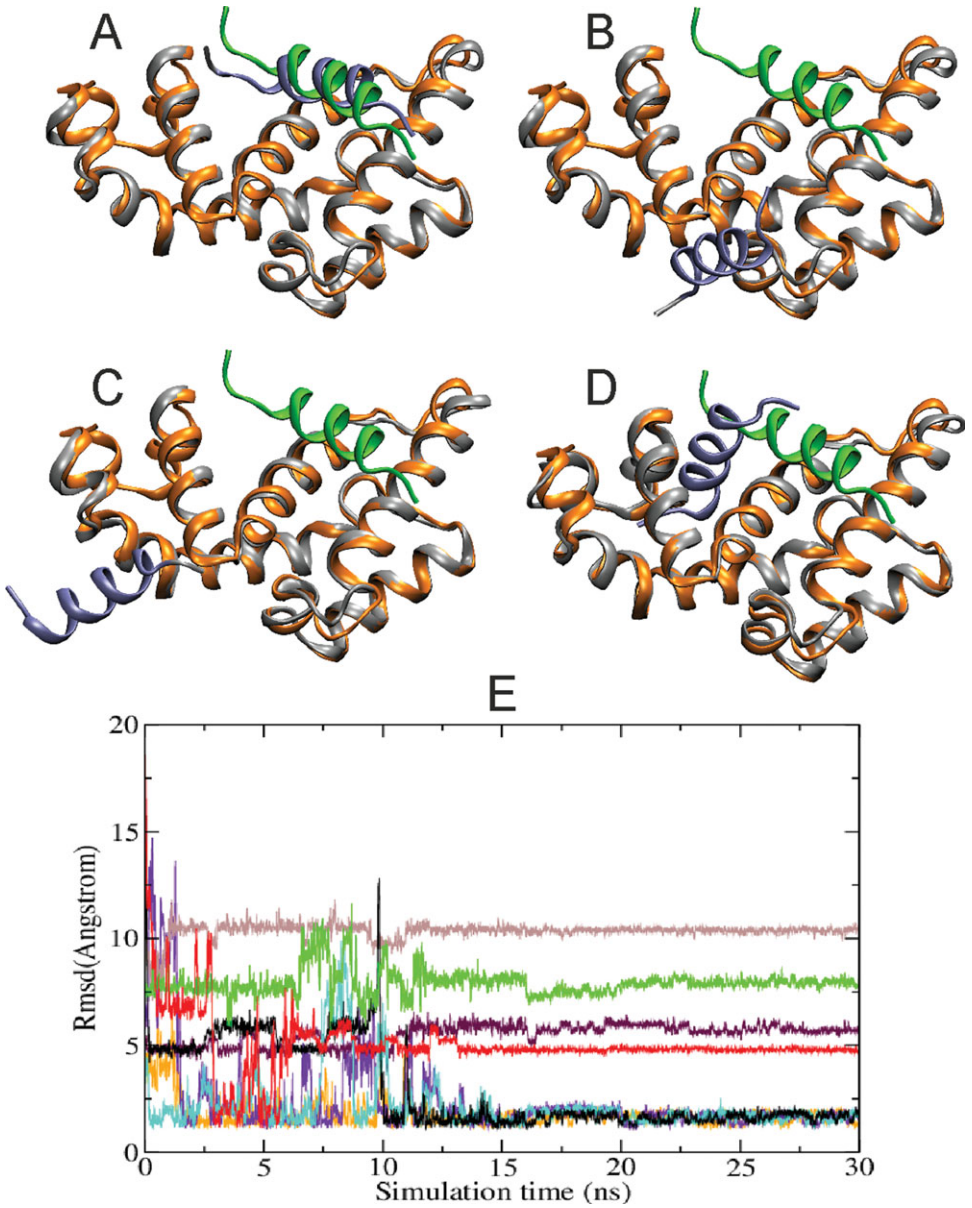

Figure 5

(**A**–**D**) Final peptide–receptor complexes obtained after MD–SA simulations of the axin-peptide–ADC receptor system (pdb1EMU) with the CG/AT model starting from several peptide starting positions distributed over the surface of the ADC receptor (gray cartoon represents ADC receptor; iceblue cartoon is the peptide). Structures were superimposed onto the native complex (with respect to the ADC receptor; orange cartoon represents native ADC structure and green cartoon corresponds to the native Axin peptide placement). The placement in (A) comes closest to the native structure (representative of four solutions with smallest $Rmsd_{C\alpha}$ from experiment). (**E**) $Rmsd_{C\alpha}$ versus simulation time for the three-stage MD-SA simulations (Fig. 4). Each color represents a different starting geometry. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.].

the total energy is affected by many parts of the structures not participating at the interface, it is expected that it may not correlate perfectly well with a favorable or unfavorable peptide–protein interaction. Interestingly, the average total energy of the final part of the simulation (10 ns at 300 K) allowed the best distinction of near-native complex geometries from alternative binding modes (lower average total potential energy for each near-native complex compared with all alternative binding states, see Table I). However, it has also the largest standard deviation. As the average total potential energy is (besides of entropy) the driving force for the peptide–protein complex formation, this result explains why the near-native solution was found more frequently than other possible solutions. Instead of using any energetic evaluation for distinguishing near-native from incorrect binding modes, it is in this case

Table I
Molecular Dynamics Docking Simulation on Axin-peptide-ADC Complex

Structures	$<E_{pot}>_{300\ K}$	$<E_{gr\_int}>_{300\ K}$	E_{min}	$E_{gr\_int\_min}$	$Rmsd_{final}$	$Rmsd_{lig\_final}$	$Rmsd_{start}$	$Rmsd_{lig\_start}$
1	**626 (±21)**	**−27.8 (±3)**	**−586.3**	**−27.7**	**1.3**	**3.4**	**3.6**	**11.0**
2	**625 (±20)**	**−27.4 (±3)**	**−595.3**	**−31.1**	**1.4**	**3.7**	**6.8**	**21.5**
3	644 (±21)	−27.0 (±4)	−586.1	−27.4	5.8	17.6	5.7	17.3
4	**629 (±20)**	**−26.6 (±3)**	**−585.2**	**−28.7**	**1.8**	**4.9**	**5.5**	**16.8**
5	640 (±20)	−22.1 (±3)	−585.6	−26.2	10.6	33.1	5.7	17.1
6	639 (±21)	−21.2 (±3)	−581.9	−22.5	7.2	22.8	6.6	20.1
7	647 (±20)	−20.3 (±3)	−570.3	−28.1	10.4	32.6	7.4	23.3
8	**629 (±21)**	**−28.2 (±3)**	**−595.8**	**−33.7**	**1.75**	**4.9**	**7.8**	**23.4**
9	647 (±20)	−19.7 (±3)	−565.1	−23.3	7.9	24.7	9.1	25.7
10	633 (±20)	−28.7 (±3)	−578.6	−33.6	4.8	14.0	10.8	33.1

First column corresponds to start structure number; $<E_{pot}>_{300K}$and $<E_{gr\_int}>_{300K}$ are average (over 10 ns at 300 K) total potential energy and average CG interaction energy (in kcal mol^{-1}) between peptide and receptor protein, respectively, after a simulated annealing MD simulation. The next two columns represent the same energies after energy minimization of the final complexes. $Rmsd_{final}$ and $Rmsd_{lig\_final}$ indicate the backbone Rmsd (in Å) and the $Rmsd_{Lig}$ (backbone Rmsd of the peptide after best superposition of the receptor with respect to the native complex), respectively, of the final energy-minimized complexes from the native complex. The last two columns indicate the corresponding Rmsds for the start structures. Data for the four cases that resulted in near-native final complex structures are in bold.

possible to simply use the most frequently found binding geometry (or a representative cluster) as the putative correct peptide–protein binding complex after performing several MD-SA runs.

In addition to flexible peptide–protein docking applications, the approach was also used to refine incorrectly docked complexes obtained from systematic protein–protein docking searches (using rigid partner structures during the initial docking) involving larger interfaces than in case of peptide–protein complexes. However, in this case frequently, the MD refinement of incorrect non-native binding geometries gave lower energies than of the native complex (not shown). The ATTRACT force field has been shown to distinguish near-native docking geometries quite efficiently form non-native geometries in case of docking rigid proteins or including a few degrees of flexibility.[24,32–35] It is likely that the inclusion of restricted backbone flexibility and full side chain flexibility as in the present MD-refinement procedure results in many alternative complex structures with high surface complementarity and similar or better scoring as the native complex.

To illustrate the difficulties in case of protein–protein complexes, the results for refinement of eight docking solutions of the colicin E9 enzyme in complex with the Immunity protein 9 (E9/Im9-complex, pdb1EMV)[36] are presented in Figure 6. The eight starting geometries were obtained from rigid docking of the two partner proteins in the bound form and corresponded to solutions relatively close to the native geometry with deviations from the native complex ranging from an $Rmsd_{C\alpha}$ ~1.5–3.5 Å ($Rmsd_{lig}$~2.5–6.5 Å, see Fig. 6). The refinement procedure corresponded to a short MD simulation of 100–200 ps at 450 or 500 K followed by energy minimization of the final structure. During the simulations, distance restraints between Cα backbone atoms of the smaller ligand protein (Im9) and positional restraints on the Cα atoms of the receptor protein were included as well as a center-of-mass restraint between the centers of both proteins similar to the conditions described for the peptide–protein complexes. For each of the eight starting complexes, nine refinement simulations with different initial velocities and slight variations in simulation length (150–200 ps) or temperature (450–500 K) were performed.

As illustrated in Figure 6(A), the procedure resulted in several solutions with lower deviation from the native complex geometry than the starting structure [points below the diagonal of the Rmsd plot in Fig. 6(A)], even sometimes starting from a complex with significantly larger Rmsd than the final complex. However, about half of the generated final structures deviated more from the native structure than the corresponding start structure. In addition, neither the total energy nor the interaction energy of CG centers of the final energy minimized complexes showed a clear correlation with respect to the deviation of the complex from the native complex structure [Fig. 6(B,C)]. It might be possible to design a composite score (with weighted contributions from different energy terms) that may show a more reasonable correlation. However, an optimal composite score is not universal for each protein–protein complex. A promising result on the E9/Im9 test case is that the procedure at least in several cases samples complex geometries close to the native complex even if one starts relatively far from the native structure. Nevertheless, an improved parameterization of the nonbonded CG interactions as well as the bonded terms is necessary to apply the approach systematically in flexible refinement of protein–protein complexes in future studies.

CONCLUSIONS

A new hybrid AT (for bonded) and CG model (for nonbonded interactions) and several initial tests of protein MD simulation and docking search and refinement

simulations have been presented. An advantage of the method compared with other CG representations is the fact that the CG centers correspond to virtual sites and are always compatible with the movements of the AT model. Therefore, no particular and often difficult parameterization of force field terms that describe an accurate bonded geometry at the level of the CG centers is necessary. For example, secondary structural elements such as α-helices and β-sheets are represented with the same or a similar accuracy as a pure AT force field. This comes at the price that the computational savings of the method compared with an AT representation is considerably less compared with other CG approaches.[17–19] Also, the simulation time step is not longer than in an AT MD simulation although this could still be possible by using bond constraints during the simulations in future studies. It should be emphasized that the main advantage of the CG representation for the nonbonded interactions is the generation of a smoother energy landscape compared with pure AT force fields. This allows more rapid structural transitions and conformational sampling of available states. One should also keep in mind that due to this smoother energy surface and the limited viscous dampening of motions, the time scales for transitions observed in the present simulations do not correspond to realistic time ranges.

It was shown that the methodology with the current force field parameterization preserves the folded geometry of three test cases to within ~3–4 Å of the native start structure in simulations of up to 100 ns. Most current "pure" CG force fields achieve similar results only in applications to one type of protein (e.g., only α-helical proteins) or upon including restraints, for example, to keep a certain secondary structure or to preserve a set of distances or contacts.[16,17] However, it should nevertheless be emphasized that the current parameterization is not useful to perform protein folding simulations since efforts to fold extended protein structures to near-native structures failed. Presumably, the native folded structure is only one of several subminima and therefore only locally stable. Note that the parameters for the ATTRACT CG model have been optimized for protein–protein docking with rigid or semiflexible partner structures and are not necessarily optimal for protein folding simulations.

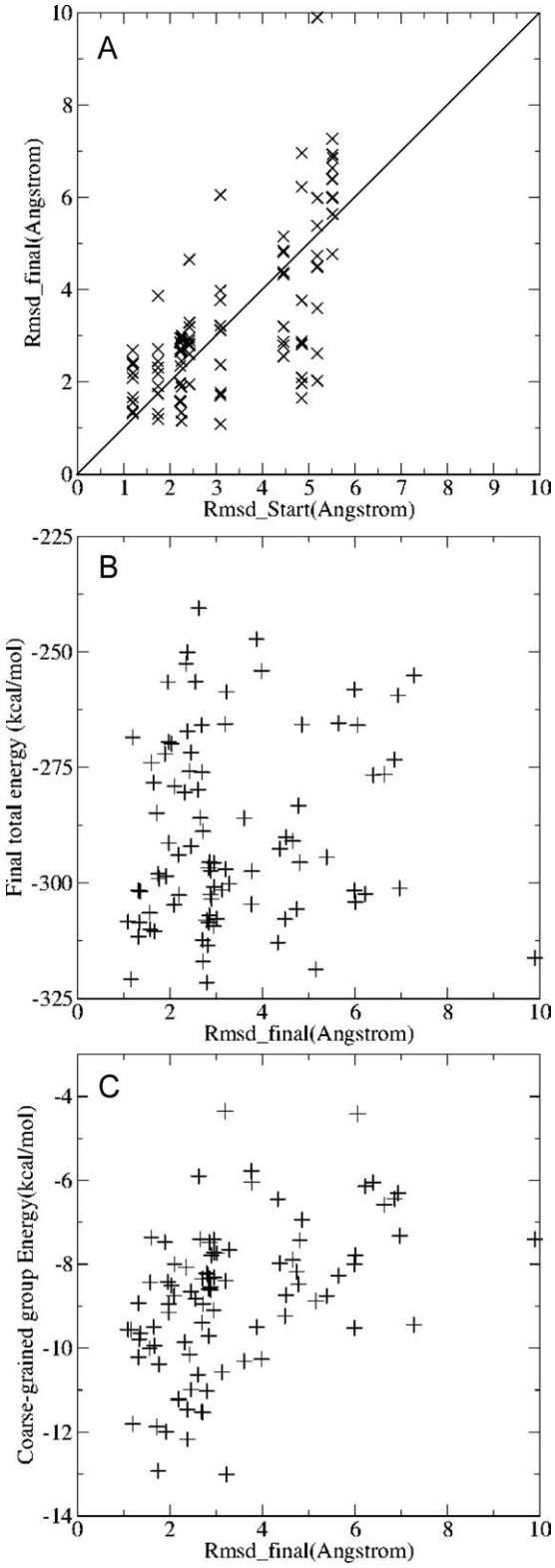

Figure 6

(**A**) Deviation ($Rmsd_{C\alpha}$) of final colicin E9/immunity protein Im9 complex structures after MD refinement of eight start complexes versus $Rmsd_{C\alpha}$ of the corresponding starting structures from the native complex (pdb1EMV). In each case, different starting velocities and/or slightly different simulation conditions were used followed by final energy minimization. The diagonal line corresponds to the limiting condition below which the refinement simulation resulted in an improved placement of the final structure compared with the starting complex. (**B**) Total energy of energy minimized final colicin E9/Im9 complexes versus $Rmsd_{C\alpha}$ from experiment (**C**) interaction energy of coarse-grained centers between protein partners after MD refinement and energy minimization.

The initial applications on the sampling of peptide–protein interaction geometries looked promising. During such simulated annealing simulations restraints to keep the peptide backbone near the binding geometry were necessary to avoid unfolding of the partners or strong deviation from a given reference structure. In many cases of practical interest, the assumption that the partner structures (or parts of it) do not undergo major backbone conformational changes upon binding is reasonable as long as small backbone adjustment and some side chain rearrangements are allowed. For a peptide–protein system, near-native binding geometries within an $Rmsd_{C\alpha} < 2$ Å were sampled even in some cases starting from docked placements of the peptide on the receptor surface far away from the native binding site (e.g., opposite side of the protein). The high initial search temperature allowed a rapid search on the surface of the receptor protein and identification of a possible binding region with modest computational demand. For a protein–protein test complex starting from initial placements of the smaller ligand protein that deviated from the native placement by an $Rmsd_{lig}$ between 5 and 10 Å resulted in sampling of configurations significantly closer to the native complex structures in some MD runs. However, further refinement of the force field to accurately distinguish between realistic and incorrect protein–protein interaction geometries is necessary in future studies as well as a subsequent test application to a large set of native and decoy protein–protein complexes.

ACKNOWLEDGMENTS

The author thanks Drs C. Beier, P. Setny, and S. de Vries for helpful discussions. This work was performed using supercomputer resources of the EMSL (Environmental Molecular Science Laboratories) at the PNNL (Pacific Northwest National Laboratories, USA).

REFERENCES

1. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. Annu Rev Biophys 2012;41:429–452.
2. Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253:694–698.
3. Liwo A, Khalili M, Scheraga HA. Ab initio simulations of protein folding pathways by molecular dynamics with united-residue model of polypeptide chains. Proc Natl Acad Sci USA 2005;102:2362–2367.
4. Gibson KD, Scheraga HA. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. Proc Natl Acad Sci USA 2005;102:7547–7552.
5. Ayton GS, Noid WG, Voth GA. Multiscale modeling of biomolecular systems: in serial and in parallel. Curr Opin Struct Biol 2007;17:192–198.
6. Tozzini V, Trylska J, Chang CE, McCammon JA. Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. J Struct Biol 2007;157:606–615.
7. Maupetit J, Tuffery P, Derreumaux P. A coarse-grained protein force field for folding and structure prediction. Proteins 2007;69:394–408.
8. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The Martini coarse-grained force field: extension to proteins. Chem Theory Comput 2008;4:819–834.
9. Enisco M, Rey A. Simple model for the simulation of peptide folding and aggregation with different sequences. J Chem Phys 2012;136:215103.
10. Potestio R, Pontiggia F, Micheletti C. Coarse-grained description of protein internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. Biophys J 2009;96:4993–5002.
11. DeVane R, Shinoda W, Moore PB, Klein ML. Transferable coarse grain nonbonded interaction model for amino acids. J Chem Theory Comput 2009;5:2115–2124.
12. Betancourt MR. Coarse-grained protein model with residue orientation energies derived from atomic force fields. J Phys Chem B 2009;113:14824–14830.
13. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. J Phys Chem A 2010;114:4471–4485.
14. Hills RD Jr, Lu L, Voth GA. Multiscale coarse-graining of the protein energy landscape. PLoS Comput Biol 2010;6:e1000827.
15. Kmiecik S, Kolinski A. Simulation of chaperonin effect on protein folding: a shift from nucleation-condensation to framework mechanism. J Am Chem Soc 2011;133:10283–10289.
16. Clementi C. Coarse-grained models of protein folding: toy models or predictive tools? Curr Opin Struct Biol 2008;18:10–15.
17. Orozco M, Orellana L, Hospital A, Naganathan AN, Emperador A, Carrillo O, Gelpí JL. Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings. Adv Protein Chem Struct Biol 2011;85:183–215.
18. Liwo A, He Y, Scheraga HA. Coarse-grained force field: general folding theory. Phys Chem Chem Phys 2011;13:16890–16901.
19. Ołdziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kaźmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Kamerlin SC, Vicatos S, Dryga A, Warshel A. Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. Annu Rev Phys Chem 2011;62:41–64.
20. Zimmermann MT, Leelananda SP, Gniewek P, Feng Y, Jernigan RL, Kloczkowski A Free energies for coarse-grained proteins by integrating multibody statistical contact potentials with entropies from elastic network models. J Struct Funct Genomics 2011;12:137–147.
21. Gniewek P, Leelananda SP, Kolinski A, Jernigan RL, Kloczkowski A. Multibody coarse-grained potentials for native structure recognition and quality assessment of protein models. Proteins 2011;79:1923–1929.
22. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268:209–225.
23. Zacharias, M. Protein-protein docking using a reduced model. Protein Sci 2003;12:1271–1282.
24. Fiorucci S, Zacharias M. Binding site prediction and improved scoring during flexible protein-protein docking with ATTRACT. Proteins 2010;78:3131–3139.
25. van Gunsteren WF, Berendsen HJC. Groningen molecular simulation GROMOS. Library Manual, BIOMOS b.v, University of Groningen, Groningen: The Netherlands, 1987.
26. Andersen HC. Molecular dynamics simulations at constant pressure and/or temperature. J Chem Phys 1980;72:2384–2393.
27. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.

28. Zacharias, M. Accounting for conformational changes during protein-protein docking. Curr Opin Struct Biol 2010;16:194–200.
29. Levine AJ. P53, the cellular gatekeeper for growth and division. Cell 1997;88:323–331.
30. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP. Structure of the MDM2 oncoprotein bound to the p 53 tumor suppressor transactivation domain. Science 1996;274: 948–953.
31. Spink KE, Polakis P, Weis WI. Structural basis of the Axin-adenomatous polyposis coli interaction. EMBO J 2000;19:2270–2279.
32. May A, Zacharias M. Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. Proteins 2008,70:794–809.
33. Schneider S, Zacharias M. Scoring optimisation of unbound protein-protein docking including protein binding site predictions. J Mol Recognit 2012;25:15–23.
34. Li Y, Cortés J, Siméon T. Enhancing systematic protein-protein docking methods using ray casting: application to ATTRACT. Proteins 2011;79:3037–3049.
35. Sacquin-Mora S, Carbone A, Lavery R. Identification of protein interaction partners and protein-protein interaction sites. J Mol Biol 2008;382:1276–1289.
36. Kuhlmann UC, Pommer AJ, Moore GR, James R, Kleanthous C. Specificity in protein-protein interactions: the structural basis for dual recognition in endonuclease colicin-immunity protein complexes. J Mol Biol 2000;301:1163–1178.