Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations

Philippe Ferrara, Joannis Apostolakis, and Amedeo Caflisch*
Department of Biochemistry, University of Zürich, Zürich, Switzerland

ABSTRACT **A solvation term based on the solvent accessible surface area (SASA) is combined with the CHARMM polar hydrogen force field for the efficient simulation of peptides and small proteins in aqueous solution. Only two atomic solvation parameters are used: one is negative for favoring the direct solvation of polar groups and the other positive for taking into account the hydrophobic effect on apolar groups. To approximate the water screening effects on the intrasolute electrostatic interactions, a distance-dependent dielectric function is used and ionic side chains are neutralized. The use of an analytical approximation of the SASA renders the model extremely efficient (i.e., only about 50% slower than *in vacuo* simulations). The limitations and range of applicability of the SASA model are assessed by simulations of proteins and structured peptides. For the latter, the present study and results reported elsewhere show that with the SASA model it is possible to sample a significant amount of folding/unfolding transitions, which permit the study of the thermodynamics and kinetics of folding at an atomic level of detail. Proteins 2002;46:24–33.** © 2001 Wiley-Liss, Inc.

Key words: implicit solvent; solvent accessible surface area; CHARMM force field; protein folding

INTRODUCTION

The very demanding computational requirement of molecular dynamics (MD) in explicit water prohibits the simulations of large conformational transitions between states separated by energy barriers of even a few kTs. The omission of the solvent degrees of freedom and interaction centers reduces drastically the required computational power, but simulations *in vacuo* are well known to suffer from serious artifacts, such as an excessive deviation from the native conformation and a too large number of hydrogen bonds. A way of reducing the computational cost without sacrificing the accuracy of the results is to incorporate the properties of the solvent into the energy function by means of a potential of mean force. Many approaches have been proposed to derive such potentials (see Refs. 1 and 2 for a review).

Accurate continuum dielectric models based on the finite difference solution of the Poisson or Poisson-Boltzmann equations have been widely used.[3–5] These methods can provide the electrostatic forces on atoms[6,7] and can therefore be used in MD simulations.[8–11] However, finite difference approaches are still too slow for a wide range of biochemical applications. A simpler and faster continuum dielectric approximation is the so-called generalized Born (GB) model.[12] Numerical implementations are more accurate than analytical ones, but cannot be used efficiently for MD simulations.[13,14] Analytical solutions for the GB model have been proposed[15–17] and used to determine the thermodynamic properties of structured peptides[18,19] and to simulate conformational transitions of 10-base pair duplexes of DNA and RNA.[20] In a more simplified class of implicit models the mean solvation term is assumed to be proportional to the solvent accessible surface area (SASA)[21] or accessible volume.[22] Many parameterizations have been proposed in the past.[22–26] Two important issues are how many solvation parameters should be used and how to calibrate them.

Here an implicit solvent model based on a very efficient analytical evaluation of the SASA is combined with the CHARMM force field.[27] The present work is based on the EEF1 model proposed by Lazaridis and Karplus[26] for dielectric shielding due to the solvent and the surface area approximation for the hydrophobic effect introduced by Eisenberg and McLachlan.[21] Electrostatic screening effects are approximated by a distance-dependent dielectric function and a set of partial charges with neutralized ionic groups.[26] Because an exact analytical or numerical computation of the SASA is too slow to compete with simulations in explicit solvent, we make use of an approximate analytical expression.[28] This drastically reduces the computational cost with respect to an explicit solvent simulation. The model used here is based on the assumptions that most of the solvation energy arises from the first water shell around the protein[21] and that two atomic solvation parameters are sufficient to describe these effects at a qualitative level of accuracy.[25] Within these assumptions, the SASA energy term approximates the solute-solvent interactions [i.e., it should account for the energy of cavity

Grant sponsor: Swiss National Science Foundation.

J. Apostolakis' present address is GMD-SCAI, Schloss Birlinghoven, D-53754, St. Augustin, Germany.

*Correspondence to: A. Caflisch, Department of Biochemistry, University of Zurich, Winterthurstrasse 190, CH-8507 Zurich, Switzerland. E-mail: caflisch@bioc.unizh.ch

Received 5 April 2001; Accepted 19 July 2001

DOI 10.1002/prot.10001

formation, solute-solvent dispersion interactions, and the direct (or Born) solvation of polar groups]. The two atomic solvation parameters were optimized by performing 1-ns MD simulations at 300 K on six small proteins. The results are compared with those obtained by the *vacuo* energy function and EEF1, a previously published implicit solvent approximation based on a Gaussian solvent-exclusion model for the solvation free energy.[26] The comparison is extended to two 20-residue three-stranded antiparallel β-sheets not used for the calibration of the SASA atomic solvation parameters. Finally, the limitations of the SASA model are shown for two designed proteins. Taken together with previous applications, the present study indicates that the SASA model is a realistic approximation for small proteins in aqueous solution and is particularly useful for simulating large conformational transitions in structured peptides.

MODEL AND METHODS

Description of the Model

In most empirical force fields, the Hamiltonian of the solute-solvent system is additive and consists of the sum of solute-solute, solute-solvent, and solvent-solvent terms. After integration over the solvent coordinates, the potential of mean force $W(\mathbf{r})$, or effective energy, can be divided into two contributions:

$$W(\mathbf{r}) = V_{\text{solute}}(\mathbf{r}) + V_{\text{solvation}}(\mathbf{r}) \tag{1}$$

for a solute having M atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_M)$. The term *effective energy* for $W(\mathbf{r})$ is used here as in Ref. 26; it is the sum of intra-solute and mean solvation terms. In the present study, we assume that the mean solvation energy is linearly related to the SASA of the solute:

$$V_{\text{solvation}}(\mathbf{r}) = \sum_{i=1}^{M} \sigma_i A_i(\mathbf{r}) \tag{2}$$

where σ_i and $A_i(\mathbf{r})$ are the atomic solvation parameter and SASA of atom i, respectively. $A_i(\mathbf{r})$ is computed by an approximate analytical expression:[28]

$$A_i(\mathbf{r}) = S_i \prod_{j \neq i}^{M} [1 - p_i p_{ij} b_{ij}(r_{ij})/S_i] \tag{3}$$

where S_i denotes the SASA of an isolated atom i of radius R_i:

$$S_i = 4\pi(R_i + R_{\text{probe}})^2 \tag{4}$$

and R_{probe} is the radius of the solvent probe. In Eq. 3, $b_{ij}(r_{ij})$ represents the SASA removed from S_i due to the overlap between atoms i and j separated by a distance $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, and is given by

$$b_{ij}(r_{ij}) = \begin{cases} 0 & r_{ij} > R_i + R_j + 2R_{\text{probe}} \\ \pi(R_i + R_{\text{probe}})(R_i + R_j + 2R_{\text{probe}} - r_{ij})[1 + (R_j - R_i)r_{ij}^{-1}] & \text{otherwise} \end{cases} \tag{5}$$

Using 270 small molecules, the atom type parameters p_i and connectivity parameters p_{ij} have been optimized to reproduce the exact SASA with $R_{\text{probe}} = 1.4$ Å.[28] The parameter p_{ij} is 0.8875 if the atoms i and j are covalently bonded and 0.3516 otherwise. The values of R_i's and p_i's according to Hasel et al.[28] and the σ_i's optimized in this work are listed in Table I.

The SASA model includes the free energy cost of burying a charged residue in the interior of a protein. However, it does not take into account the solvent screening on the interactions between solute charges. This effect is approximated here using a distance-dependent dielectric function, $\epsilon(r) = 2r$. $\epsilon(r) = 2r$ was chosen instead of $\epsilon(r) = r$ mainly to reduce the strength of the hydrogen bonds. Larger values of the dielectric constant are expected to lead to partial unfolding of proteins in room temperature simulations. Because a cutoff for long-range interactions is used (see below), a linear distance-dependent dielectric function does not differ significantly from a more sophisticated one, such as a sigmoidal function,[29–31] because the deviation from linearity is negligible for distances smaller than 10 Å.[31] A distance-dependent dielectric function is a very simplified way of accounting for the solvent polarization effects on the solute. In particular, the screening of the electrostatic interactions between charged groups is insufficient, as shown by the formation of too stable salt bridges in MD simulations of the RGDW peptide.[32] The limitations of this approximation can be partly overcome by using a set of partial charges with a zero total charge for every residue. Here, the ionizable amino acids were neutralized as in Ref. 26.

The solvation model has been implemented in CHARMM and is used with a united-atom parametrization (PARAM19), where the only modified parameters are the partial charges of the ionic side chains. The CHARMM PARAM19 default cutoffs for long-range interactions were used (i.e., a shift function[27] was used with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms). This cutoff length was chosen to be consistent with the original parametrization of CHARMM PARAM19.[27] Even though the SASA solvation term is calculated at every step of dynamics, the CPU time required for simulations with SASA is only about 50% larger than that for a vacuum simulation with the same cutoff (7.5 Å).

Simulation Protocols

The temperature was kept constant by weak coupling to an external bath[33] with a coupling constant of 0.5 ps for the simulations of native proteins and 5 ps for the three-stranded antiparallel β-sheets simulations. The SHAKE algorithm was used to fix the length of the covalent bonds where hydrogen atoms are involved, which allows an

TABLE I. CHARMM Atom Types and Mean Solvation Parameters

Atom type	R_{min}^{vdW} [a] (Å)	R_i [b] (Å)	p_i [b]	σ_i [c] (kcal mol^{-1} Å^{-2})	Description
C	2.1	1.72	1.554	0.012	Carbonyl carbon
CH1E	2.365	1.80	1.276	0.012	Extended aliphatic carbon with 1 hydrogen
CH2E	2.235	1.90	1.045	0.012	Extended aliphatic carbon with 2 hydrogens
CH3E	2.165	2.00	0.880	0.012	Extended aliphatic carbon with 3 hydrogens
CR1E	2.1	1.80	1.073	0.012	Extended aromatic carbon with 1 hydrogen
NH1	1.6	1.55	1.028	−0.060	Amide nitrogen
NR	1.6	1.55	1.028	−0.060	Aromatic nitrogen with no hydrogens
NH2	1.6	1.60	1.215	−0.060	Nitrogen bound to two hydrogens
NH3	1.6	1.60	1.215	−0.060	Nitrogen bound to three hydrogens
NC2	1.6	1.55	1.028	−0.060	Guanidinium nitrogen
N	1.6	1.55	1.028	−0.060	Proline nitrogen
OH1	1.6	1.52	1.080	−0.060	Hydroxyl oxygen
O	1.6	1.50	0.926	−0.060	Carbonyl oxygen
OC	1.6	1.70	0.922	−0.060	Carboxyl oxygen
S	1.89	1.80	1.121	0.012	Sulphur
SH1E	1.89	1.80	1.121	0.012	Extended sulphur with 1 hydrogen
H	0.8	1.10	1.128	0.000	Polar hydrogen
HC	0.6	1.10	1.128	0.000	Polar hydrogen (in Arg, Lys and N-term)

[a] The CHARMM PARAM19 van der Waals radii are given as a basis of comparison but are not used in the solvation term.
[b] Solvation parameters according to Ref. 28.
[c] Optimal σ parameters (this work).

integration time step of 2 fs. The nonbonded interaction list was updated every 10 dynamics steps, and coordinate frames were saved every 4 and 10 ps for the simulations of native proteins and three-stranded antiparallel β-sheets, respectively.

Analysis

A reasonably accurate effective energy function should avoid the main problem which arises in *in vacuo* simulations (i.e., too large deviations from the native conformation). One quantity of interest is therefore the C_α root mean square deviation (RMSD) from the native state, which is calculated after optimal superposition using a least-squares fit of C_α atom coordinates. The behavior of the number of hydrogen bonds is also analyzed. A geometric criterion for the presence of a hydrogen bond is used: the hydrogen-acceptor distance must be smaller than 2.6 Å and the angle formed by the donor-hydrogen · · · acceptor triplet larger than 120°.

The results obtained by the SASA model are compared with three other approximations: the *vacuo* force field (VAC), the *vacuo* force field with neutralized ionic side chains (VACNEUT), and EEF1.[26] The latter is an implicit model that approximates the effects of an aqueous solvent by a Gaussian term for the first solvation shell. EEF1 also uses a distance-dependent dielectric function [$\epsilon(r) = r$] and neutralization of the ionic side chains.[26]

RESULTS

Simulations of Proteins Starting from the Native State

Previous studies have shown that implicit solvent models based on the SASA can be used in MD simulations of different proteins to avoid the main difficulties that arise in *in vacuo* simulations.[25] Following that work, we considered only two σ parameters: one for carbon and sulfur atoms ($\sigma_{C,S} > 0$), and one for nitrogen and oxygen atoms ($\sigma_{N,O} < 0$). The solvation parameter of the hydrogen atoms is set to zero. The two σ parameters were optimized in this work by a trial-and-error approach, performing 1-ns MD simulations at 300 K on six small proteins (Tables II and III). In spite of the fact that the model is now based on a different force field (CHARMM instead of GROMOS[25]) and that it uses a distance-dependent dielectric for the electrostatic interaction (a constant dielectric of 1.0 was used in Ref. 25), we have found that the two σ values that give the minimal C_α RMSD correspond to those determined previously by others[25] (0.012 kcal mol^{-1} Å^{-2} for carbon and sulfur atoms, −0.060 kcal mol^{-1} Å^{-2} for nitrogen and oxygen atoms). With these values of σ, the C_α RMSD from the native conformation averaged over the six proteins and the 0.5–1.0-ns time interval is 1.79 Å (2.00 Å including flexible loops and termini; Table II), which is satisfactory. Table II shows that already the neutralization of the ionizable groups (VACNEUT) decreases significantly the deviations from the native conformation compared with the *vacuo* simulations with the formal charges on Asp, Glu, Arg, and Lys side chains. More recently, 10-ns simulations were performed at 300 K on the same set of proteins. With respect to the 1-ns trajectories, higher C_α RMSDs from the native conformation are observed for the four energy functions, which shows that relatively long (at least a few nanoseconds) MD simulations are needed to validate an energy function. Moreover, the deviations obtained by the SASA and EEF1 models are comparable to the ones obtained by VACNEUT, which indicates that the neutralization of the ionic side chains is responsible for

TABLE II. Deviation from the Native Conformation during MD Simulations at 300 K[†]

		VAC[c]		VACNEUT[d]		EEF1[e]		SASA[f]	
PDB code[a]	Nres[b]	$\langle\ \rangle_{0.5-1}$[g]	$\langle\ \rangle_{5-10}$[h]	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$
1crn	46	1.27 (1.43)	1.53 (1.62)	1.01 (1.33)	1.17 (1.43)	1.05 (2.47)	1.12 (1.62)	1.19 (1.50)	1.46 (2.87)
1bpi	58	2.33	3.43	2.82	2.71	2.05	3.89	1.89	3.47
2ci2	64	2.13 (2.56)	2.74 (2.99)	1.50 (1.82)	1.78 (1.91)	1.54 (1.69)	1.31 (1.62)	1.77 (1.97)	1.41 (1.68)
1ubq	76	2.05 (2.90)	2.37 (3.39)	1.83 (2.92)	2.74 (4.10)	1.36 (1.89)	1.84 (2.02)	1.70 (1.79)	1.93 (2.61)
1pht	83	2.81 (3.36)	2.97 (3.48)	1.94 (3.28)	2.79 (4.17)	2.44 (3.08)	2.45 (2.91)	1.60 (2.26)	1.79 (2.52)
1hdn	85	2.63	3.02	1.85	2.38	2.27	3.34	2.59	3.52
$\langle\ \rangle_{\text{proteins}}$		2.20 (2.54)	2.68 (2.99)	1.83 (2.34)	2.26 (2.78)	1.79 (2.24)	2.33 (2.57)	1.79 (2.00)	2.26 (2.78)

[†]The C_α RMSD (Å) from the native conformation of different proteins obtained using the *vacuo* energy function and three solvation models is shown. It was computed using all the residues, except for the C-terminus 41–46 in 1crn, the loop 34–44 in 2ci2, the C-terminus 73–76 in 1ubq, the N-terminus 1–2 and the C-terminus 81–83 in 1pht. The corresponding values obtained using all the residues are given in parentheses.
[a]1crn: crambin, 1bpi: trypsin inhibitor, 2ci2: chymotrypsin inhibitor 2, 1ubq: ubiquitin, 1pht: SH3 domain of the p85α subunit of bovine phosphatidylinositol 3-kinase, 1hdn: histidine-containing phosphocarrier protein.
[b]Number of residues.
[c]VAC, *vacuo* force field.
[d]VACNEUT, *vacuo* force field with the ionic side chains neutralized.
[e]Effective energy function described in Ref. 26.
[f]SASA: Eq. 2, with the σ's specified in Table 1, $\epsilon(r) = 2r$, and the ionic side chains neutralized.
[g]Average over the 0.5–1.0-ns time interval.
[h]Average over the 5–10-ns time interval.

TABLE III. Number of Hydrogen Bonds during MD Simulations at 300 K[†]

		VAC		VACNEUT		EEF1		SASA	
PDB code	X-ray	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$	$\langle\ \rangle_{0.5-1}$	$\langle\ \rangle_{5-10}$
1crn	37 (23)	47 (24)	50 (25)	44 (28)	44 (28)	42 (24)	42 (26)	31 (21)	34 (22)
1bpi	36 (23)	75 (22)	75 (22)	57 (33)	57 (31)	54 (26)	60 (25)	42 (22)	41 (21)
2ci2	46 (30)	91 (29)	92 (26)	60 (36)	60 (36)	56 (36)	60 (34)	44 (28)	45 (29)
1ubq	63 (44)	108 (39)	108 (36)	84 (48)	77 (44)	80 (42)	79 (43)	61 (37)	60 (38)
1pht	66 (45)	120 (29)	122 (28)	84 (45)	87 (41)	88 (47)	88 (40)	66 (41)	66 (39)
1hdn	70 (51)	107 (47)	109 (42)	95 (59)	95 (56)	87 (49)	93 (47)	69 (48)	66 (43)

[†]The number of hydrogen bonds is shown with the number of backbone-backbone hydrogen bonds in parentheses. A hydrogen bond is present when the O···H distance is smaller than 2.6 Å and the N—H···O angle is larger than 120°. See caption of Table II for further explanations.

most of the improvement of the SASA and EEF1 models with respect to the *vacuo* force field.

Simulations *in vacuo* suffer also from too small atomic fluctuations. Figure 1 shows the RMS fluctuations of C_α atoms (C_α RMSF) in the simulations of chymotrypsin inhibitor 2. Apart from the EEF1 simulation, the C_α RMSF for the first nanosecond and the whole simulation are rather different. The average C_α RMSF over the whole simulation are 0.85, 0.81, 0.92, and 1.22 Å with VAC, VACNEUT, EEF1, and SASA, respectively. The C_α RMSF obtained from the simulations are compared with those derived from the crystallographic *B*-factors using the relationship C_α RMSF = $(3B_{C_\alpha}/8\pi)^{1/2}$. The loop (residues 34–44) is the region of highest mobility (excluding the N-terminus). This is observed only in the SASA simulation, although the C_α RMSF in this region are somewhat overestimated by this model.

Another quantity of interest is the number of hydrogen bonds. This is particularly important for folding/unfolding simulations. The strength of a hydrogen bond should be reasonably described to obtain meaningful energetics along the folding process. Furthermore, the global minimum of the effective energy should represent the native conformation. The total number of hydrogen bonds is too large in *vacuo* (Table III). As the number of backbone hydrogen bonds is even smaller than in the native state, this discrepancy originates from the side chains at the surface that form intra-solute hydrogen bonds. Part of this problem is solved by neutralizing the charged side chains. EEF1 yields a number of hydrogen bonds, which is slightly lower than with VACNEUT but 34% higher on average than in the native conformation. The number of hydrogen bonds in the X-ray structure is best reproduced by the SASA model. This indicates that the strength of the hydrogen bond is described reasonably. However, the SASA model slightly underestimates the number of backbone-backbone hydrogen bonds, and some side chains form intra-solute hydrogen bonds that are not present in the native state.

Three-stranded Antiparallel β-sheets Simulations

As a further evaluation of the SASA model, two designed three-stranded antiparallel β-sheet peptides were studied.[34,35] They consist of 20 residues each and have a sequence identity of 15%. One peptide (called hereafter GS peptide) has Gly-Ser at both turns,[36] whereas the other

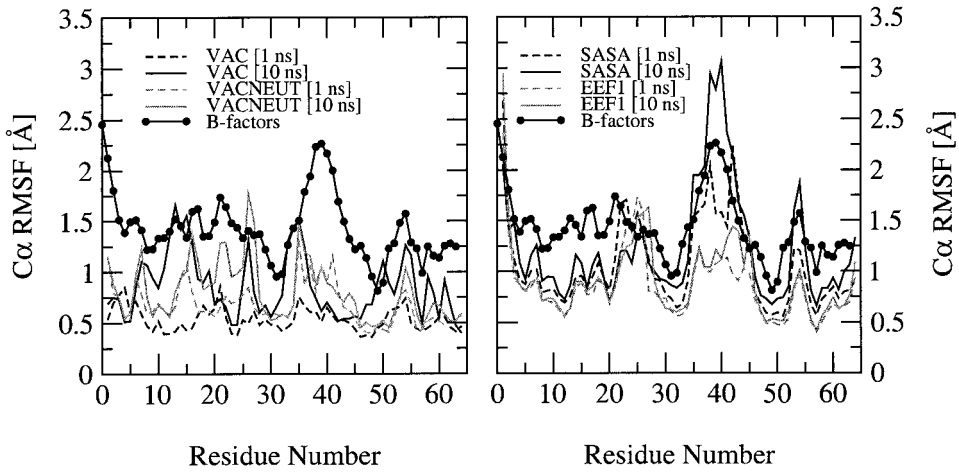

Fig. 1. RMS fluctuations of the C_α atoms of CI2 from the average 300 K conformation as a function of residue number. Left: VAC simulation (black lines) and VACNEUT simulation (gray lines). Right: SASA simulation (black lines) and EEF1 simulation (gray lines). The dashed and solid lines represent the fluctuations averaged over the first nanosecond and the whole 10-ns simulation, respectively. The bold line with circles represents the fluctuations derived from the crystallographic *B*-factors.[50]

(called hereafter ^{D}PG peptide) has D-Pro-Gly.[37] Circular dichroism and chemical shift data have provided evidence that GS and ^{D}PG adopt the expected three-stranded antiparallel β-sheet conformation in aqueous solution. Moreover, both peptides were shown to be monomeric in aqueous solution by equilibrium sedimentation. For the GS peptide, the β-sheet population was 13–31% based on NOE intensities and 30–55% based on the chemical shift data.[36] For ^{D}PG the percentage of β-sheet population was not estimated, but Nuclear Overhauser enhancements (NOEs) indicate that both hairpins are highly populated at 24°C.[37]

A detailed analysis of the MD results obtained by the SASA model for the GS and ^{D}PG peptides has been presented elsewhere.[34,35] Briefly, it was shown that the two peptides fold reversibly to the correct NMR conformation, irrespective of the starting conformation. They have similar free energy surfaces and folding pathways that involve first the almost complete formation of either the 1–2 β-hairpin (i.e., the hairpin consisting of strands 1 and 2) or the 2–3 β-hairpin, followed by consolidation of the unstructured strand. However, for the GS peptide, the most predominant pathway involves first the formation of the 2–3 β-hairpin, whereas for the ^{D}PG peptide the statistical predominance is opposite. It is interesting to note that NMR data of another designed 24-residue three-stranded antiparallel β-sheet peptide indicate the presence of two folding pathways starting by the almost complete formation of either of the two β-hairpins.[38]

It is useful to compare the results obtained by SASA with VAC, VACNEUT, and EEF1. For this purpose, trajectories were generated at 300 K and 360 K starting from the antiparallel three-stranded β-sheet and at 360 K with the completely extended conformation as initial structure (Figs. 2 and 3). The simulations were run for 200 and 100 ns for the GS and ^{D}PG peptides, respectively.

The trajectories at 300 K started from the folded state were used to calculate average inter-proton distance violations, d_{viol}, which are defined as $\langle r(t)^{-6}\rangle^{-1/6} - r_{exp}$, where $r(t)$ is the inter-proton distance at simulation time t, r_{exp} is the NOE upper distance limit, and $\langle\ \rangle$ represents a time average. The smallest number of violations are observed in the SASA trajectory for the GS peptide and the SASA and EEF1 runs for the ^{D}PG peptide (Table IV). The fraction of native contacts Q is a progress variable used often to monitor folding.[39] Because the SASA trajectories at 300 K yield the results closest to the NMR data, they were used to define the native contacts. There are 10 native hydrogen bonds in each of the two peptides, and 16 and 19 native side chain contacts for the GS and ^{D}PG peptide, respectively.[34,35] It is preferable to use the contacts populated along the 300 K MD trajectory because the small number of NOEs and their inhomogeneous distribution along the sequence do not allow for a clear definition of the folded structure.[36]

For the GS peptide the C_α RMSD between the average structure over the whole 200-ns simulation at 300 K and the average NMR conformation is 1.7, 1.8, 1.9, and 2.8 Å using VAC, VACNEUT, SASA, and EEF1, respectively. Despite a low C_α RMSD value, eight NOE restraints are violated by more than 1 Å in the VAC simulation (Table IV). The time evolution of Q differs clearly in the simulations with VAC with respect to VACNEUT and SASA [Fig. 2(A, D)]. With VAC, a Q value of about 0.5 is obtained at the end of the simulation. This is due to the loss of contacts between strands 2 and 3, most of which break either very early or at about 100 ns. Figure 2(A) and Table IV show that the neutralization of the ionic side chains suffices to solve part of this problem. The average Q value is about 0.9 and 0.8 with VACNEUT and SASA, respectively. With EEF1, most of the native contacts between strands 1 and 2 in GS broke at about 33 ns, and the peptide stays partly

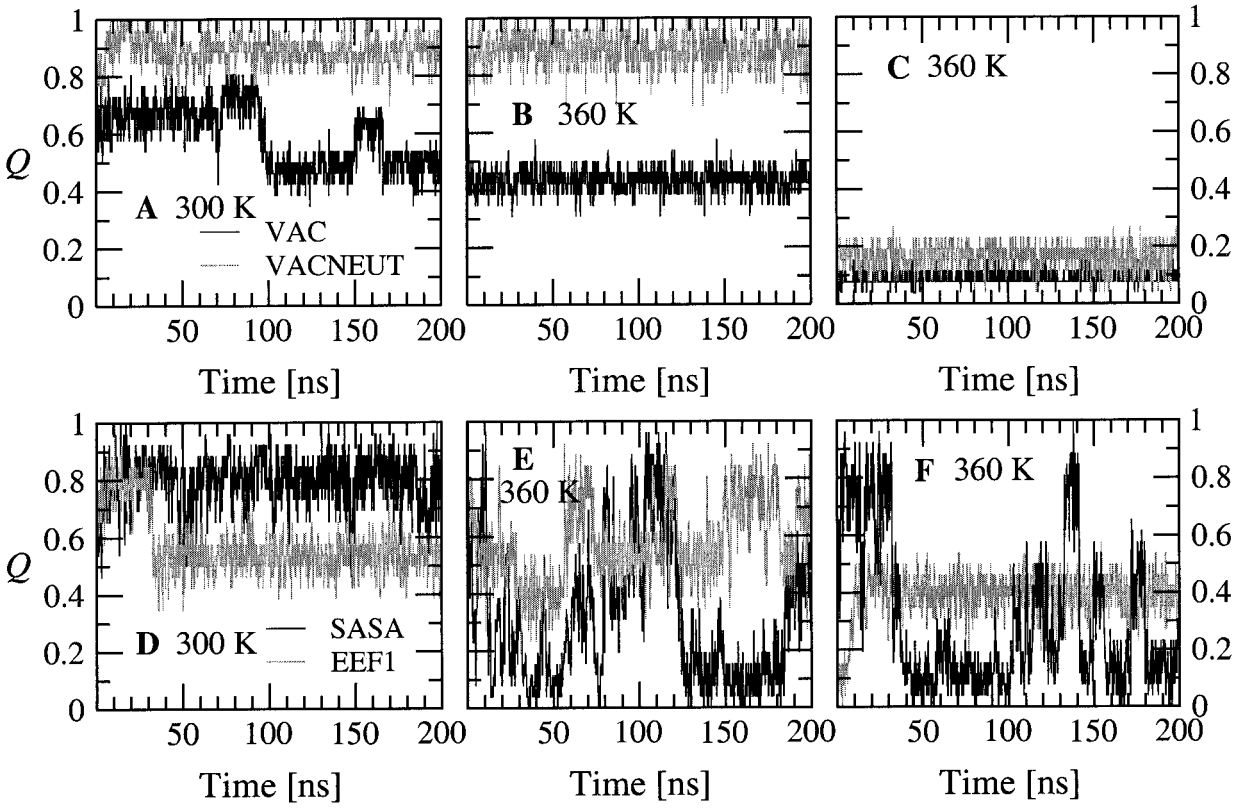

Fig. 2. Time dependence of the fraction of native contacts Q in the GS peptide simulations. (**A, D**) The 300 K simulation started from the folded conformation. (**B, E**) The 360 K simulation started from the folded conformation. (**C, F**) The 360 K simulation started from the extended conformation. The VAC and VACNEUT simulations are shown in the panels (**A**) through (**C**) with black and gray lines, respectively. Panels (**D**) through (**F**) show the SASA and EEF1 simulations with black and gray lines, respectively.

unfolded until the end of the simulation. Moreover, most of the NOEs are violated in the EEF1 trajectory (Table IV). At 360 K with VAC starting from the folded state, GS unfolds at the beginning of the simulation and does not fold again [Fig. 2(B)]. The evolution of Q with VACNEUT is very similar at 300 and 360 K. Although the temperature dependence of the β-sheet population has not been measured experimentally, it is likely that this result is an artifact of the *vacuo* force field, because the percentage of folded structures is expected to decrease with increasing temperature. Several partial unfolding and refolding events are observed with EEF1, but the fully unfolded state is never reached. Only with SASA does the peptide undergo a series of complete folding ($Q > 0.8$) and unfolding ($Q < 0.2$) transitions. Most of these happen in the first 100 ns, whereas in the second half of the simulation the folded state is present from about 105 to 120 ns, and GS is unfolded from about 125 to 200 ns. When the simulations were initiated from the extended conformation, no folding event was observed with VAC, VACNEUT, and EEF1 [Fig. 2(C, F)]. A conformation, where most of the contacts between strands 2 and 3 are formed and most of the interactions between the first and second strands are broken, was reached at around 10 ns with EEF1 and was then stable until the end of the simulation. With the SASA model, the peptide folded into an antiparallel three-stranded β-sheet at about 1 ns and stayed folded for nearly 30 ns. It was partly or completely unfolded until the end of the simulation, except between 132 and 141 ns. In the two folding transitions one observes first formation of most of the contacts between strands 2 and 3, followed by the association of the N-terminal strand onto the preformed 2–3 β-hairpin. The two unfolding transitions are essentially the reverse of the two folding events.

For the ^{D}PG peptide the results obtained by VAC, VACNEUT, and SASA are similar to the ones obtained for the GS peptide (Fig. 3 and Table IV). The evolution of Q with EEF1 at 300 K is comparable to the one obtained by VACNEUT and SASA. However, it is very similar at 300 and 360 K. Starting from the extended conformation, the folded state is not sampled using VAC or VACNEUT, whereas it is reached in about 42 ns with EEF1. In this folding event, the formation of most of the contacts between strands 2 and 3 precedes the appearance of the 1–2 interstrand contacts.

In summary, these results show that the neutralization of the charged residues contribute significantly to stabilize the native fold of the two peptides at room and elevated temperatures. However, the three-stranded antiparallel β-sheet was never reached in the VACNEUT simulations

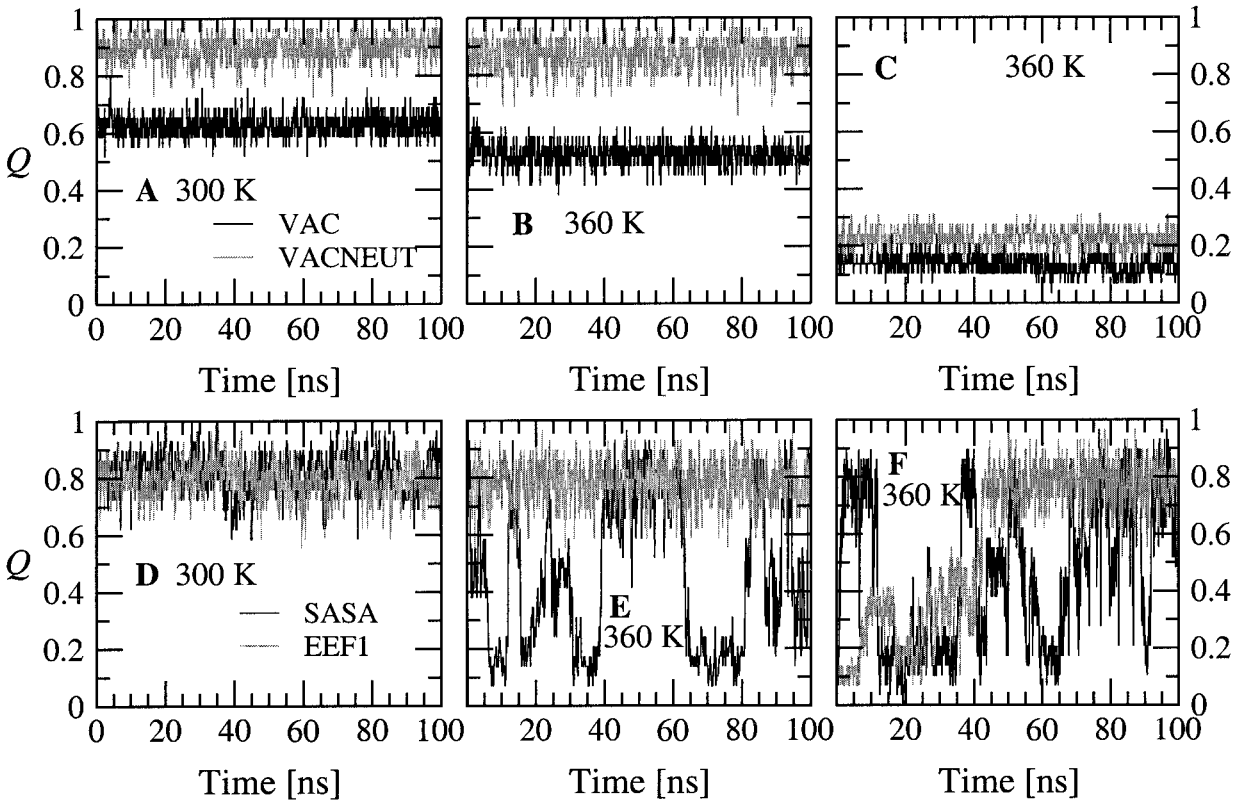

Fig. 3. Time dependence of the fraction of native contacts Q in the ^{D}PG peptide simulations. Panels (**A**) through (**F**) are as described in Figure 2.

TABLE IV. Average Inter-proton Distance Violations from the MD Simulations at 300 K†

	GS peptide (200 ns)				^{D}PG peptide (100 ns)			
	VAC	VACNEUT	SASA	EEF1	VAC	VACNEUT	SASA	EEF1
$d_{viol} < 0.0$ Å	11	13	16	6	32	33	36	36
$d_{viol} < 1.0$ Å	7	9	7	9	4	6	4	6
$d_{viol} > 1.0$ Å	8	4	3	11	8	5	4	2

†d_{viol} represents the average inter-proton distance violation, which is defined as $\langle r(t)^{-6}\rangle^{-1/6} - r_{exp}$, where $r(t)$ is the inter-proton distance at simulation time t, r_{exp} is the NOE upper distance limit, and $\langle\ \rangle$ represents a time average.

started from the extended conformation. Only the SASA model is able to simulate the reversible folding of the two peptides, whereas the EEF1 model seems to get trapped even at elevated temperature.

Limitations of the SASA Implicit Solvent Model

The SASA model is not expected to describe correctly the stability of large proteins. One of the reasons is that in this approximation the screening between partial charges does not depend on the environment. Starting from the X-ray conformation, 1-ns simulations were performed at 300 K on barnase (1a2p, 110 residues), hen egg-white lysozyme (1hel, 129 residues), and cutinase (1cus, 197 residues). The C_α RMSD from the native conformation averaged over the 0.5–1.0-ns time interval was 3.81, 3.53, and 4.20 Å, respectively. The deviations are not localized in the termini or flexible loops. To reduce such deviations, a longer nonbonding cutoff and an explicit dependence of the screening of the partial charges on the surrounding volume (low/high dielectric) could be considered.

For small systems, the limitations of the SASA implicit solvent model are exemplified by two designed miniprotein motifs, BBA5 and αtα. BBA5 is a 23-residue peptide with a ββα architecture,[40] whereas αtα is a 38-residue peptide designed to adopt a helical hairpin conformation in aqueous solution.[41] Simulations were performed at 280 K starting from the NMR conformation. For BBA5 the C_α RMSD from the native state reaches rapidly 5 Å and fluctuates around this value until the end of the simulation (Fig. 4). The β-hairpin was very stable, with an average C_α RMSD value of 1.0 Å, whereas the α-helix showed a significant deviation from the native α-helix, with an average C_α RMSD of 1.9 Å. However, most of the deviation from the NMR structure originates from the loss of the packing of the hairpin against the helix. The percentage of α-helicity for the residues 11–19 calculated

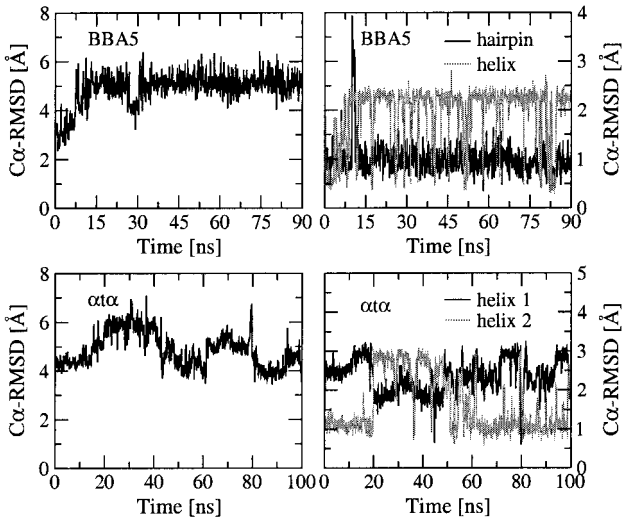

Fig. 4. Evolution of the C_α RMSD from the native state as a function of time. Top: BBA5: left, all residues; right, β-hairpin (residues 1–8) and α-helix (residues 11–19). Bottom: αtα: left, all residues; right, α-helix 1 (residues 3–15) and α-helix 2 (residues 23–33).

over the whole simulation was 25%, whereas the percentage of π-helicity was 52%. This is an unexpected result because π-helices are very rare in proteins. A π-helix is favored with respect to an α-helix by van der Waals side chain interactions and disfavored by electrostatic interactions. For BBA5, it is difficult to say if the unnatural π-helix propensity is the origin or the consequence of the loss of the packing between the β-hairpin and the α-helix. An unusually high (26% at 300 K) percentage of π-helicity was observed in simulations of the helical peptide $Y(MEARA)_6$.[42]

For αtα a C_α RMSD value of 4 Å was obtained after heating and equilibration. It fluctuated between 4 and 6 Å along the simulation. Of the two α-helices the second one was more stable, with an average value of 1.6 Å for the C_α RMSD (2.3 Å for the first helix). As for BBA5, the loss of the packing between the two helices is at the origin of most of the deviation from the NMR structure. The percentages of α-helicity and π-helicity were 34% and 35% for helix 1, respectively, and 57% and 19% for helix 2. In summary, the tertiary structure was not preserved in both proteins and the helices showed a too large presence of π-hydrogen bonds.

As a basis of comparison, simulations of BBA5 and αtα with VACNEUT and EEF1 gave also C_α RMSDs between 3.5 and 5 Å, apart from BBA5 with EEF1 that deviated by 3.1 Å over a 100-ns run at 280 K. For BBA5 with VACNEUT, both the β-hairpin and the α-helix were very stable, with an average C_α RMSD value of 1.1 and 0.7 Å, respectively. These values were 1.4 and 1.2 Å, respectively, in the EEF1 run. For αtα with VACNEUT, the average C_α RMSD was 1.2 and 1.3 Å for the first and second helix, respectively. Moreover, the percentage of π-helicity was smaller than 3% in the VACNEUT runs. Therefore, the regular elements of secondary structure are more stable with the VACNEUT approximation than the SASA model. The tertiary structure was not preserved in any simulation of BBA5 and αtα.

It is likely that the largest error in implicit solvent models originates from the treatment of the charged groups. VACNEUT, EEF1, and SASA have in common the neutralization of the ionic side chains, which is a rather crude approximation. In this respect, it is interesting to note that αtα and BBA5 have more ionizable side chains (15 and 7, respectively) than the GS and ^{D}PG peptide (2 and 4, respectively). Recent experimental data indicate that at pH 10.5 αtα, which has four lysine residues, is largely disordered even at a temperature of 5°C.[43] This shows that changes in the interactions between ionizable groups can have significant effects on stability. Finally, BBA5 was not stable in a 300 K MD simulation with explicit water and the AMBER force field with an 8-Å cutoff, whereas it was stable with a 10-Å cutoff.[44]

CONCLUDING DISCUSSION

An efficient, analytical implicit solvation model for MD simulations of peptides and small proteins was evaluated. It is based on a SASA term that complements the CHARMM polar hydrogen force field. A linear distance-dependent dielectric function and neutralized ionic side chains are used to approximate the electrostatic screening effects. The model contains only two atomic solvation parameters that were calibrated by performing 1-ns MD simulations at 300 K for six small proteins. Larger deviations from the native conformation were observed in 10-ns simulations, which shows that relatively long MD simulations are needed to validate an energy function. The typical problems which arise in *in vacuo* MD simulations, large deviations from the native conformation and an excessive number of intrasolute hydrogen bonds, are reduced when including the mean solvation term. In previous studies, the usefulness of the SASA model for simulating the reversible folding of structured peptides was shown for two three-stranded antiparallel β-sheets,[34,35] five alanine-based α-helices,[42,45] and a β-hairpin of 12 residues.[45] The simulations successfully reproduced the native conformation, irrespective of the starting structure. In a more recent study, the relative importance of amino acid sequence and native state topology in the unfolding process of two SH3 domains and two circular permutants has been investigated.[46] The statistically significant number of folding/unfolding events sampled by MD with the SASA model has allowed to study the main folding pathways and energetics. In agreement with experimental data,[38,47] one interesting insight from the simulation results is that at least for small β-sheet proteins the folding mechanism seems to be primarily defined by the native state topology, whereas specific interactions determine the statistically predominant folding route.[35,46]

There are several approximations in both EEF1[26] and SASA whose treatment of the electrostatic screening effects due to the solvent is based on EEF1. In both EEF1 and SASA the assumption is made that the main effects of

an aqueous solvent on the solute are additive and most of the direct solvation free energy is due to the first solvation shell. Another important approximation common to both models is that the effective dielectric constant between partial charges does not depend on the environment. This is reasonable only for peptides and very small proteins where most of the charges are on the surface and experience a similar environment. Further, to keep the SASA model as simple as possible, the atomic solvation parameters do not have a temperature dependence. Nevertheless, MD simulations with the SASA model were used to demonstrate the non-Arrhenius temperature dependence of the folding rate of structured peptides[45] in agreement with experimental data.[48,49]

It is likely that a more precise treatment of the dielectric screening effects is needed to improve the SASA model. The present and previously published results indicate that despite its drawbacks, such as the instability of large proteins, the SASA model is useful for studying thermodynamics and kinetics of folding of structured peptides. Its application to larger proteins could be limited by the lack of an environment-dependent dielectric function and the rather crude treatment of ionic side chains.

ACKNOWLEDGMENTS

The authors thank Urs Haberthür for subsequently implementing the SASA module in CHARMM (version 28), and Prof. T. Lazaridis for providing us the partial charges of the neutralized ionic side chains.

REFERENCES

1. Roux B, Simonson T. Implicit solvent models. Biophys Chem 1999;78:1–20.
2. Cramer CJ, Trulhar DG. Implicit solvation models: equilibria, structure, spectra, and dynamics. Chem Rev 1999;99:2161–2200.
3. Warwicker J, Watson HC. Calculation of the electric potential in the active site cleft due to α-helix dipoles. J Mol Biol 1982;157:671–679.
4. Gilson MK, Rashin A, Fine R, Honig BH. On the calculation of electrostatic interactions in proteins. J Mol Biol 1985;184:503–516.
5. Davis ME, Madura JD, Luty BA, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. Comp Phys Comm 1991;62:187–197.
6. Zauhar RJ, Morgan RS. Computing the electric-potential of biomolecules—application of a new method of molecular-surface triangulation. J Comput Chem 1990;11:603–622.
7. Davis ME, McCammon JA. Calculating electrostatic forces from grid-calculated potentials. J Comput Chem 1990;11:401–409.
8. Wan SZ, Wang CX, Xiang ZX, Shi YY. Stochastic dynamics simulation of alanine dipeptide: including solvation interaction determined by boundary element method. J Comput Chem 1997; 18:1440–1449.
9. Smart JL, Marrone TJ, McCammon JA. Conformational sampling with poisson-boltzmann forces and a stochastic dynamics/monte carlo method: application to alanine dipeptide. J Comput Chem 1997;18:1750–1759.
10. Im W, Beglov D, Roux B. Continuum solvation model: computation of electrostatic forces from numerical solutions to the poisson-boltzmann equation. Comput Phys Commun 1998;111:59–75.
11. David L, Luo R, Gilson MK. Comparison of generalized Born and Poisson models: energetics and dynamics of HIV Protease. J Comput Chem 2000;21:295–309.
12. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
13. Scarsi M, Apostolakis J, Caflisch A. Continuum electrostatic energies of macromolecules in aqueous solutions. J Phys Chem A 1997;101:8098–8106.
14. Ghosh A, Rapp CS, Friesner RA. Generalized Born model based on a surface integral formulation. J Phys Chem B 1998;102:10983–10990.
15. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
16. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. J Phys Chem A 1997;101:3005–3014.
17. Dominy BN, Brooks CL III. Development of a generalized born model parametrization for proteins and nucleic acids. J Phys Chem B 1999;103:3765–3773.
18. Schaefer M, Bartels C, Karplus M. Solution conformation and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. J Mol Biol 1998;284:835–848.
19. Bursulaya BD, Brooks CL III. Folding free energy surface of a three-stranded β-sheet protein. J Am Chem Soc 1999;121:9947–9951.
20. Tsui V, Case DA. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. J Am Chem Soc 2000;122: 2489–2498.
21. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
22. Stouten PFW, Froemmel C, Nakamura H, Sander C. An effective solvation term based on atomic occupancies for use in protein simulations. Mol Simul 1993;10:97–120.
23. Ooi T, Oobatake M, Némethy M, Scheraga HA. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc Natl Acad Sci USA 1987;84:3086–3090.
24. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Prot Sci 1992;1:227–235.
25. Fraternali F, van Gunsteren WF. An efficient mean solvation force model for use molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
26. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Prot Struct Funct Genet 1999;35:133–152.
27. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
28. Hasel W, Hendrickson TF, Still WC. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Comput Methodol 1988;1:103–116.
29. Mehler EL, Eichele G. Electrostatic effects in water-accessible regions of proteins. Biochemistry 1984;23:3887–3891.
30. Ramstein J, Lavery R. Energetic coupling between DNA bending and base pair opening. Proc Natl Acad Sci USA 1988;85:7231–7235.
31. Mehler EL. Comparison of dielectric response modles for simulating electrostatic effects in proteins. Prot Eng 1990;3:415–417.
32. Bartels C, Stote R, Karplus M. Characterization of flexible molecules in solution: the RGDW peptide. J Mol Biol 1998;284:1641–1660.
33. 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.
34. Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel β-sheet peptide. Proc Natl Acad Sci USA 2000;97: 10780–10785.
35. Ferrara P, Caflisch A. Native topology or specific interactions: what is more important for peptide folding? J Mol Biol 2001;306: 837–850.
36. De Alba E, Santoro J, Rico M, Jiménez MA. De novo design of a monomeric three-stranded antiparallel β-sheet. Prot Sci 1999;8: 854–865.
37. Schenck HL, Gellman SH. Use of a designed triple-stranded antiparallel β-sheet to probe β-sheet cooperativity in aqueous solution. J Am Chem Soc 1998;120:4869–4870.
38. Griffiths-Jones SR, Searle MS. Structure, folding, and energetics of cooperative interactions between the β-strands of a de novo

designed three-stranded antiparallel β-sheet peptide. J Am Chem Soc 2000;122:8350–8356.
39. Dobson CM, Šali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed 1998;37:869–893.
40. Struthers MD, Ottesen JJ, Imperiali B. Design and nmr analyses of compact, independently folded bba motifs. Folding Design 1998;3:95–103.
41. Fezoui Y, Connolly PJ, Osterhout JJ. Solution structure of αtα, a helical hairpin peptide of de novo design. Prot Sci 1997;6:1869–1877.
42. Hiltpold A, Ferrara P, Gsponer J, Caflisch A. Free energy surface of the helical peptide $Y(MEARA)_6$. J Phys Chem B 2000;104:10080–10086.
43. Fezoui Y, Hartley DM, Walsh DM, Selkoe DJ, Osterhout JJ, Teplow DB. A de novo designed helix-turn-helix peptide forms nontoxic amyloid fibrils. Nat Struct Biol 2000;7:1095–1099.
44. Wang L, Duan Y, Shortle R, Imperiali B, Kollman PA. Study of the stability and unfolding mechanism of BBA1 by molecular dynamics simulations at different temperatures. Prot Sci 1999;8:1292–1304.
45. Ferrara P, Apostolakis J, Caflisch A. Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. J Phys Chem B 2000;104:5000–5010.
46. Gsponer J, Caflisch A. Role of native topology investigated by multiple unfolding simulations of four SH3 domains. J Mol Biol 2001;309:285–298.
47. McCallister EL, Alm E, Baker D. Critical role of β-hairpin formation in protein G folding. Nat Struct Biol 2000;7:669–673.
48. Muñoz V, Thompson PA, Hofrichter J, Eaton WA. Folding dynamics and mechanism of β-hairpin formation. Nature 1997;390:196–199.
49. Lednev IK, Karnoup AS, Sparrow MC, Asher SA. α-Helix peptide folding and unfolding activation barriers: a nanosecond UV resonance raman study. J Am Chem Soc 1999;121:8074–8086.
50. McPhalen CA, James MNG. Crystal and molecular-structure of the serine proteinase-inhibitor CI-2 from barley-seeds. Biochemistry 1987;26:261–269.