PROTEINS: Structure, Function, and Genetics 25:89–103 (1996)

On the Sensitivity of MD Trajectories to Changes in Water-Protein Interaction Parameters: The Potato Carboxypeptidase Inhibitor in Water as a Test Case for the GROMOS Force Field

Xavier Daura,[1] Baldomero Oliva,[1] Enrique Querol,[1] Francesc X. Avilés,[1] and Orlando Tapia[2]
[1]Institut de Biologia Fonamental and Departament de Bioquímica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; [2]Fysikalisk Kemiska Institutionen, Uppsala Universitet, S-751 21 Uppsala, Sweden

ABSTRACT **A critical evaluation is presented of the sensitivity of the results of molecular dynamics simulations of proteins to changes in the parameters describing water–protein and protein–protein van der Waals interactions in the GROMOS force field. The origin of the van der Waals and electrostatic parameters of the GROMOS standard force field is reviewed, and possible weaknesses are discussed. Four alternate sets of van der Waals parameters for the oxygen types of the GROMOS force field that have been suggested by different authors are then tested against the original force field. Six 500 ps molecular dynamics simulations of the potato carboxypeptidase inhibitor (PCI) in solution using the different parameter sets are analyzed and the results compared with the available X-ray and NMR data. It is shown that the behavior of the molecular system is very sensitive to changes in the van der Waals parameters of the oxygens, especially when affecting the interactions between water and aliphatic or aromatic groups. It is also shown that correction of just the repulsive van der Waals parameter of the water oxygen for its interactions with nonpolar groups is sufficient to correct the main deficiency of the original GROMOS parameter set. Nevertheless, the present study suggests that further refinement of the current parameters is still needed for a proper representation of nonbonded interactions.** © 1996 Wiley-Liss, Inc.

Key words: computer simulation, force-field analysis, molecular dynamics of proteins, water solvation properties, van der Waals interactions

INTRODUCTION

Molecular dynamics (MD) simulations are becoming an increasingly important tool in the study of protein dynamics and structure/function relationships at an atomic level.[1,2] The approach is particularly useful in situations where direct experimental structural or dynamic information is difficult to obtain, for example, in the study of mechanisms of protein-ligand interaction[3,4] or in the study of protein folding/unfolding processes.[5–7] The results obtained from the simulations when addressing these challenging questions is, however, critically dependent on the properties of the underlying force field. One key aspect is the representation of the nonbonded interactions between the protein atoms and the surrounding water molecules. This is because protein stability depends intimately on the balance between protein–protein intramolecular interactions and protein–water intermolecular interactions.

Recently, different authors, based on different observations, have reported the failure of the GROMOS standard force field to properly model water–protein interactions. Åqvist et al.[8] noted that the GROMOS standard van der Waals parameters for O_w (water oxygen)–CH_n (united carbon) interactions yielded an incorrect value of free energy of hydration for a model of methane. These authors had earlier suggested a revision of the van der Waals parameters for charged carboxylate groups to better reproduce experimental free energies of hydration for these species.[9] van Buuren et al.[10] showed that the standard GROMOS force field does not yield correct decane/water interface behavior: in MD simulations of surfactant/oil/water systems, the solubility of decane in water appeared to be too high. These authors also tested different interaction models on the stability of a small α-helical peptide in water.[11] When using the standard parameters of GROMOS, they observed this helix starting to unfold after about 100ps. Mark et al.[12] found, from free energy calculations, the transformation of p-chlorophenol to p-methylphenol to be exothermic in water, against experimental and common-sense predictions. Smith et al.[13] also detected a loss of close packing, particularly of hydrophobic residues, in

Received June 9, 1995; revision accepted October 31, 1995.
Address reprint requests to F.X. Avilés, Institut de Biologia Fonamental, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain.

simulations of lysozyme in water, and reported the penetration of water into the core of the protein when using the standard GROMOS parameters.

Corrections to the water–protein interactions have been proposed by these three authors. In all cases, changes have been made to the short-range repulsion parameters for pair interactions involving oxygens, leading, primarily, to a higher repulsion between aliphatic or aromatic groups and surrounding water molecules. The revised parameter sets that have been proposed are not, however, the same, and contain internal inconsistencies. We therefore present here a critical evaluation of the sensitivity of the average structure of a protein and of the water solvation properties in molecular dynamics simulations to changes in the parameters describing water–protein and protein–protein van der Waals interactions. The origin of the van der Waals and electrostatic parameters of the GROMOS standard force field is first reviewed, and possible weaknesses are discussed. Four new sets of van der Waals parameters for the oxygen types of the force field, suggested by the above-mentioned authors as corrections to the original set, are then presented and discussed. To test these alternative parameter sets against the original force field, six molecular dynamics simulations, 500 ps-long each, have been performed for the potato carboxypeptidase inhibitor (PCI) in solution. The relative reliability of the different parameter sets has then been assessed by comparison of the results from the simulations to the available X-ray and NMR data.

Standard Force Field

In GROMOS, the nonbonded pair interaction potential consists of a 6-12 Lennard-Jones term plus a Coulombic term, and has the form:[14]

$$V(i,j) = -\frac{C6(i,j)}{r^6(i,j)} + \frac{C12(i,j)}{r^{12}(i,j)} + \frac{q(i)q(j)}{4\pi\varepsilon_0\varepsilon_r r(i,j)}, \quad (1)$$

where r(i,j) is the distance between the centres of atoms i and j, C6(i,j) and C12(i,j) are the van der Waals parameters for the long-range dispersion energy and the short-range repulsion energy, respectively, q(i) and q(j) are the atomic charges, $\varepsilon_r(=1)$ is the relative permittivity of the medium and ε_0 is the permittivity of vacuum.

The van der Waals parameters are defined in terms of the self-interaction between pairs of identical atoms, that is, C6(i,i) and C12(i,i). The parameters for crossed interactions (between different atom types) are generated by taking the geometric mean of the self-interaction values, such that $C6(i,j) = C6^{1/2}(i,i).C6^{1/2}(j,j)$ and $C12(i,j) = C12^{1/2}(i,i).C12^{1/2}(j,j)$. The application of such a combination rule is, however, not necessarily appropriate in all cases. It has recently been shown that, though the geometric mean yields a good estimate for C6(i,j), this combination rule significantly overestimates the well depth ε of the potential and underestimates the equilibrium distance σ for interactions between dissimilar atom types.[15] The combination rule $C12(i,j) = C12^{1/2}(i,i).C12^{1/2}(j,j)$ has also been observed to fail when a single set of $C12^{1/2}(i,i)$ parameters is used to model interactions of a different nature.[16] In the GROMOS force field, more than one $C12^{1/2}(i,i)$ value can be defined for each atom, the value that is used in conjunction with the combination rule depending on the nature of the interaction. In principle, this means that the Lennard-Jones parameters for any pair of atoms can be separately defined. For example, the GROMOS force field contains no special term for hydrogen-bonding interactions. Instead, the interaction between hydrogen-bonding pairs is modeled using a combination of Lennard-Jones and Coulombic interactions. The partial charges and the $C12^{1/2}(i,i)$ repulsive term are adjusted accordingly to reproduce experimental hydrogen-bonding distances and energies.

TABLE I. van der Waals Parameters of the GROMOS 37C4 Force Field for Selected Atom Types*

Atom type	$C6^{1/2}(i,i)$	$C12^{1/2}(i,i)$	$C12^{1/2}(i,i)_{hb}$	$C12^{1/2}(i,i)_{ch}$
O_{cn}	23.25	421.0	550.0	
O_{cx}	23.25	421.0	900.0	1,500.0
O_h	23.25	421.0	600.0	
O_w	25.01	421.0	793.3	
N_p	24.13	636.0	950.0	
N_{ad}	24.13	636.0	1,100.0	
N_{h3}	24.13	636.0	1,500.0	
N_{ab}	24.13	636.0	900.0	
N_{ah}	24.13	636.0	900.0	
C_{cn}	23.65	898.0		
C_{h1}	54.65	4,141.0		
C_{h2}	46.63	2,906.0		
C_{h3}	46.06	2,500.0		
C_{ar}	36.30	1,901.0		
C_b	23.65	898.0		
$N_{R\eta}$	24.13	636.0	1,050.0	
$N_{R\varepsilon}$	24.13	636.0	970.0	

*C6(i,i) in kcal mol^{-1} Å^6. C12(i,i), $C12(i,i)_{hb}$, and $C12(i,i)_{ch}$ in kcal mol^{-1} Å^{12}. For atom type abbreviations, see the text.

A list of $C6^{1/2}(i,i)$ and $C12^{1/2}(i,i)$ parameters from the GROMOS 37C4 force field is given in Table I for selected atom types. Where two values are given, $C12^{1/2}(i,i)_{hb}$ is used for interactions between a hydrogen-bond donor and a hydrogen-bond acceptor, and $C12^{1/2}(i,i)$ is used otherwise. $C12^{1/2}(i,i)_{ch}$ is only used for interactions between a charged carboxyl group and a charged amino or guanidino group.

Before commenting on the proposed modifications to the GROMOS force field, it is instructive to briefly indicate the origins of the standard parameters. We stress, however, that we are not the original authors

of the force field and that the following discussion is indicative but does not pretend to be a definitive discussion of the force field. For this we refer the reader to van Gunsteren and Berendsen.[14]

The $C6(O_w,O_w)$ (water oxygen) parameter was taken from a study of Zeiss and Meath,[17] based on data from photoabsorption and high-energy inelastic electron scattering experiments. The $C12(O_w,O_w)_{hb}$ parameter was derived from MD simulations of liquid water,[18] by fitting a simple three-points charge model to different experimental data. In these calculations the $C6(O_w,O_w)$ was held fixed, and $C12(O_w,O_w)_{hb}$ and $q(O_w)$ were treated as adjustable parameters. No van der Waals parameters were included for the water hydrogen. The targets were the interaction potential energy and pressure of liquid water at 300 K. The resulting water model is named the SPC (simple point charge) water model, and has been shown to adequately reproduce the shape of the radial distribution function of O_w–O_w distances as observed by X-ray diffraction, as well as a variety of dynamical and thermodynamical properties of liquid water.

The C6(i,i) and C12(i,i) parameters for the O_{cn} (carbonyl oxygen), N_{ab} (aromatic bare nitrogen), C_{cn} (carbonyl carbon), and C_b (bare carbon) were taken from a study of Nelson and Hermans.[19,20] They were derived from fitting to experimental data, such as crystal intermolecular packing and sublimation energies of small molecules containing the above-mentioned atom types and not containing hydrogen bonds. The parameters for O_{cn} and N_{ab} were then extended to the rest of the oxygen and nitrogen types of the force field. The $C12^{1/2}(O_w,O_w)$ was also assumed to be equal to that of the carbonyl oxygen.

As stated previously, the GROMOS potential does not contain an explicit hydrogen-bonding term. Instead, the interaction between hydrogen-bonding pairs is modeled using a combination of Lennard-Jones and Coulombic terms. The $C12^{1/2}(i,i)_{hb}$ parameters for O_{cn}, O_h, N_p (peptide nitrogen), and N_{ad} (amide nitrogen) were based on those proposed by Lifson et al.[21,22] The parameters of Lifson were derived from fitting to experimental data such as crystal intermolecular packing, sublimation energies, and dipole moments of small hydrogen-bonded molecules. The charge distributions proposed by Lifson for the peptide group (C_{cn}, O_{cn}, N_p, H_n) and amide group (C_{cn}, O_{cn}, N_{ad}, H_n) were directly transferred to GROMOS. The partial charges on C_{cn} and H_n (hydrogen bonded to nitrogen) were determined by assuming electroneutrality of the chemical groups CO, NH, and NH_2. The $C12^{1/2}(i,i)_{hb}$ parameters for the O_{cx} (ionized carboxyl oxygen), N_{h3} (ionized terminal amino), N_{ab}, N_{ah} (aromatic nitrogen), $N_{R\eta}$ (arginine nitrogen η), and $N_{R\varepsilon}$ (arginine nitrogen ε) were adjusted with respect to the rest of the force field to obtain appropriate hydrogen-bond equilibrium distances and energies.

Charge distributions for groups bearing a net charge, and for indole (tryptophan) and imidazol (histidine) rings were based on a study of Poland and Scheraga.[23] These authors calculated the charge distributions in polyaminoacids, obtaining the σ charges by molecular orbital methods, and the π charges by fitting to empirical dipole moment data of small molecules (in the case of the peptide and carboxyl groups) or from the literature (aromatic side chains). The charge distribution of the hydroxyl group was based on that of the SPC water.

C_{h1} (aliphatic CH), C_{h2} (aliphatic CH_2), C_{h3} (aliphatic CH_3), and C_{ar} (aromatic CH) were taken from the "united atom" model of Dunfield et al.[24] In this model, the nonpolar hydrogens are not explicitly included in the calculations. Rather, the hydrogens are included implicitly by representing the aliphatic or aromatic carbons and their attached hydrogens as a single mass centred in the carbon atom. The C6(i,i) and C12(i,i) parameters for these groups were derived from fitting to experimental data on crystal structures of hydrocarbons, and were tested, in part, in simulations of amino acid crystal structures. The charge on the united atoms was assumed to be zero.

Clearly, not all of the parameters are known with the same precision. The parameters have been obtained by separate fitting to different experimental and theoretical calculations and then combined with the aim of obtaining a self-consistent force field. In some cases, however, the self-consistency of the original parameters can be questioned. Probably, the most important example is the transfer of the $C6^{1/2}(i,i)$ and $C12^{1/2}(i,i)$ parameters of O_{cn} and N_{ab}, taken from Nelson and Hermans,[19] to the rest of the protein oxygen and nitrogen types. Logically, one would expect the polarizability of the different atom types to roughly follow the following order $O_{cn} < O_h,O_w < O_{cx}$, in the case of oxygen atom types, and $N_{ab},N_{ah},N_p,N_{R\varepsilon} < N_{ad},N_{R\eta} < N_{h3}$, in the case of nitrogen atom types. In accordance with the Slater-Kirkwood formula, this order would also be apparent in their $C6^{1/2}(i,i)$ coefficients (see for example Gelin and Karplus[25]). In addition, because in the GROMOS force field polar hydrogens have no van der Waals interactions, oxygens and nitrogens bearing polar hydrogens would be expected to have an increased van der Waals radius to prevent neighboring atoms from sampling the position of the polar hydrogen. This fact has implicitly been taken into account when adjusting the $C12^{1/2}(i,i)_{hb}$ parameters, but it has been clearly omitted when deriving the $C12^{1/2}(i,i)$ parameters. To be consistent with the model, one would expect the $C12^{1/2}(i,i)$ of the O_h to be larger than that of the O_{cn}, while the ratio between the $C12^{1/2}(i,i)$ of the O_h and that of the O_{cx} would be dependent on the weight given to their polarizabilities. The same reasoning can be applied to the different nitrogen types. The consequences of this underestimation of the repulsion energy of in-

TABLE II. Repulsion van der Waals Parameters for the Four Oxygen Types of the Force Field in the Tested Parameter Sets*

	psG		psA		psW		psD		psB	
Atom type	$C12^{1/2}$ (i,i)	$C12^{1/2}$ $(i,i)_{hb}$	$C12^{1/2}$ (i,i)	$C12^{1/2}$ $(i,i)_{hb}$	$C12^{1/2}$ (i,i)	$C12^{1/2}$ $(i,i)_{hb}$	$C12^{1/2}$ (i,i)	$C12^{1/2}$ $(i,i)_{hb}$	$C12^{1/2}$ (i,i)	$C12^{1/2}$ $(i,i)_{hb}$
O_{cn}	421.0	550.0	793.3	550.0	421.0	550.0	793.3	550.0	550.0	550.0
O_{cx}	421.0	900.0	793.3	956.0	421.0	900.0	793.3	900.0	900.0	900.0
O_h	421.0	600.0	793.3	600.0	421.0	600.0	793.3	956.0	600.0	600.0
O_w	421.0	793.3	793.3	793.3	793.3	793.3	793.3	793.3	793.3	793.3

*C12(i,i) and $C12(i,i)_{hb}$ in kcal mol^{-1} $Å^{12}$. psG corresponds to the standard force field,[14] psA has been proposed by Åqvist et al.,[8,9] psW has been proposed by Mark et al.,[12,13] psD is defined in this work, and psB has been proposed by van Buuren et al.[10,11] For atom type abbreviations, see the text.

teraction between H_2O, $-OH$, or $-NH_n$ groups and nonpolar atoms is expected to be particularly relevant in the case of interactions between aliphatic or atomatic groups and surrounding water molecules, since this could directly affect the stability of hydrophobic packing in the protein core. The effect on the intramolecular interactions within the protein are expected to be less important. First, the more abundant O_{cn} is the oxygen type for which Nelson and Hermans derived the parameters. In this case the self-interaction as well as interactions with nonpolar groups should be well represented. Second, hydrophobic interactions between CH_n groups, as well as hydrogen-bonding interactions (which use the adjusted $C12^{1/2}(i,i)_{hb}$) may also be assumed to be well parameterized. Finally, some of the affected atom types, such as O_h or N_{ad}, are found in a relatively small proportion in the protein.

Proposed Modifications

The observations made by Åqvist et al.,[8,9] van Buuren et al.,[10,11] and Mark et al.,[12,13] suggest the same deficiency discussed above: the repulsion energy of, at least, interactions between water molecules and CH_n groups is underestimated by the GROMOS 37C4 force field. These authors centred their attention in the $C12^{1/2}(O_w,O_w)$ parameter (and in the $C12^{1/2}(i,i)$ of the other oxygen types by extension) and suggested different corrections. The revised $C12^{1/2}(i,i)$ and $C12^{1/2}(i,i)_{hb}$ parameters for the 4 oxygen types are shown in Table II. The values corresponding to the standard force field (psG) are also included in Table II for comparison.

The set psA contains the corrections suggested by Åqvist and co-workers. Using the perturbation formula, they adjusted the $C12^{1/2}(O_w,O_w)$ parameter in order to obtain a value of free energy of hydration for a model of CH_4 that was in close agreement with the experimental one.[8] The CH_4 model was taken *equivalent to the* C_{h3} united atom of GROMOS. They found the value 793.3 (kcal mol^{-1} $Å^{12})^{1/2}$ to be more appropriate than the original 421.0 (kcal mol^{-1} $Å^{12})^{1/2}$ for the O_w-CH_4 interaction, and assigned 793.3 (kcal mol^{-1} $Å^{12})^{1/2}$ also to the $C12^{1/2}(i,i)$ of the other oxygen types. Comparing the parameters for the GROMOS C_{h3} to a well parameterized model of CH_4,[26] it is evident that although the effective diameters are similar, with $\sigma(CH_4)$ = 3.730 Å and $\sigma(C_{h3})$ = 3.786 Å, the self-interaction energies differ substantially, being $\varepsilon(CH_4)$ = 0.294 kcal mol^{-1} and $\varepsilon(C_{h3})$ = 0.180 kcal mol^{-1} and $V_{LJ} = 4\varepsilon\,[(\sigma^{12}/r^{12}) - (\sigma^6/r^6)]$, such that the GROMOS C_{h3} is less repulsive than the reference CH_4. Therefore, it is to be expected that the calculations of Åqvist and co-workers would overestimate the effective van der Waals radius of O_w for interactions with nonpolar carbon atoms. The assignment of the value 793.3 (kcal mol^{-1} $Å^{12})^{1/2}$ to the $C12^{1/2}(i,i)$ of all protein oxygen types is also questionable. For one, it assumes, as does the original GROMOS force field, that there is no difference in the effective van der Waals radii of different oxygen atom types. In the original force field the $C12^{1/2}(i,i)$ values were all assigned based on parameterization of the O_{cn}-O_{cn} interaction. By parameterizing all interactions based on the CH_4-O_w interaction, Åqvist and co-workers substitute the only other experimentally fitted value in the series, the $C12^{1/2}(O_{cn},O_{cn})$, illogically making it higher than the $C12^{1/2}(O_{cn},O_{cn})_{hb}$. This theoretical mismatching is also evident between the $C12^{1/2}(O_h,O_h)$ and the $C12^{1/2}(O_h,O_h)_{hb}$.

The set psB contains the corrections suggested by van Buuren and co-workers. They observed that the use of a single repulsion parameter for each oxygen type, that is, the original $C12^{1/2}(i,i)_{hb}$ parameters, gives an improved picture for decane/water interface behavior[10] and an increased stability for small α-helical peptides in water.[11] This parameterization contradicts the principle that the $C12^{1/2}(i,i)$ values should be smaller than the $C12^{1/2}(i,i)_{hb}$ values, since the latter were specifically increased to compensate for the attractive electrostatic interactions at short-range in hydrogen bonding, and thus are expected to overestimate the repulsion if used for interactions lacking this attractive term. In addition, the differences in the $C12^{1/2}(i,i)$ parameters of the four oxygen types do not reflect the expected differences in the effective van der Waals radii, i.e., oxygens to which

hydrogens are attached (O_w and O_h) should have increased van der Waals radii.

The set psW contains the correction suggested by van Gunsteren and co-workers, based on their own observations and on the work of Åqvist and that of van Buuren. These authors only changed the $C12^{1/2}(O_w,O_w)$ parameter, taking the value proposed by Åqvist. They tested this change by repeating the calculation of the free-energy change associated with the transformation of p-chlorophenol to p-methylphenol in water[12] and by running a new nanosecond simulation of lysozyme,[13] in both cases obtaining substantially better agreement with the available experimental data.

The set psD has been included to test the sensitivity of the trajectories to small changes in force-field parameters not involving the water-protein interaction. The parameters are derived from the set given by Åqvist et al., but with the corrections proposed for $C6^{1/2}(O_{cx},O_{cx})$ and $C12^{1/2}(O_{cx},O_{cx})_{hb}$ being applied to the O_h instead. This parameter change affects only a small number of interactions and so can be used to assess the reproducibility of the results.

Taking psG (the standard parameter set) as reference, each of the other four sets is expected to result in a lower energy of interaction between water molecules and nonpolar groups, thus stabilizing the hydrophobic packing inside the protein. Sets psA and psD are also expected to result in a weaker interaction between protein oxygens and nonpolar groups, as well as between pairs of carbonyl oxygens. This would decrease some of the intramolecular interactions and thus result in some degree of destabilization of the protein structure. A similar but less extensive effect would be expected in relation to set psB.

Test System

To investigate the effects of the five different parameter sets on the simulation of protein dynamics and stability, we have chosen the carboxypeptidase inhibitor from potato (PCI) as the test system. PCI is a protein particularly suitable for molecular dynamics studies because it is small (39 residues) and its structure is known both in aqueous solution[27] and in crystal complex with carboxypeptidase A (CPA).[28] PCI has previously been extensively studied in our laboratory using a variety of experimental techniques[29,30] to determine key elements of its structure, function, and folding pathway. The 27-residues globular core of PCI, organized in three loops, is stabilized by three disulphide bridges and lacks regular secondary structure, except for a short 5-residues 3_{10}-helix.[28] From this globular core protrudes a 7-residues N-terminal tail for which no functional role has been so far assigned, and a 5-residues C-terminal tail that docks into the active site of CPA leading to a stopper-like inhibition mechanism.[28] Although a large degree of disorder in the conformation of the C-terminal tail has been suggested by NMR measurements, it is believed that the inhibitor most of the time adopts a conformation similar to that found when forming complex with CPA. MD simulations have previously been used to map the conformational space accessible to the C-tail in vacuo, for the native PCI and for two mutants. The results have been correlated with inhibitory activities obtained experimentally.[31] The structural and dynamic properties of the inhibitor in water, as shown in a MD simulation, have also been studied.[32]

METHODS

A series of six molecular dynamics simulations was carried out for PCI with an explicit solvent representation. The GROMOS package[14] was used to perform all the calculations, the conformational energies being computed with its standard potential energy function. Five different sets of van der Waals parameters were used for the oxygen atom types, while all the other force-field parameters were taken from the standard 37C4 file. Two models of bulk water, the SPC[18] and the SPC/E,[33] were used. All simulations were run on a CRAY Y-MP.

Molecular Model

The crystallographic coordinates of PCI in complex with CPA, obtained at a 2.5 Å resolution,[28] were used as the initial molecular structure for the MD calculations, with removal of the N-terminal glutamic residue, undefined in the crystallographic structure and also lacking in some isoforms,[34] and the C-terminal glycine residue, not bound to the PCI in its complex with CPA and which was not found to be necessary for the inhibitory activity.[29,35] The coordinates of polar hydrogens were generated from standard geometries.

The X-ray structure of PCI was initially energy minimized; 900 steps of steepest descent were performed without explicit water molecules (37D4 parameter set[14]) in order to relax the strain. The potential energy decreased by 1,040 kcal/mol, and reached a negative value; the RMS deviation from the crystallographic structure was 0.36 Å for the backbone and 0.51 Å for all non-hydrogen atoms.

Two models for bulk water were considered. The already commented SPC model was used together with the parameter sets psG, psA, psW, psD, and psB, in MD simulations that we call, respectively, SimG, SimA, SimW, SimD, and SimB. The SPC/E model[33] was only used together with the parameter set psB, in a simulation that we call SimB/E. The extended simple point charge water model (SPC/E) is a variant of the SPC one, with the application of a polarization correction to its effective pair potential. This model has an enhanced dipole moment, and was reported to give better agreement with experimental values for the potential energy of liquid wa-

ter at 300 K and for its diffusion coefficient, being within the region of parameter space that produces well-defined second-neighbor peaks in the radial distribution function of O_w–O_w distances. We decided to use both water models when testing the set psB in order to see whether they produce distinguishable differences in the trajectory or not, and because of the observations of van Buuren and coworkers, who found that while the best equilibrium distributions at interfaces were obtained with the SPC water, the best dynamics in the aqueous phase were obtained with the SPC/E one.[10]

The energetically minimized PCI was placed in the center of a rectangular box. The dimensions were chosen such that the minimum distance of any protein atom to the wall was 5 Å. The solvent was inserted into the box by using a cubic box containing 216 equilibrated SPC or SPC/E (depending on the case) water molecules as a building block. All water molecules with the oxygen lying within 2.3 Å of a non-hydrogen protein atom were removed. The dimensions of the box were 30.4 Å × 32.2 Å × 43.6 Å for the six molecular models, and the number of water molecules in it was 1,259. In order to relax strong nonbonded interactions generated between the water molecules and the protein, 300 steps of steepest descent energy minimization were performed while keeping the protein atoms positionally restrained using a harmonic potential with a force constant of 21.5 kcal $Å^{-2}$ mol^{-1}.

In order to neutralize the charges of PCI, a total of $4Na^+$ and $4Cl^-$ were added to the system as follows: the electrostatic potential was calculated at all water-oxygen positions, and the water molecule at the lowest or highest potential was replaced with a Na^+ or Cl^- ion accordingly, unless this water oxygen lay within 3.5 Å of a non-hydrogen protein atom or another ion. This process was repeated until all ions were added. This was again followed by 300 steps of steepest descent energy minimization, keeping the protein atoms positionally restrained, in order to relax unfavorable short-range ion–water interactions. At this stage, the protein concentration was 0.038 M, while the salt concentration was 0.155 M.

MD Trajectories

Each of the six molecular models was then used to run a 500 ps simulation at constant temperature and pressure under periodic boundary conditions. Initial velocities for the atoms were taken in all simulations from the same Maxwellian distribution at 293 K. The protein and the solvent were, independently, strongly coupled to a temperature of 293 K with a relaxation time of 0.01 ps[36] for an initial period of 10ps, while a relaxation time of 0.05 ps was used to have the pressure at around 1 atm. After this time span, and throughout the rest of the simulation, the protein and the solvent were independently weakly coupled to the temperature of the bath with a relaxation time of 0.1 ps, while a relaxation time of 0.5 ps was chosen for the pressure. Bond lengths were constrained to equilibrium values using the SHAKE algorithm[37] allowing us to use an integration time step of 2 fs. The nonbonded interactions were evaluated using a twin-range method: the short-range van der Waals and electrostatic interactions were evaluated at each time step using an 8 Å cut-off radius; long-range electrostatic interactions were evaluated at each time step, within a cut-off radius of 13 Å, using a charge group pair list that was updated every 10 time steps. The cut-off radii were applied to the centers of geometry of charge groups and to the oxygen atom of the water molecule.

Analysis

1. Mean RMS deviation: For the calculation of the RMS deviation of two conformations, j and k, of PCI, the centroids of these conformations are superimposed and the rotation **R** that minimizes the RMS difference between the vectors $\mathbf{r}_{ji}$ and $\mathbf{r}_{ki}$, defined from the common centroid to each of the N non-hydrogen atoms, is found. Then:

$$\mathrm{RMSD}_{jk} = \left[\frac{1}{N}\min_{R}\sum_{i=1}^{N}\|\vec{r}_{ji} - \mathbf{R}\vec{r}_{ki}\|^2\right]^{1/2} \quad (2)$$

We have calculated two types of mean RMSD: (a) an average of the RMSD between every two conformations from a group of 36 conformations extracted from the trajectory (time interval 150–500 ps), and (b) an average of the RMSD between every one of these conformations and the X-ray structure or the NMR structure.

(a)<MD> vs. <MD>:

$$\text{mean RMSD} = \frac{2}{M(M-1)}\sum_{j=1}^{M}\sum_{k=j+1}^{M}\mathrm{RMSD}_{jk}, \quad (3)$$

(b) <MD> vs. X-ray and <MD>vs. NMR:

$$\text{mean RMSD} = \frac{1}{M}\sum_{k=1}^{M}\mathrm{RMSD}_{jk}, \quad (4)$$

where M is the number of conformations extracted from the MD simulation.

2. RMS structural displacement: 100 conformations extracted at intervals of 5 ps from the MD trajectory have been used for the representation of the RMS structural displacement from the crystallographic structure over simulation time. After a least square fitting between each of these conformations and the X-ray conformation, based on a set of 19 C_α atoms that showed a low B-factor in all simulations, the RMS difference between a conformation j from the trajectory and the X-ray conformation (k), N being the number of atoms, is calculated as:

$$\mathrm{RMSD}_{jk} = \left[\frac{1}{N}\sum_{i=1}^{N}\|\vec{r}_{ji} - \vec{r}_{ki}\|^2\right]^{1/2}. \quad (5)$$

3. Percentage of solvent-accessible surface area (SASA): the SASA of polar and nonpolar protein atoms has been calculated here by using the program GEPOL,[38] with a rolling sphere of 1.4 Å. It is plotted against simulation time as a percentage of the maximum SASA of the polar and of the nonpolar atoms, the maximum being the sum over the SASA of the respective isolated atoms.

4. Water density around polar and nonpolar protein atoms: the water density has here been calculated as the number of water atoms lying in a volume defined by a cut-off radius of 3.5 Å around the van der Waals volume of every protein atom, and averaging it over 175 conformations extracted from the time interval 150–500 ps. This cut-off radius has been chosen such that the differences in density in the simulations are amplified, that is, by using a shorter radius the densities are too low and because of that more homogeneous, and by using a longer radius the densities tend to be closer to that of bulk water and, because of that, again more homogeneous.

RESULTS AND DISCUSSION

A common measure of the quality of a given MD trajectory is the mean RMS positional deviation with respect to a reference conformation, for example, the starting structure or a structure determined from X-ray crystallography or NMR. In the present case, such a comparison is not straightforward. One would expect that, as the simulations were performed in solution, the structure of PCI would evolve from the initial X-ray structure to become closer to the NMR structure, as protein-protein interactions present in the complex with CPA and crystal packing forces are removed. The NMR structure is, however, an early-derived structure[27] and has to be used with caution. In particular, the N-(residues 1–7) and C-(residues 35–39) terminal tails are poorly defined by the experimental data. For this reason, although data relating to comparisons based on all atoms and for the entire molecule (residues 2–38) are presented and discussed for completeness, detailed analysis will concentrate on the backbone atoms of the core of the protein (residues 8–34). Mean RMS positional differences between the simulated and reference structures are given in Table III. The RMS positional difference between the NMR structure and the X-ray structure is 2.54 Å and 1.66 Å for all atoms and for backbone atoms, respectively (residues 2–38). Thus, although they have the same fold, there are significant conformational differences between the two experimental structures. These differences are particularly clear when comparisons are made based on subsets of residues. By comparing the RMS positional differences for the core residues, the helix (residues 14–21), and the regions before and after the helix, one can easily see that the most significant difference between the backbone of the core residues in the X-ray and NMR structures resides in the helical region. In fact, the helix is defined between residues 14–18 by X-ray and between residues 16–21 by NMR. For the core, the RMS positional difference is 1.16 Å, compared with 1.47 Å for the helical region (residues 14–21) and 0.70 Å for the rest of the core.

When comparing the MD ensembles to the crystallographic structure, the picture that is obtained depends on whether the protein is considered as a whole or in sections. Considering the backbone atoms of residues 2 to 38 it can be seen that <SimG>, <SimD>, <SimB/E>, and <SimB> all lie in a narrow range slightly below the 2.10 Å measured for the NMR ensemble. <SimA> and <SimW> lie approximately 0.5 Å above the <NMR>, but the values are probably highly influenced by the mobility of the tails, especially that of the N-terminal one. In this context, we are considering the values of the NMR ensemble as an indication of what might be a reasonable upper limit for the RMS. Only <SimA> shows a difference larger than <NMR> from the X-ray structure for the protein core. If a small RMS deviation from the X-ray structure is characteristic of a good simulation, then <SimW> and <SimD> are the ensembles that most closely reproduce the reference structure in the core region.

Comparison between the MD ensembles and the NMR structure shows that the mean RMS deviation vs. the NMR structure is larger for any simulation than for the NMR ensemble. <SimD> (1.1 Å for the core backbone) is the MD ensemble that best reproduces the NMR structure.

Comparing the mean RMS positional difference in the helix region (residues 14–21) between the simulations and both the X-ray and NMR structures, it can be seen that the only ensemble closer to the NMR structure than to the X-ray structure in this region is <SimD>. Furthermore, its mean RMS positional deviation vs. the NMR structure (0.72 Å) is the smallest in Table III. This might be taken as indicating an evolution from the X-ray to the NMR structure but should be interpreted with caution. The smallest mean RMS positional deviation vs. the X-ray structure for the helix section is that of <SimW> (0.91 Å). If we consider the rest of the core (residues 8–13 and 22–34), we find very similar values for <SimD> and <SimW> in both comparisons with the X-ray and NMR structures.

The mean RMS positional deviation within the simulation ensemble gives an indication of the structural stability in the dynamics. It is for all the simulations smaller than for <NMR>. The smallest values are in general for <SimW> and <SimD>, but this cannot be taken as a measure of quality.

The time series of the RMS structural displace-

TABLE III. Mean RMS Positional Deviations*

	Mean RMS deviation (Å)							
	Residues 2–38		Residues 8–34		Residues 14–21		Residues 8–13 and 22–34	
Comparison	Backbone	All atoms†	Backbone	All atoms†	Backbone	All atoms†	Backbone	All atoms†
NMR vs. X-ray	1.66	2.54	1.16	2.27	1.47	2.25	0.70	2.13
<NMR> vs. X-ray	2.10 ± 0.20	3.00 ± 0.30	1.50 ± 0.20	2.40 ± 0.30				
<SimG> vs. X-ray	1.81 ± 0.27	2.86 ± 0.30	1.58 ± 0.16	2.67 ± 0.21	1.29 ± 0.14	1.92 ± 0.15	1.12 ± 0.15	2.61 ± 0.29
<SimA> vs. X-ray	2.65 ± 0.32	3.18 ± 0.31	1.86 ± 0.32	2.54 ± 0.27	1.46 ± 0.29	2.16 ± 0.35	1.20 ± 0.16	2.22 ± 0.20
<SimW> vs. X-ray	2.58 ± 0.16	3.66 ± 0.18	0.91 ± 0.12	2.13 ± 0.29	0.91 ± 0.20	1.52 ± 0.20	0.73 ± 0.07	2.23 ± 0.39
<SimD> vs. X-ray	1.89 ± 0.23	2.90 ± 0.18	1.13 ± 0.10	2.13 ± 0.17	1.35 ± 0.16	1.85 ± 0.14	0.71 ± 0.09	2.06 ± 0.19
<SimB/E> vs. X-ray	1.62 ± 0.27	2.79 ± 0.26	1.22 ± 0.25	2.42 ± 0.28	1.00 ± 0.14	1.65 ± 0.22	0.82 ± 0.07	2.45 ± 0.22
<SimB> vs. X-ray	1.86 ± 0.26	2.67 ± 0.19	1.46 ± 0.14	2.48 ± 0.13	1.22 ± 0.17	1.84 ± 0.14	1.14 ± 0.11	2.47 ± 0.13
<NMR> vs. NMR	1.50 ± 0.30	2.10 ± 0.30	1.00 ± 0.20	1.50 ± 0.20				
<SimG> vs. NMR	2.10 ± 0.10	2.88 ± 0.13	1.66 ± 0.11	2.36 ± 0.14	1.77 ± 0.18	2.25 ± 0.17	1.23 ± 0.09	2.28 ± 0.21
<SimA> vs. NMR	2.63 ± 0.36	3.23 ± 0.32	1.78 ± 0.19	2.40 ± 0.13	1.84 ± 0.18	2.26 ± 0.19	1.37 ± 0.13	2.32 ± 0.15
<SimW> vs. NMR	2.47 ± 0.16	3.60 ± 0.15	1.22 ± 0.10	2.41 ± 0.16	1.20 ± 0.14	1.76 ± 0.17	0.86 ± 0.07	2.38 ± 0.12
<SimD> vs. NMR	1.96 ± 0.15	2.81 ± 0.15	1.08 ± 0.09	1.93 ± 0.09	0.72 ± 0.12	1.39 ± 0.12	0.95 ± 0.09	2.00 ± 0.09
<SimB/E> vs. NMR	2.09 ± 0.30	3.05 ± 0.27	1.49 ± 0.21	2.48 ± 0.12	1.36 ± 0.31	2.13 ± 0.21	1.04 ± 0.10	2.41 ± 0.11
<SimB> vs. NMR	2.11 ± 0.13	2.87 ± 0.13	1.62 ± 0.11	2.40 ± 0.10	1.75 ± 0.20	2.36 ± 0.16	1.26 ± 0.10	2.30 ± 0.11
<NMR> vs. <NMR>	2.10 ± 0.40	2.80 ± 0.40	1.30 ± 0.20	2.00 ± 0.30				
<SimG> vs. <SimG>	1.35 ± 0.41	2.06 ± 0.58	0.91 ± 0.22	1.64 ± 0.39	0.79 ± 0.24	1.23 ± 0.31	0.76 ± 0.18	1.61 ± 0.42
<SimA> vs. <SimA>	1.54 ± 0.52	2.11 ± 0.60	1.00 ± 0.33	1.51 ± 0.35	0.96 ± 0.49	1.41 ± 0.52	0.76 ± 0.19	1.40 ± 0.31
<SimW> vs. <SimW>	0.93 ± 0.21	1.69 ± 0.46	0.69 ± 0.18	1.56 ± 0.50	0.66 ± 0.23	0.99 ± 0.28	0.56 ± 0.14	1.62 ± 0.55
<SimD> vs. <SimD>	1.11 ± 0.26	1.68 ± 0.39	0.71±0.14	1.33 ± 0.33	0.53 ± 0.16	0.87 ± 0.19	0.66 ± 0.14	1.38 ± 0.36
<SimB/E> vs. <SimB/E>	1.36 ± 0.50	1.91±0.64	0.97 ± 0.33	1.49 ± 0.40	0.93±0.41	1.23 ± 0.48	0.71 ± 0.18	1.44 ± 0.36
<SimB> vs. <SimB>	1.22 ± 0.29	1.75 ± 0.33	0.85 ± 0.17	1.47 ± 0.27	0.74 ± 0.24	1.04 ± 0.25	0.74 ± 0.16	1.50 ± 0.30

*X-ray is the structure of the PCI as extracted from its crystal complex with the CPA.[28] <NMR> comprises 11 converged distance-geometry structures subjected to restrained energy minimization and restrained molecular dynamics.[27] NMR has been obtained by fitting and averaging the 11 structures in <NMR>, with a further restrained energy minimization. The simulation time interval 150–500 ps has been used for the statistics. The selection of one conformation every 10 ps gives six sets of 36 structures, noted as <SimG>, <SimA>, <SimW>, <SimD>, <SimB/E>, and <SimB>.
†Non-hydrogen atoms.

ment from the starting X-ray structure for the backbone atoms of the whole protein (residues 2–38) and of the core (residues 8–34) is displayed for each of the simulations in Figure 1. In terms of the RMS structural displacement from the starting structure, the simulations SimG, SimW, and SimD have apparently reached equilibrium within about 150 ps. In SimA and SimB large fluctuations in the RMS are observed. In SimB/E, the RMS for both the whole structure and the core is stable during the first 300

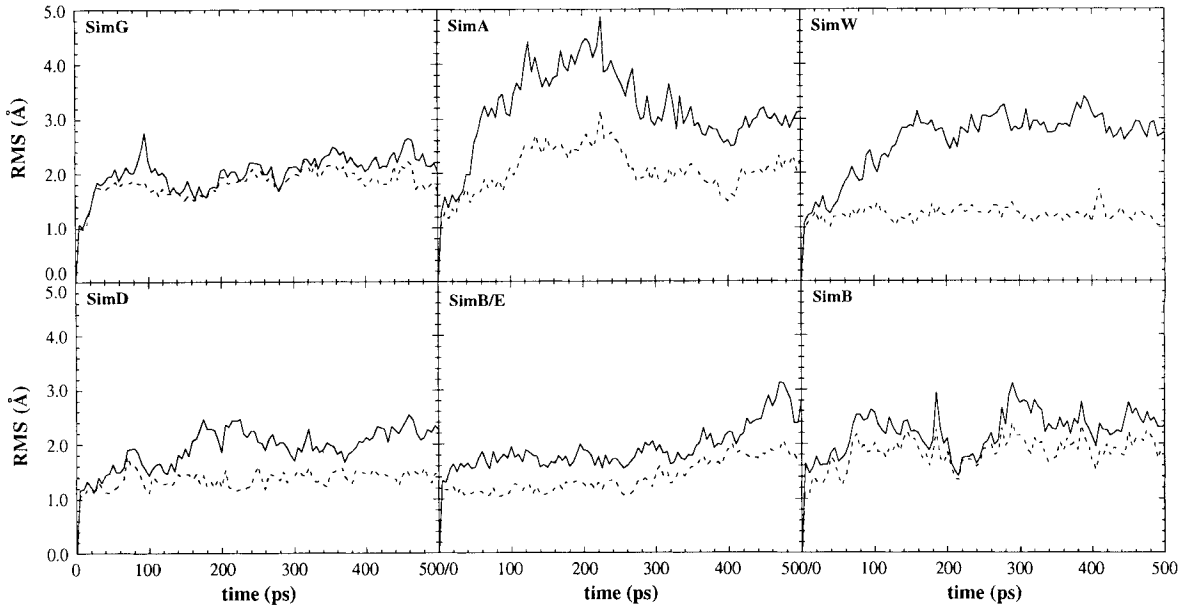


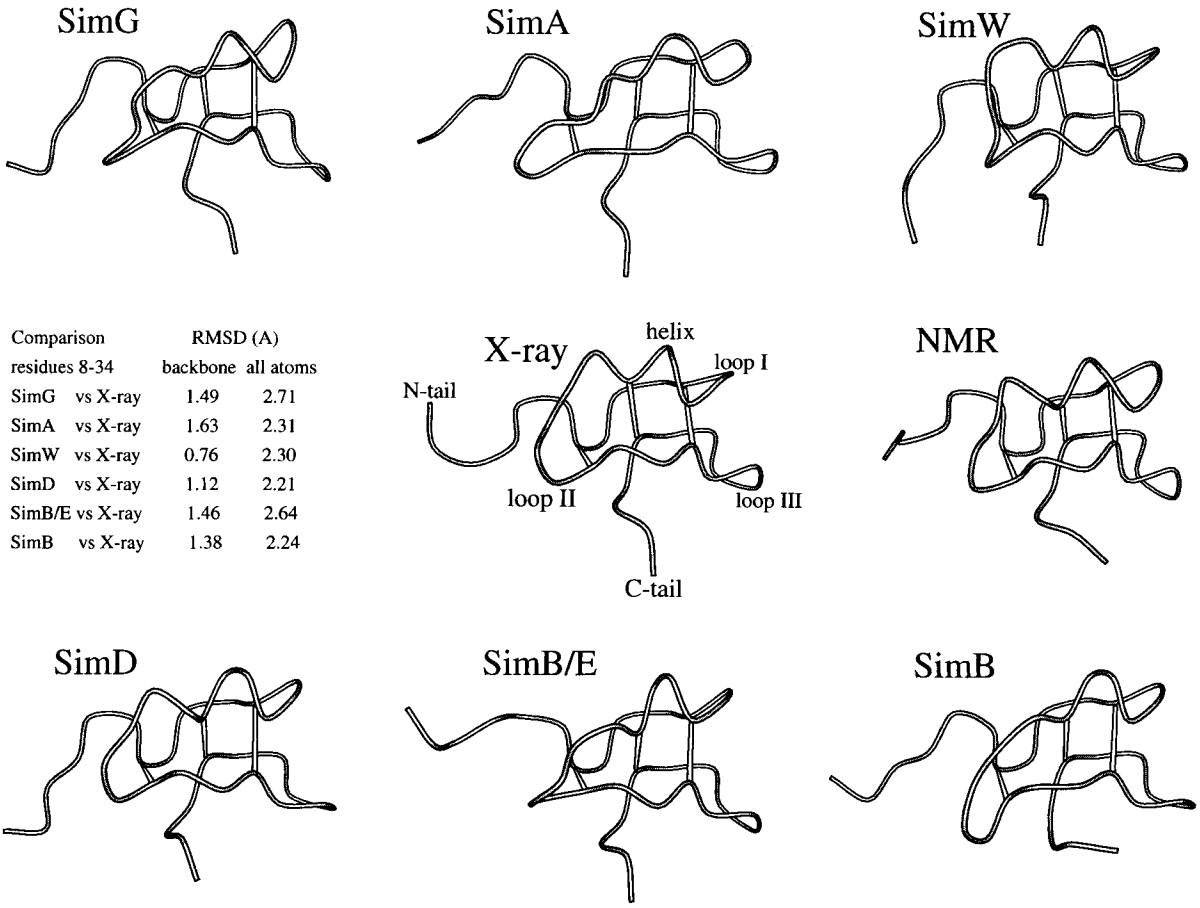
Fig. 1. RMS structural displacement from the X-ray structure along simulation time. The solid lines show the time evolution of the RMS calculated for the backbone atoms of the whole molecule (residues 2–38), while the broken lines show the time evolution of the RMS calculated for the backbone atoms of the core (residues 8–34).

ps. After this point, the RMS shows a marked increase indicating the system is not yet in equilibrium. Structurally, SimA is clearly the most unstable, having a large RMS value with large fluctuations. Figure 2 shows the average structure from the last 100 ps of each simulation, and their RMS positional deviations vs. the X-ray structure. It can be seen that loop II, after the short helix, is the region in the core of the inhibitor which appears, in general, more deformed after 400 ps of simulation. The average structures from SimD and SimW are the ones that remain closest to the crystallographic conformation in the region of this loop. The figure also shows that in SimW the protein tends to maximize its packing, with a compact core and the N-tail partially folded against it.

Table IV shows the average number of hydrogen bonds per atom type and conformation, in terms of hydrogen bonds with the solvent and intramolecular hydrogen bonds. The ratio of the total number of protein–water hydrogen bonds to the total number of protein–protein hydrogen bonds per conformation is smaller in SimW, SimD, and SimB (3.6, 4.2, and 4.8, respectively) than in SimB/E, SimA, and SimG (5.8, 5.9, and 6.0, respectively). The trend in the first group of simulations to favor the intramolecular hydrogen bonds may have a direct influence in the stability of the inhibitor, which in fact is higher in these three simulations. The effect is specially important when involving the atom types O_{cn} and N_p, which in turn are the most important acceptor and donor of intramolecular hydrogen bonds.

Table V lists the NH···OC backbone hydrogen bonds as found in the X-ray and NMR structures, as well as the frequency of their occurrence in the simulations (>50%). For those hydrogen bonds that appear in at least 50% of the conformations extracted from the trajectory, the patterns evident in the different simulations are quite similar. Those hydrogen bonds that are present in both the X-ray and NMR structures, 12CysN–32ArgO, 34CysN–10LysO, 35GlyN–26AlaO, and 26AlaN–35GlyO, are also observed with high frequency in all simulations except SimD and SimW in which 26AlaN–35GlyO occurs in close to, or less than, 50% of the configurations. Table V also shows that the two hydrogen bonds which define the 3_{10}-helix in the crystallographic structure, 17AspN–14ThrO and 18CysN–15HisO, occur more frequently in SimW, SimG, and SimD than in the other three simulations.

The percentage of solvent-accessible surface area of polar and nonpolar atoms as a function of simulation time is shown in Figure 3. The most striking results occur in SimG using the unmodified GROMOS force field. During the first 10 ps of this simulation, the percentage of SASA increases rapidly for the nonpolar atoms while decreasing for the polar ones, the difference being maintained along the rest of the trajectory. The fact that psG favors the



Comparison residues 8-34		RMSD (A) backbone	all atoms
SimG	vs X-ray	1.49	2.71
SimA	vs X-ray	1.63	2.31
SimW	vs X-ray	0.76	2.30
SimD	vs X-ray	1.12	2.21
SimB/E	vs X-ray	1.46	2.64
SimB	vs X-ray	1.38	2.24

Fig. 2. Average structures from the time interval 400–500 ps and their RMS positional deviation from the X-ray structure. "X-ray" is the structure of the PCI as extracted from its crystal complex with the CPA.[28] "NMR" is the structure of the PCI as determined in solution. For its derivation, 11 converged distance-geometry structures were subjected to restrained energy minimization and restrained molecular dynamics, and were afterwards fitted and averaged; the generated structure was then subjected to a further restrained energy minimization.[27]

exposure of hydrophobic atoms to the solvent water, relative to the exposure found in the starting X-ray crystallographic structure, is in agreement with the conclusions of previous studies discussed in the introduction.[8–13] The percentage of SASA cannot be used, however, to establish the comparative reliability of the other simulations. In principle, all of them, with the possible exception of SimB, are at least not in contradiction with the X-ray SASA values, which are 7.17% for the polar atoms and 7.05% for the nonpolar atoms. The NMR values are 5.45 and 7.27%, respectively, for the polar and nonpolar atoms. However, as the conformation of residues in the tails and the positions of most of the side chains, which are the biggest contributors to the SASA, are not accurately known, these values are not considered reliable.

The average water density around the polar and nonpolar protein atoms is shown in Table VI. In SimG, the measured water densities are considerably higher than in any other simulation (by a factor of two around the nonpolar atoms). It should be noted that the water density has been determined by counting water molecules which fall within a cut-off radius of 0.35 nm from a given protein atom. The exact value of the density is very sensitive to the cut-off. Choosing 0.35 nm highlights the fact that in SimG the water molecules can approach closer to, and interact more strongly with, nonpolar atoms than using any of the modified parameter sets. Interestingly, the water density around the polar atoms is also significantly higher in SimG compared to the other simulations although the water–polar atoms interaction is unchanged. This illustrates the extent to which the alteration of one parameter can result in substantial indirect effects on the interactions in other parts of the system. The other five simulations all have similar values for the water density around polar and nonpolar protein atoms. Thus, although the water density is sensitive to the

TABLE IV. Average Number of Hydrogen Bonds per Atom Type and Conformation*

	SimG		SimA		SimW		SimD		SimB/E		SimB	
Atom type	HBS†	HBP‡	HBS	HBP	HBS	HBP	HBS	HBP	HBS	HBP	HBS	HBP
O_{cn}	19.6	7.3	25.3	9.0	22.0	14.5	25.1	12.0	25.3	10.2	24.0	11.0
O_{cx}	15.2	3.0	20.5	4.7	18.5	4.8	17.8	4.9	15.6	2.7	15.0	3.7
O_h	5.6	0.8	8.2	0.3	7.8	0.8	5.9	0.9	8.1	0.4	8.2	0.4
N_p	9.2	9.0	9.9	13.4	6.0	18.0	7.7	15.2	10.7	12.3	9.5	13.5
N_{ad}	3.9	0.4	6.0	0.2	5.2	0.7	5.2	0.7	5.8	0.3	5.3	0.6
N_{h3}	5.1	0.0	7.0	0.2	7.1	0.3	7.3	0.1	7.0	0.5	7.3	0.2
N_{ab}	2.1	0.2	2.6	0.4	2.2	1.0	2.1	0.9	2.5	0.6	2.3	0.8
N_{ah}	1.4	0.3	2.3	0.2	2.1	0.4	1.7	0.5	2.6	0.1	2.1	0.6
$N_{R\eta}$	1.6	0.3	1.9	0.2	2.4	0.2	2.1	0.4	2.4	0.2	2.1	0.5
$N_{R\varepsilon}$	0.6	0.0	0.8	0.0	0.7	0.1	0.5	0.2	0.5	0.3	0.6	0.2
Total	64.3	10.7	84.7	14.3	74.0	20.4	75.5	17.9	80.6	13.8	76.3	15.8

*The simulation time interval 150–500 ps has been used for the statistics.
†Average number of hydrogen bonds with solvent atoms.
‡Average number of hydrogen bonds with other protein atom types. The quotient HBS/HBP is 6.0 in SimG, 5.9 in SimA, 3.6 in SimW, 4.2 in SimD, 5.8 in SimB/E, and 4.8 in SimB. For atom type abbreviations see the text.

TABLE V. Intramolecular Hydrogens Bonds in the X-Ray and NMR Structures and Their Occurrences in the Simulations*

		Occurrence†		Occurrence (%)‡					
Donor atom	Acceptor atom	X-ray	NMR	SimG	SimA	SimW	SimD	SimB/E	SimB
Backbone-backbone hydrogen bonds									
8CysN	5AspO	+	−	—	—	—	69	—	—
9AsnN	34CysO	+	−	68	66	84	84	85	57
10LysN	7IleO	+	−	—	58	77	67	73	53
12CysN	32ArgO	+	+	86	85	94	91	90	80
17AspN	14ThrO	−	−	49	11	74	66	51	24
18CysN	15HisO	−	−	60	72	46	44	32	67
19SerN	16AspO	−	+	—	—	—	—	—	—
21AlaN	18CysO	−	+	—	—	—	—	—	—
21AlaN	19SerO	−	−	—	—	—	55	—	—
23PheN	21AlaO	+	−	—	—	—	—	—	—
26AlaN	35GlyO	+	+	77	86	—	51	85	86
28TrpN	26AlaO	−	+	—	—	—	—	—	—
28TrpN	31AlaO	−	−	—	59	—	—	—	—
28TrpN	33ThrO	−	−	90	—	65	54	74	71
30SerN	28TrpO	+	−	—	—	—	—	—	—
31AlaN	28TrpO	−	+	—	—	—	—	—	—
32ArgN	28TrpO	−	+	—	—	—	—	—	—
32ArgN	29AsnO	−	−	—	—	60	—	—	—
33ThrN	28TrpO	−	−	—	—	61	—	—	—
33ThrN	31AlaO	−	−	68	80	—	57	62	72
34CysN	10LysO	+	+	87	90	98	95	89	93
35GlyN	26AlaO	+	+	94	88	95	96	96	95
37TyrN	23PheO	−	−	—	—	75	—	—	—

*The simulation time interval 150–500 ps has been used for the statistics.
†Presence (+) or absence (−) of a particular hydrogen bond.
‡Only those hydrogen bonds appearing in at least 50% of the analyzed conformations have been shown. For the hydrogen bonds 17AspN–14ThrO and 18CysN–15HisO, corresponding to a 3_{10}-helix, the occurrences are shown independently of the value.

change in the water–nonpolar atom interaction, it is insensitive to the other force-field modifications tested.

In Figure 4, the volume of the protein and the volume of the box are plotted against simulation time. The simulations have been grouped in pairs to facilitate the comparison of the effects of closely related parameter sets. For instance, psG and psW only differ in the $C12^{1/2}(O_w,O_w)$ parameter; psA and psD differ in the $C12^{1/2}(O_{cx},O_{cx})_{hb}$ and $C12^{1/2}(O_h,O_h)_{hb}$ parameters, and SimB/E and SimB use the same van der Waals parameters but a different water model. Units are given here in nm^3 for convenience. The initial values are 7.90 nm^3 for the pro-

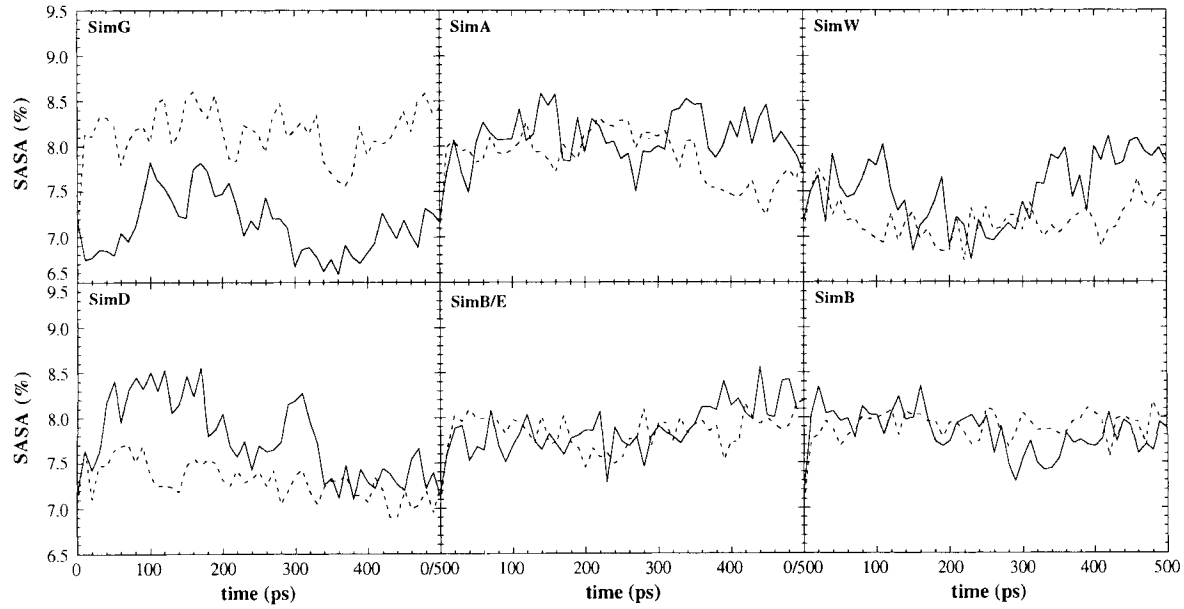


Fig. 3. Percentage of solvent-accessible surface area along simulation time. The solid lines show the percentage of SASA of protein polar atoms with respect to their total surface area, which is the sum of the surface areas of the isolated atoms. The broken lines show the percentage of SASA of protein nonpolar atoms with respect to their total surface area. The crystallographic values are 7.17 and 7.05%, respectively.

TABLE VI. Average Water Density Around the Protein Polar Atoms and Around the Protein Nonpolar Atoms

Class of atoms	SimG	SimA	SimW	SimD	SimB/E	SimB
Polar	57	42	40	39	43	40
Non-polar	51	24	23	21	25	24

The simulation time interval 150–500 ps has been used for the statistics. A cut-off radius of 0.35 nm beyond the van der Waals radius of every protein atom has been used to define the volume over the solvent-accessible surface area in which the number of water atoms is counted. The units are in water atoms per nm^3.

tein volume, and 42.73 nm^3 for the box volume. In SimG the increase of the volume of PCI (about 0.4 nm^3) is not reflected in the volume of the box. This indicates that the swelling of the protein is accompanied by water molecules entering the resulting free volume, or the expansion of the protein might even be induced by the penetration of water molecules into the core. In SimW, where the repulsion between water and CH_n groups has been enhanced, the situation is completely different. A slight increase in the volume of the PCI (about 0.2 nm^3) is accompanied by a larger increase of the volume of the box (about 0.9 nm^3). This is in part expected because of the increase in the distance between the protein surface atoms and the first shell of water molecules. The differences between SimA and SimD are predominantly in the volume of the inhibitor. The increase in the volume of the box is similar in both simulations and close to the change observed in SimW. The PCI increases in volume by about 0.6 nm^3 in SimA and 0.3 nm^3 in SimD. SimD follows a pattern similar to SimW. In SimA, the expansion of the protein has less influence on the volume of the box. In SimB/E and SimB, the influence of the difference in the magnitude of the dipole moment in the two water models can be sensed. The volume of PCI is similar in both simulations; in contrast, the volume of the box increases by about 0.8 nm^3 in SimB and slightly decreases in SimB/E. It is expected that the volume of the box is smaller in SimB/E than in SimB as the density of the bulk SPC/E water is slightly higher than that of bulk SPC water.[33] In addition, interactions with polar groups of the protein are slightly more favored when using the SPC/E water model. Thus, the water molecules may be able to approach closer to the protein surface, although this is not obvious from the measure of the water density around the polar atoms.

To investigate the variability that might be expected from different equivalent finite-time simulations, we have performed a control simulation ($SimG_c$), 500 ps long, using the same force-field parameters and settings than in SimG but different starting velocities. In this control simulation the mean RMS positional deviation of the backbone atoms of residues 8–34, calculated for the time interval 150–500 ps, is 0.99 Å with respect to the X-ray

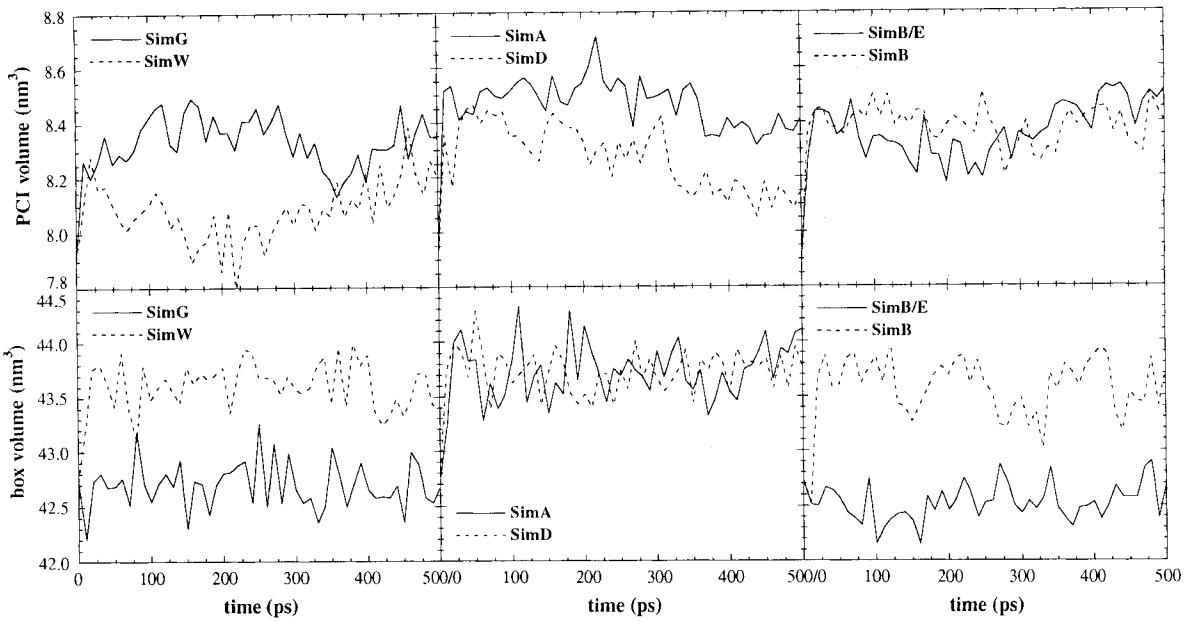


Fig. 4. Volume of the PCI and volume of the box along simulation time. The initial values are 7.90 nm^3 for the volume of the PCI, and 42.73 nm^3 for the volume of the box.

structure (1.58 Å in SimG, Table III) and 1.24 Å with respect to the NMR structure (1.66 Å in SimG). According to the RMS measurements, the different starting velocities led to the sampling of different regions of the conformational space in simulations SimG and $SimG_c$, denoting, as expected, that a 500 ps simulation is not enough to characterize the complete conformational space accessible to the PCI under a given force field. Nevertheless, other properties of the molecular system relax faster under the force-field influence and are less dependent on the size of the sampled conformational space, for example, properties related with the solvation of the protein. Thus, the percentage of solvent-accessible surface area of polar and nonpolar atoms, the water density around polar and nonpolar atoms, the volume of PCI, the volume of the box, and the ratio of the number of protein–water intermolecular hydrogen bonds to the number of protein–protein intramolecular hydrogen bonds per conformation, all averaged over the time interval 150–500 ps, are very similar in $SimG_c$ and SimG: 7.55 vs. 7.11% (SASA of polar atoms, Fig. 3), 8.35 vs. 8.15% (SASA of nonpolar atoms, Fig. 3), 56 vs. 57 water-atoms/nm^3 (density around polar atoms, Table VI), 55 vs. 51 water-atoms/nm^3 (density around nonpolar atoms, Table VI), 8.5 vs. 8.3 nm^3 (volume of PCI, Fig. 4), 42.5 vs. 42.7 nm^3 (volume of the box, Fig. 4), and 6.7 vs. 6.0 intermolecular/intramolecular hydrogen bonds (Table IV), respectively. These results show that the differences observed between the trajectories obtained with the various studied sets of parameters (SimG, SimA, SimW, SimD, SimB/E, and SimB) are significant.

CONCLUSIONS

From the analysis of the trajectories it is evident that, of the properties of the system that have been investigated, only when evaluating the solvent-accessible surface area of polar and nonpolar atoms, or possibly the water density around these atoms, is the advantage of using one of the modified parameter sets, as opposed to the original van der Waals parameters (psG) of the GROMOS force field, obvious.

Comparison of mean RMS positional deviations from the crystallographic structure of the globular core of the protein suggests that use of the parameter set psW best preserves the native structure of PCI. In addition, comparison of the protein–water/protein–protein hydrogen-bond ratios and protein/box volume ratios shows that in SimW internal packing is most favored. Nevertheless, there is evidence, for example from the behavior of the N-terminal tail, that the stability in SimW can be partially due to an excessive repulsion between CH_n groups and water molecules. On the other hand, the increased $C12^{1/2}(i,i)$ parameters used for the protein oxygens in simulations SimA, SimB/E, and SimB, appear to produce the inverse effect. A loss of the original internal packing is observed. This could be due to an overweighed repulsion between protein oxygens and CH_n groups and between pairs of carbonyl oxygens, which would introduce internal instabilities.

psA and psD differ only in the $C12^{1/2}(i,i)_{hb}$ parameters of the carboxyl and hydroxyl oxygens. The differences in the trajectories obtained in SimA and

SimD are therefore surprising. SimA shows the greatest degree of denaturation, while SimD best reproduces the experimental data overall. Despite the agreement between the results of SimD and the experimental data, we cannot recommend the use of psD since the choice of the parameters was in some respects arbitrary, and it does not represent a consistent parameter set. The differences between SimA and SimD and between SimG and $SimG_c$ highlight, from different perspectives, the degree of structural variability that can be observed in two different trajectories starting from the same conformation, and caution against the over-interpretation of results from a single simulation albeit 500 ps long.

Simulations SimB/E and SimB show that, for the molecular system studied, using the parameter set psB there is no apparent advantage in using either the SPC or the SPC/E water models. The only simulation results that could be directly related to the different dipole moment of the SPC and SPC/E models could be found in that in SimB/E hydrogen bonds between the protein and water are slightly more favored than in SimB, and in that the volume of the box is smaller in SimB/E.

In summary, we have seen that while the effects of changing the $C12^{1/2}$(i,i) van der Waals parameter of the water oxygen can be observed by comparison of the simulation results, the effects of changes in the van der Waals parameters of the protein oxygen types are more difficult to characterize. This is especially true when more than one parameter is changed at the same time. Although we must be cautious about basing general conclusions on only one protein case and one set of simulation conditions, the simulations suggest that taking a value of 793.3 (Kcal mol^{-1} $Å^{12})^{1/2}$ for the $C12^{1/2}$(i,i) van der Waals parameter of the water oxygen most probably yields an overestimation of the repulsion energy between water and aliphatic/aromatic groups. It is, however, a clear improvement of the treatment of these interactions with respect to the original 421.0 (Kcal mol^{-1} $Å^{12})^{1/2}$ value.

From both the results of the simulations studied here and the above discussed internal coherency of the analyzed parameter sets, we conclude that the set referred to as psW—the one currently in use in GROMOS—should be recommended until new parameters become available.

ACKNOWLEDGMENTS

The authors thank Dr. A.E. Mark for critical revision of this manuscript, and for providing access to reference 13 before publication. We also thank Dr. M. Clore for providing the coordinates of the NMR structure of PCI. X. Daura is a fellowship recipient from the Programa de Química Fina (CIRIT). This work has been supported by grants BIO92-0458, BIO94-0912, and BIO95-0848 from the CICYT (Ministerio de Educación y Ciencia). The authors also acknowledge support from Fundación Francisca de Roviralta, from the Centre de Supercomputació de Catalunya (CESCA), and from the Centre de Referència en Biotecnologia (Generalitat de Catalunya). O. Tapia acknowledges grants from the Swedish Research Council (NFR).

REFERENCES

1. Karplus, M., Petsko, G.A. Molecular dynamics simulations in biology. Nature 347:631–639, 1990.
2. van Gunsteren, W.F. Molecular dynamics studies of proteins. Curr. Opin. Struct. Biol. 3:277–281, 1993.
3. Kollman, P.A., Merz, K.M. Computer modeling of the interactions of complex molecules. Acc. Chem. Res. 23:246–252, 1990.
4. Burt, S.K., Hutchins, C.W., Greer, J. Predicting receptor-ligand interactions. Curr. Opin. Struct. Biol. 1:213–218, 1991.
5. Brooks III, C.L. Molecular simulations of peptide and protein unfolding: In quest of a molten globule. Curr. Opin. Struct. Biol. 3:92–98, 1993.
6. Alonso, D.O.V., Daggett, V. Molecular dynamics simulations of protein unfolding and limited refolding: Characterization of partially unfolded states of ubiquitin in 60% methanol and in water. J. Mol. Biol. 247:501–520, 1995.
7. Hünenberger, P.H., Mark, A.E., van Gunsteren, W.F. Computational approaches to study protein unfolding: Hen egg white lysozyme as a case study. Proteins 21:196–213, 1995.
8. Åqvist, J., Medina, C., Samuelsson, J.-E. A new method for predicting binding affinity in computer-aided drug design. Protein Eng. 7:385–391, 1994.
9. Åqvist, J., Fothergill, M., Warshel, A. Computer simulation of the CO2/HCO3-interconversion step in human carbonic anhydrase I. J. Am. Chem. Soc. 115:631–635, 1993.
10. van Buuren, A.R., Marrink, S.-J., Berendsen, H.J.C. A molecular dynamics study of the decane/water interface. J. Phys. Chem. 97:9206–9212, 1993.
11. van Buuren, A.R., Berendsen, H.J.C. Molecular dynamics simulation of the stability of a 22-residue α-helix in water and 30% trifluoroethanol. Biopolymers 33:1159–1166, 1993.
12. Mark, A.E., van Helden, S.P., Smith, P.E., Janssen, L.H.M., van Gunsteren, W.F. Convergence properties of free energy calculations: α-Cyclodextrin complexes as a case study. J. Am. Chem. Soc. 116:6293–6302, 1994.
13. Smith, L.J., Mark, A.E., Dobson, C.M., van Gunsteren, W.F. Comparison of MD simulations and NMR experiments for hen lysozyme. Analysis of local fluctuations, cooperative motions and global changes. Biochemistry 34: 10918–10931, 1995.
14. van Gunsteren, W.F., Berendsen, H.J.C. Groningen Molecular Simulation (GROMOS) Library Manual, BIOMOS, Nijenborgh 4, 9747 AG Groningen, The Netherlands, 1987.
15. Waldman, M., Hagler, A.T. New combining rules for rare gas van der Waals parameters. J. Comput. Chem. 14: 1077–1084, 1993.
16. Hermans, J., Berendsen, H.J.C., van Gunsteren, W.F., Postma, J.P.M. A consistent empirical potential for water-protein interactions. Biopolymers 23:1513–1518, 1984.
17. Zeiss, G.D., Meath, W.J. The H2O-H2O dispersion energy constant and the dispersion of the specific refractivity of dilute water vapour. Mol. Phys. 30:161–169, 1975.
18. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J. Interaction models for water in relation to protein hydration. In: "Intermolecular Forces." Pullman, B., (ed.). Dordrecht: Reidel, 1981:331–342.
19. Nelson, D.J., Hermans, J. Non-bonded interatomic potential functions and crystal structure. Correction of the functions for use with macromolecules and application to polipeptide helixes. Biopolymers 12:1269–1284, 1973.
20. Ferro, D.R., Hermans, J. Nonbonded interatomic potential functions and crystal structure: Non hydrogen-bonded organic molecules. In: "Liquid Crystals and Ordered Fluids."

Johnson, J.F., Porter, R., (eds.). New York: Plenum, 1970: 259–275.
21. Lifson, S., Hagler, A.T., Dauber, P. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O···H-hydrogen bonds. J. Am. Chem. Soc. 101:5111–5121, 1979.
22. Hagler, A.T., Lifson, S., Dauber, P. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 2. A benchmark for the objective comparison of alternative force fields. J. Am. Chem. Soc. 101:5122–5130, 1979.
23. Poland, D., Scheraga, H.A. Energy parameters in polypeptides. I. Charge distributions and the hydrogen bond. Biochemistry 6:3791–3800, 1967.
24. Dunfield, L.G., Burgess, A.W., Scheraga, H.A. Energy parameters in polypeptides. 8. Empirical potential energy algorithm for the conformational analysis of large molecules. J. Phys. Chem. 82:2609–2616, 1978.
25. Gelin, B.R., Karplus, M. Side-chain torsional potentials: Effect of dipeptide, protein, and solvent environment. Biochemistry 18:1256–1268, 1979.
26. Jorgensen, W.L., Madura, J.D., Swenson, C.J. Optimized intermolecular potential functions for liquid hydrocarbons. J. Am. Chem. Soc. 106:6638–6646, 1984.
27. Clore, G.M., Gronenborn, A.M., Nilges, M., Ryan, C.A. Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. Biochemistry 26:8102–8023, 1987.
28. Rees, D.C., Lipscomb, W.N. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. J. Mol. Biol. 160:475–498, 1982.
29. Molina, M.A., Marino, C., Oliva, B., Avilés, F.X., Querol, E. C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. J. Biol. Chem. 269:21467–21472, 1994.
30. Chang, J.-Y., Canals, F., Schindler, P., Querol, E., Avilés, F.X. The disulfide folding pathway of potato carboxypeptidase inhibitor. J. Biol. Chem. 269:22087–22094, 1994.
31. Querol, E., Molina, M.A., Daura, X., Oliva, B., Marino, C., Canals, F., Crane-Robinson, C., Tapia, O. Protease inhibitors from vegetables as a target for protein engineering: Application to the potato carboxypeptidase inhibitor. In: "Innovations in Proteases and Their Inhibitors." Avilés, F.X., (ed.). Berlin: Walter de Gruyter, 1993:477–493.
32. Oliva, B., Daura, X., Querol, E., Avilés, F.X., Tapia, O. Structure and atomic fluctuation patterns of potato carboxypeptidase A inhibitor protein. A molecular dynamics study in water. Eur. Biophys. J. 24:1–11, 1995.
33. Berendsen, H.J.C., Grigera, J.R., Straatsma, T.P. The missing term in effective pair potentials. J. Phys. Chem. 91:6269–6271, 1987.
34. Hass, G.M., Nau, H., Biemann, K., Grahn, D.T., Ericsson, L.H., Neurath, H. The amino acid sequence of carboxypeptidase inhibitor from potatoes. Biochemistry 14:1334–1342, 1975.
35. Hass, G.M., Ryan, C.A. Cleavage of the carboxypeptidase inhibitor from potatoes by carboxypeptidase A. Biochem. Biophys. Res. Commun. 97:1481–1486, 1980.
36. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81:3684–3690, 1984.
37. Ryckaert, J.-P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23:327–341, 1977.
38. Pascual-Ahuir, J.L., Silla, E., Tomasi, J., Bonaccorsi, R. Electrostatic interaction of a solute with a continuum. Improved description of the cavity and of the surface cavity bound charge distribution. J. Comput. Chem. 8:778–787, 1987.