The Effects of Internal Water Molecules on the Structure and Dynamics of Chymotrypsin Inhibitor 2

Hongxing Lei and Paul E. Smith*

Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506-3702

Received: October 25, 2002

Chymotrypsin Inhibitor 2 (CI2) is a small inhibitor that blocks the action of serine proteases by positioning a reactive loop in the protease active site. NMR studies of this inhibitor indicate that the reactive loop residues are quite rigid (high order parameters) and yet there is a shortage of the required NOE data to fully describe the structure of the loop in solution. In addition, four internal water molecules are observed in the crystal structure, resulting in an irregular sheet structure linking the ends of the reactive loop. To further understand the role of the internal water molecules and the nature of this loop region, we have performed molecular dynamics simulations of CI2 using the GROMOS force field. Simulations of 10 ns each were performed starting from two different initial conformations corresponding to the X-ray and NMR structures. Both simulations converged to a common structure. All secondary structural elements remained intact with order parameters in agreement with experiment. Loop regions displayed larger fluctuations than described by the experimental NMR order parameters. One-half of the reactive loop is supported by a hydrophilic arginine cluster and was stable during the simulations. The second half of the loop, which is supported by a hydrophobic cluster, deviated from the X-ray and NMR derived structures. This difference could be directly traced to the loss of internal water molecules during the simulation. A simulation performed with one of the internal water molecules restrained to occupy a hydration site between the two irregular sheet strands resulted in stability of both the hydrophobic cluster and the reactive loop. Hence, the internal water molecules of CI2 appear to be essential for maintaining the structure of the reactive loop, even though they are spatially well removed from the loop itself.

Introduction

Serine proteases have multiple biological activities including the digestion of proteins and hormone processing.[1] Chymotrypsin inhibitor 2 (CI2) is a single chain 83 residue protein that belongs to the potato inhibitor family that regulates the activity of serine proteinases.[2] As CI2 is small in size with a well characterized function, it has been extensively studied by structural techniques,[3−9] mutagenesis,[10−14] folding and unfolding experiments,[15−17] and simulation.[18−32] Of particular importance for understanding the mechanism of inhibition, and hence our ability to design improved proteinase inhibitors, is the structure of one of the major loops in the protein (the reactive loop) that interacts directly with the proteinase and contains the hydrolysis site.[10] The structure and flexibility of the reactive loop in this family of inhibitors has been the subject of several experimental studies,[9,10,33] which have demonstrated that the hydrogen bonding and salt bridge network supporting the reactive loop are essential for the inhibitory activity of the protein.[10,33] Even so, the exact structure and dynamics of the reactive loop currently remain uncertain.

The structure of CI2 has been determined by X-ray crystallography at 2.0 Å resolution[3] and is essentially identical to the X-ray structure of CI2 bound to subtilisin.[34] The structure of CI2 is displayed in Figure 1. The first 19 N terminal residues are unstructured in both crystals. The rest of the protein includes a core α-helix (residues 31−43), two major hydrogen bonded β-strands (residues 47−53 and 64−71), and two short hydrogen bonded β-strands (residues 22−24 and 80−83). A long reactive loop (residues 54−63) with sequence GTIVTMEYRI links the two major β-strands. The peptide bond between Met59 and Glu60 sits in the active site of the protease and is slowly hydrolyzed.[35] The Arg65, Arg67, and the Gly83 C terminus form a salt bridge cluster helping to support the loop, as illustrated in Figure 1. The two arginines are well conserved among other

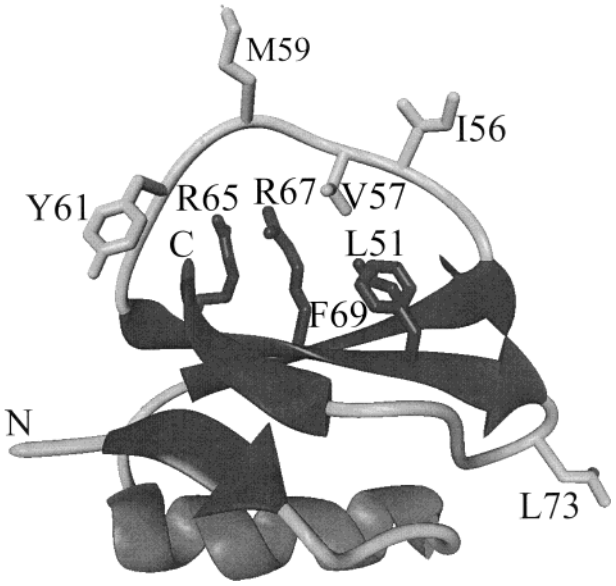

Figure 1. Cartoon representation of the crystal structure of CI2. Some structurally important residues are represented as sticks. Arg65, Arg67, and the C terminus form a salt bridge network. Leu51, Val57, and Phe69 form a hydrophobic core. Ile56, Met59, and Leu73 make contacts with other residues in the crystal unit cell.

* Corresponding author. Fax: 785-532-7278. E-mail: pesmith@ksu.edu.

10.1021/jp0223105 CCC: $25.00
Published on Web 01/16/2003

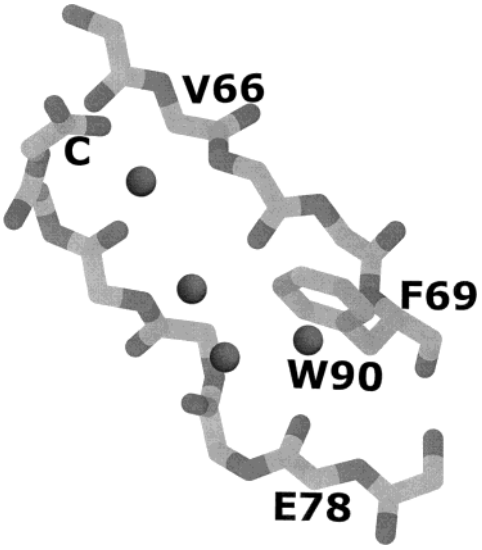

Figure 2. Internal waters of CI2 as found in the crystal structure. Single spheres represent water oxygens. The four internal water molecules bridge two β-strands located between the ends of the reactive loop. The most buried water molecule is residue 90, which bridges residues Phe69 and Glu78 (actually Gln78 in the NMR structure and current simulations). The aromatic ring of Phe69 sits above water residue 90.

members of the potato inhibitor I family[10] and are thought to play an important role in the stability and activity of the reactive loop.[36] Also helping to support the loop are Leu51, Val57, and Phe69, which form a hydrophobic core of residues (see Figure 1). Four internal water molecules are located between the major and minor β-strands leading to an irregular sheet structure as shown in Figure 2. In particular, one water molecule (residue 90 in the crystal structure) is involved in mediating hydrogen bonds between the backbone atoms of residues Phe69 and Gln78. Several crystal contacts between different protein molecules in the unit cell are observed. Most of these contacts are hydrophobic in nature and include Ile56 to Ile56, Tyr61 to Tyr61, Met59 to Phe69, and Leu73 to Val50, together with a salt bridge between Asp74 and Lys72. As expected, the B factors for the loop residues are much higher than the rest of the protein (see Figure 3).[3] Low B factors are observed for Leu51 and F69 of the hydrophobic cluster, which may serve to anchor the ends of the reactive loop.

CI2 is considered a monomer in solution,[37] although some studies suggest that the potato inhibitor I family exist as dimers.[38] The solution structure of CI2 (minus the first 19 N terminal residues) has been determined by nuclear magnetic resonance (NMR) and displays both consistency with and deviation from the crystal structure.[9,39,40] The major difference lies in the uncertainty of the loop structure, although the experimental order parameters are relatively high (typically >0.70) for all the loop residues (see Figure 3).[5] The hydrophobic cluster is defined by a strong observed nuclear Overhauser effect (NOE) between the H^{γ} of Val57 and H^{ζ} of Phe69.[40] Several side chains of the reactive loop residues are poorly defined by the NMR study, and hence the exact structure and dynamics of this important region in solution is still uncertain. The internal water molecules were not detected in early NMR studies but have been confirmed by recent techniques.[7] A detailed comparison of the crystal and solution forms of CI2 is presented by Ludvigsen et al.[9]

CI2 represents an interesting model system for study by simulation as it is: (i) small in size; (ii) has a reasonably well-defined structure; (iii) has no disulfide bonds or a net charge; (iv) displays important structural features in the functional loop

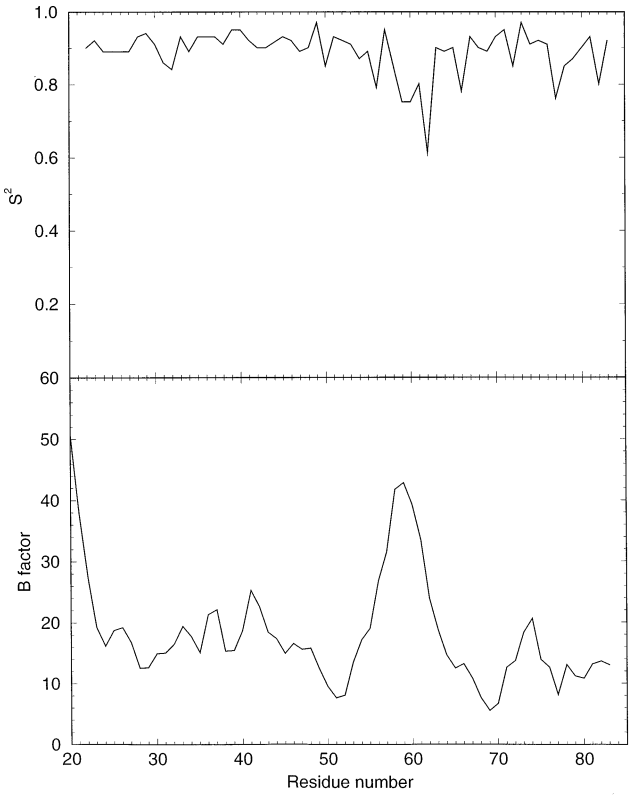

Figure 3. Experimental NMR NH order parameters (top) and isotropic crystallographic B factors for C^{α} atoms (bottom in $Å^2$) for CI2.

that are both hydrophilic and hydrophobic in nature; and (v) contains internal water molecules whose exact role is currently unclear. Here we present results from 10 ns molecular dynamics simulations of CI2 in explicit solvent in an attempt to rationalize the structure and dynamics of the reactive loop. In particular, we were interested in determining the role of the internal water molecules and understanding the structure and flexibility of the reactive loop region. The latter is especially interesting in light of the fact that the NMR order parameters for these residues are almost as high as for the secondary structure residues (suggesting rigidity), and yet the NMR refinement lacks sufficient NOE data to produce a consistent structure (suggesting flexibility).

The majority of the previous simulation studies of CI2 have focused on determining the folding and unfolding pathways for the protein.[19,20,25,26,31,32] However, the structure and dynamics of CI2 have been examined by a 5.3 ns molecular dynamics (MD) simulation using the ENCAD force field.[19] The simulation was consistent with NMR experiments as defined by the calculated order parameters, the majority of the NOEs, and the observed correlation between amide hydrogen exchange behavior and solvent accessibility. However, the strong NOE observed between the side chains of Val57 and Phe69 was lost during the simulation. Furthermore, just two of the internal water molecules were retained during the simulation, and their positions were different from those observed in the crystal structure. It is therefore interesting to determine if these problems are specific to the ENCAD force field and/or have any effect on the structurally important reactive loop.

Three simulations of CI2 were performed in an effort to rationalize the structure and dynamics of the reactive loop and investigate the role of the internal water molecules. The first (MDCRY) was initiated from the crystal structure with the internal water molecules in place. The second (MDNMR) was

initiated from one of the NMR model structures and contained no internal water molecules. As the internal water molecules were not stable during the simulation starting from the crystal structure, another simulation (MDRES) was performed in which one of the internal water molecules was constrained to remain in the crystal arrangement. A comparison of these three simulations allows one to investigate the structure and dynamics of the protein, as described by the GROMOS force field, and to examine any effects of the internal water molecules.

Methods

All the simulations were performed using the GROMOS96 program and the 43A1 force field.[41] Three simulations were performed. The initial coordinates for the first simulation corresponded to the X-ray determined structure (PDB entry 2CI2) and included all crystal water molecules. The initial structure for the second simulation corresponded to model 2 of the NMR determined ensemble (PDB entry 3CI2). Model 2 was used, as it displayed the best agreement with the crystal structure for the structurally important C terminal salt bridge cluster. A third simulation was also initiated from the X-ray structure. During the course of this simulation an internal water molecule (residue 90) was restrained to be 0.294 nm from backbone O of residue Gln78 and 0.276 nm from the backbone N of residue Phe69, as observed in the crystal structure, with a harmonic force constant of 1000 kJ/mol/nm^2.

The disordered residues (1−19) were not included in any of the simulations. Residue 78 is Glu in the X-ray structure but Gln in the NMR structure. However, the NMR sequence was used in all the simulations in an effort to maintain consistency. Charged residues were considered to be protonated (Arg and Lys) or deprotonated (Asp and Glu) corresponding to neutral pH. The net charge of the CI2 fragment is zero (using Gln78) and therefore no counterions were necessary. The SPC water model was chosen for the solvent potential and a truncated octahedron was used for the simulation box. The minimum distance from any protein atom to the edge of the box was 1.0 nm. This typically resulted in approximately 3000 water molecules around the protein and an initial box length of 6.0 nm. Each system contained approximately 10 000 atoms (641 protein atoms). The time step was 2 fs, and SHAKE[42] was used to constrain all bond lengths. A twin range cutoff of 0.8 nm/1.4 nm was employed and the nonbonded group pair list was updated every 10 steps. Long-range electrostatics were treated using a reaction field approach,[43] with a reaction field permittivity of 54.[44]

Each system was minimized using 200 steps of steepest descent minimization while the protein (and water) atom positions were restrained. Equilibration of the systems involved a 10 ps simulation at 300 K including the protein (and water) atom restraints, followed by 90 ps of free dynamics. There were no restraints during the subsequent 10 ns production period. The simulations were performed under conditions of constant temperature (300 K) and constant pressure of (1 atm) using the weak coupling approach.[45] Coordinates of the protein and water were saved every 1 ps for analysis. A summary of the simulations is displayed in Table 1.

Order parameters (S^2) were determined, after translation and rotational fitting to the initial structure, according to[46]

$$S^2 = {}^1/_2[3\sum_{\alpha=1}^{3}\sum_{\beta=1}^{3}\langle\mu_\alpha\mu_\beta\rangle - 1]$$

where μ_α or μ_β denotes the x, y, or z component of the NH unit vector and the angular brackets indicate an average over the trajectory. The above approach is statistically more reliable than fitting the long time decay of the NH orientation correlation function.[47,48] Solvent accessible surface areas (ASA) were determined using the method of Lee and Richards[49] with a probe radius of 0.14 nm. Protein atomic radii were defined by the σ_{ii} values extracted from the 43A1 force field. Uncharged atoms and aromatic carbons were classified as hydrophobic.

TABLE 1: Summary of the Molecular Dynamics Simulations[a]

	MDCRY	MDNMR	MDRES
simulation length (ns)	10	10	5
temperature (K)	300	300	300
pressure (atm)	0.9	1.1	0.9
protein−protein PE	−5.20	−5.11	−5.32
protein−solvent PE	−8.35	−8.49	−8.10
total ASA (nm^2)	45.8	45.4	45.8
hydrophilic ASA (nm^2)	13.1	12.2	12.3
hydrophobic ASA (nm^2)	32.7	33.1	33.5

[a] All the simulation properties represent an average over the last nanosecond of each simulation. Potential energies (PE) are in MJ/mol. The total, hydrophilic, and hydrophobic solvent accessible surface areas (ASA) are 48.6, 10.2, and 38.4 nm^2, respectively, for the crystal structure and 44.0, 6.5, and 37.5 nm^2, respectively, after averaging over the 20 NMR structures.

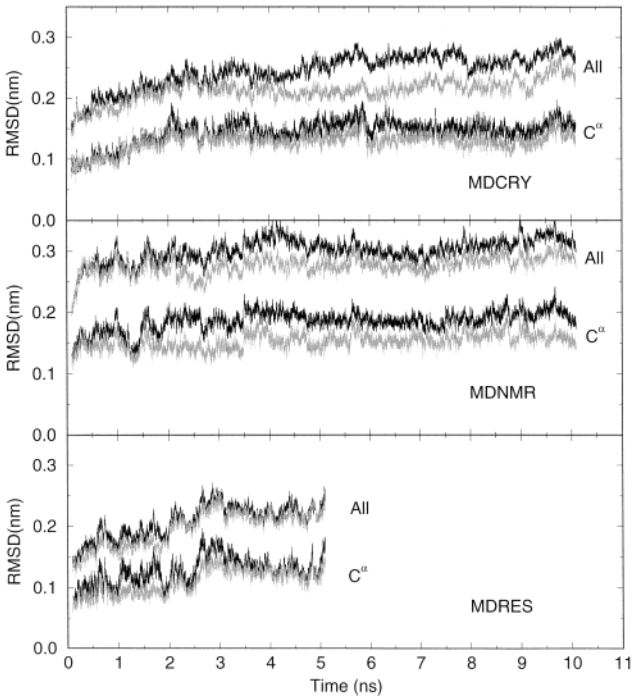

Figure 4. Root-mean-square deviation (RMSD) as a function of time for the MDCRY (top), MDNMR (middle), and MDRES (bottom) simulations. The respective minimized structures were used as references. The deviations of the whole protein are shown as dark lines, whereas the lighter lines represent the deviations excluding the reactive loop. The deviations of both all atom and Cα atoms are shown.

Results

General Features of the Simulations. The root-mean-square deviation (RMSD) of the atoms from the respective initial structures is displayed in Figure 4. The Cα RMSD for the simulation starting from the crystal structure (MDCRY) increased to 0.15 nm in the first 3 ns and then fluctuated around this value for the rest of the simulation. However, the all atom RMSD did not remain constant after 2 ns and increased

TABLE 2: Selected Distances for CI2[a]

atom pairs		crystal	NMR	MDCRY	MDNMR	MDRES
			Arginine Cluster			
Arg65(C^{ζ})	Arg67(C^{ζ})	0.41	0.66(0.19)	0.65(0.03)	0.86(0.08)	0.76(0.08)
Arg65(C^{ζ})	Gly83(C)	0.41	0.78(0.26)	0.42(0.01)	0.44(0.04)	0.58(0.12)
Arg67(C^{ζ})	Gly83(C)	0.44	0.60(0.28)	0.38(0.01)	0.53(0.05)	0.42(0.04)
			Hydrophobic Cluster			
Val57(C^{β})	Leu51(C^{γ})	0.65	0.61(0.09)	0.55(0.11)	0.58(0.12)	0.57(0.05)
Val57(C^{β})	Phe69(C^{ζ})	0.53	0.44(0.02)	0.98(0.08)	0.78(0.12)	0.51(0.13)
Leu51(C^{γ})	Phe69(C^{ζ})	0.49	0.37(0.04)	0.58(0.06)	0.57(0.06)	0.49(0.07)
			Irregular Sheet			
Gly83(N)	Arg65(O)	0.30	0.34(0.07)	0.29(0.02)	0.28(0.01)	0.34(0.06)
Arg81(N)	Arg67(O)	0.44	0.38(0.06)	0.32(0.02)	0.32(0.03)	0.33(0.03)
Arg67(N)	Arg81(O)	0.40	0.30(0.04)	0.28(0.01)	0.29(0.02)	0.28(0.02)

[a] Average distances (nm) and fluctuations in parentheses. All 20 NMR structures and the last 1 ns of the MD simulations were used to calculate the average distances. A strong NOE exists between the H^{γ} of Val57 and H^{ζ} of Phe69.

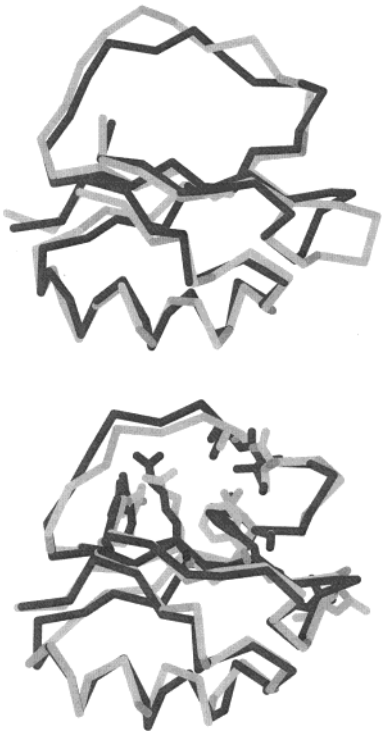

Figure 5. Superposition of the C^{α} atoms from the initial and final structure of MDCRY (top). Superposition of the C^{α} atoms from the final structures of MDCRY and MDNMR (bottom). Some side chains are shown in stick to further highlight the consistency of the two simulations.

substantially from 4 to 6 ns (0.23 to 0.28 nm). The changes were mostly due to the side chain rearrangements in the reactive loop region because the C^{α} RMSD excluding the loop residues remained stable during the same period. Slightly larger deviations were observed for the simulation starting from the NMR structure (MDNMR). Some of this deviation was undoubtedly due to the uncertainty in the loop structure as measured by NMR. The restrained water simulation appeared to converge more quickly than the other two, as the RMSD remained constant after 2.5 ns. The radius of gyration decreased slightly from 1.11 nm to approximately 1.08 nm during the simulations.

A comparison between the C^{α} trace of the crystal structure and the final MDCRY configuration is displayed in Figure 5. The two structures were very similar with the exception of a change in the small loop between residues 70 and 74, and some variation in the reactive loop between residues 54 and 59. All the secondary structural elements remained intact. A comparison of the C^{α} trace and some of the loop residues obtained from the end of the MDCRY and MDNMR simulations is also displayed in Figure 5. The agreement between the two simulations (C^{α} RMSD of 0.13 nm) was excellent, indicating that the small structural changes observed during the simulations were converged and reproducible, and hence sampling problems were not significant.

The four internal waters initially located between one of the major and minor strands at the base of the major loop did not remain in place during the MDCRY simulation. In fact, the waters were displaced within the first 100 ps after equilibration. Even though the internal waters are accessible to solvent in the crystal structure, no water molecules were ever observed in the correct binding sites for the remainder of the MDCRY simulation or the whole MDNMR simulation. Loss of the internal water molecules resulted in the Phe69 H to water hydrogen bond being replaced with a weak Gln78 O interaction (0.40 nm from 0.53 nm), resulting in the conversion of the irregular sheet region to a more regular sheet arrangement (Pro80 prevents a classical arrangement). The changes are highlighted in Table 2 where some of the sheet hydrogen bond distances are listed. Clearly, the loss of internal water resulted in shorter hydrogen bond distances as the two strands move closer together. However, all four of the internal waters remained in the vicinity of their starting positions during the restrained water simulation and the Phe69 to Val57 NOE contact was maintained (see Table 2). Even so, the irregular sheet arrangement was still unstable.

The changes in the main chain dihedral angles (with respect to the crystal structure) obtained from the experimental NMR ensemble and the three simulations are compared in Figure 6. The majority of the φ and ψ dihedral angles were very consistent between the three simulations and deviated from the crystal values to a degree similar to that in the NMR structures. This supports the view that the overall structure of the protein was well maintained during the simulations. The major differences corresponded to changes in the small loop regions (residues 44−46 and 72−74) and residues 54−59 in the major loop. The small loop between residues 72−74 was initially extended into solution. It is involved in hydrophobic and salt bridge intermolecular interaction in the crystal unit cell (Leu73 toVal50 and Asp74 to Lys72) but still remains extended in the solution structure when these interactions are removed.[9] The initially exposed loop moved out of solution within 1.5 ns and stayed close to the rest of the protein during the remaining 8.5 ns in all three simulations. The driving force for this rearrangement appeared to be a change in the classically unfavorable φ value of Asp74, which is 74° and 44° for the crystal and NMR

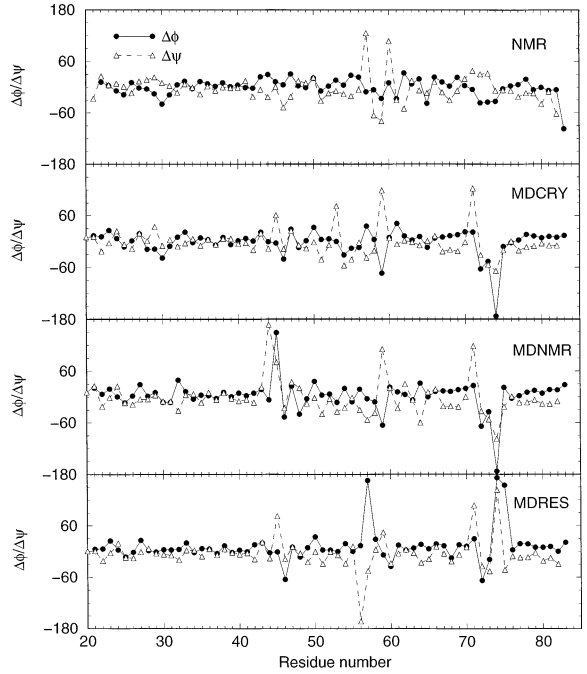

Figure 6. Deviations for the φ and ψ dihedrals from the crystal structure. The average of all the 20 structures was used for NMR data. The average from the last nanosecond of each simulation was used for MDCRY, MDNMR, and MDRES results.

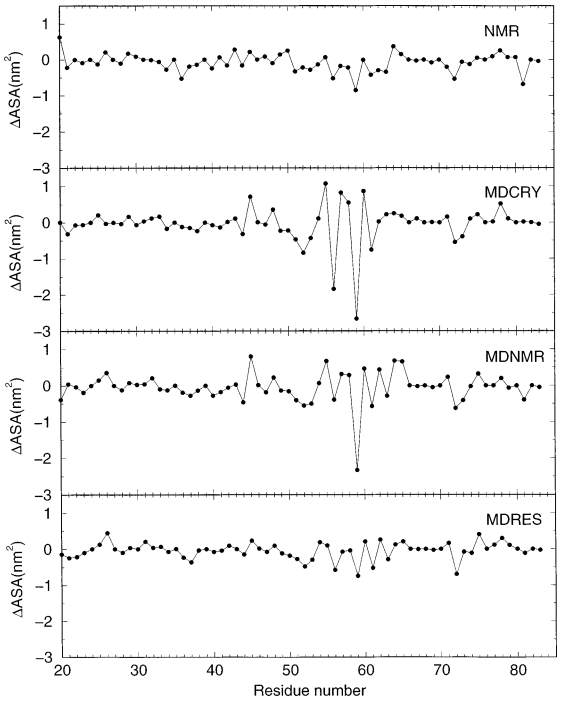

Figure 7. Deviation of the residue based average solvent accessible surface area (ΔASA) from the crystal structure. The average of all the 20 structures was used for the NMR data. The average from the last nanosecond of each simulation was used for MDCRY, MDNMR, and MDRES results.

structures, respectively. The last nanosecond of the MDCRY simulation resulted in an average value of −104° for this dihedral.

The residue based accessible surface area differences (with respect to the crystal structure) are displayed in Figure 7. It is clear from Figure 7 that several reactive loop residues were buried (Ile56, Met59, and Tyr61) at the expense of exposing other loop residues (Thr55, Val57, Thr58, and Glu60) during the unrestrained simulations. The overall effect was to decrease the hydrophobic accessible surface area, which changed from 38 to 33 nm^2 during the first 5 ns and then remained constant. In particular, the Val residue forming part of the hydrophobic core (Leu51, Val57, and Phe69) was replaced by an Ile residue leading to a Leu51, Ile56, and Phe69 cluster. The burial of both Ile56 and Met59 are expected due to their high hydrophobicity and the removal of intermolecular hydrophobic contacts in the crystal (Ile56 with Ile56 and Met59 with Phe69). Large changes in surface area of the reactive loop residues were resisted during the restrained water simulation, producing similar deviations as observed for the ensemble of NMR structures. All the simulations displayed higher hydrophilic and lower hydrophobic surface area exposure than the crystal or NMR structures (see Table 1). The presence of the restrained water appeared to maintain a larger favorable protein−protein energy, with a less favorable solvation energy (see Table 1). However, the difference in solvation energy was larger and probably promoted the release of the internal water molecules.

In summary, the simulations retain the overall structure of CI2 with all secondary structural regions being stable over 10 ns. Deviations from the experimental data were observed for a small loop and half of the reactive loop in the absence of the internal water molecules. However, restraining one of the internal water molecules was sufficient to stabilize the reactive loop structure.

Reactive Loop Structure and Dynamics. The overall structure of the loop backbone remained intact during all three simulations, as observed in Figures 4 and 5. The Arg65, Arg67, and C terminal carboxyl group form a salt bridge cluster in the crystal structure. This cluster was very stable during all the simulations. Average distances between the three residues are presented in Table 2. Both arginines remained close to the carboxylate group with the MDCRY simulation resulting in very small fluctuations in the cluster distances. The MDNMR simulation displayed more flexibility and a larger arginine to arginine distance than either the MDCRY simulation or the crystal structure. The rigidity of the hydrophilic cluster can also be quantified by the arginine side chain order parameter.[33] The N^{ϵ}-H^{ϵ} vectors of Arg65 and Arg67 both displayed order parameters above 0.8 when averaged over the last 6 ns of the MDCRY simulation. The NMR ensemble displays a large range of variation in this region due to the lack of observed NOE constraints between these residues.

The first half of the reactive loop (away from the arginine cluster) involves a series of hydrophobic interactions. Changes in this region were significant and developed between 2 and 6 ns of the unrestrained simulations. The changes were reproduced in both unrestrained simulations and involved a rearrangement of the hydrophobic core within the loop. In particular, the Val residue initially forming part of the core (Leu51, Val57, and Phe69) was replaced by an Ile residue leading to a Leu51, Ile56, and Phe69 cluster. However, after the rearrangement the distance between the Val57 C^{β} and Phe69 C^{ζ} atoms increased from 0.5 nm in the initial structure to more than 1.0 nm in the latter part of the simulation (see Table 2). This conformational change was inconsistent with the NMR data, which indicates a strong NOE between the side chains of Val57 and Phe69,[40] and was directly related to the absence of the internal water molecules. The phenyl ring of Phe69 in the final structure had moved away from the loop and occupied the space vacated by the internal water molecules. Consequently, the packing and stability of the hydrophobic core within the loop was disrupted, leading to a significant loop rearrangement. Interestingly, the disruption of

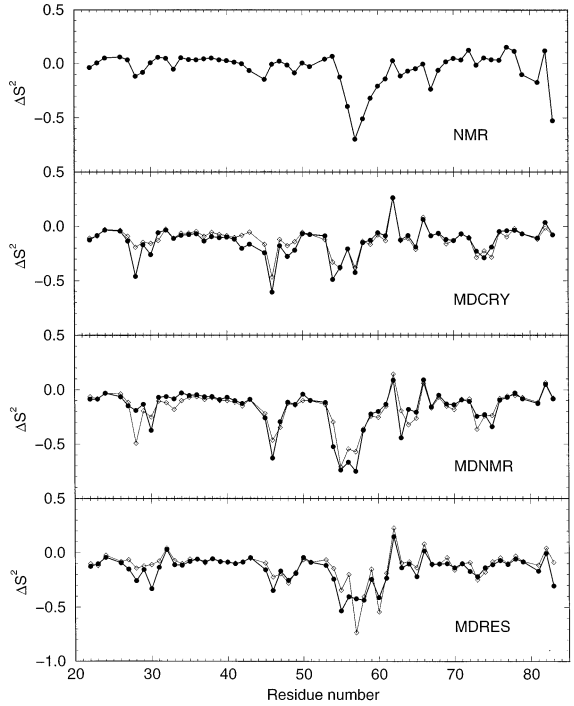

Figure 8. NH order parameter differences ($\Delta S^2 = S^2_{calc} - S^2_{exp}$) for the NMR ensemble and two 1 ns averages from the final 2 ns of the MDCRY, MDNMR, and MDRES simulations.

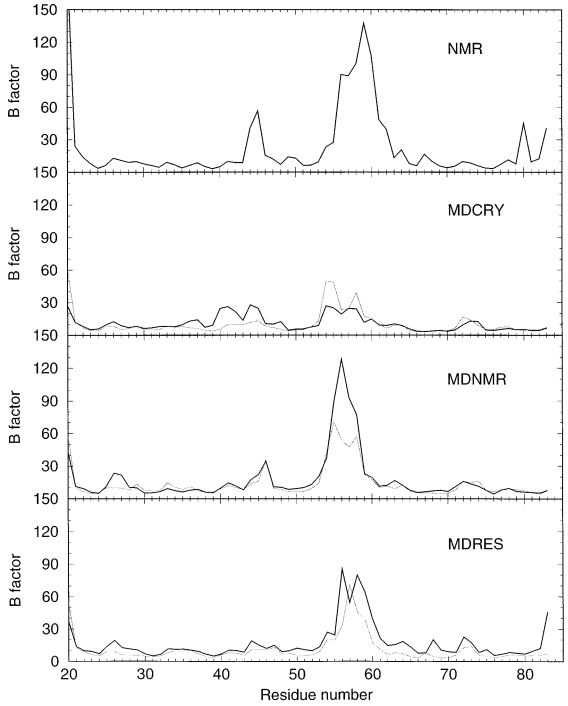

Figure 9. Isotropic C^α atomic B factors (in Å^2) calculated from the NMR ensemble and two 1 ns averages from the final 2 ns of the MDCRY, MDNMR, and MDRES simulations. Determined from the atomic fluctuations via $B_i = 8\pi^2\langle\Delta r_i^2\rangle/3$.

the hydrophobic core was not observed during the restrained water simulation. The MDRES simulation produced the best overall agreement with the distances observed in the crystal for this region. In particular, the restrained water simulation resisted any changes in accessible surface area for the loop residues (Figure 6) and maintained a reasonable Val57 to Phe69 contact distance (Table 2).

High experimental NMR order parameters (S^2) are observed for most of the NH vectors.[5] All order parameters are above 0.7, except for one residue in the reactive loop (0.6 for Ile63) and many are close to 0.9. The order parameters (1 ns averages) calculated from the simulations varied significantly as the simulation progressed, especially in the loop region, but converged to consistent values in the last few nanoseconds. The order parameters calculated from the final two nanoseconds (1 ns averages) are displayed in Figure 8. The data were consistent with the experimentally determined order parameters in that higher order parameters occurred in the helix and sheets and lower order parameters occurred in the loop and turns. There were some inconsistencies in two of the loop regions (centered on residues 46 and 56), with the simulations describing a higher degree of flexibility than that inferred from the NMR data. Calculated order parameters corresponding to the second half of the reactive loop (residues 59−63) were similar to those in the secondary structure regions and in good agreement with experiment. However, order parameters in the first half of the loop (residues 54−58) displayed significantly more motion than expected. This was even true for the restrained water simulation. Also displayed in Figure 8 are the order parameters obtained by averaging over the NMR ensemble of structures. Clearly, the NMR derived ensemble also displays more flexibility than the corresponding experimental order parameters suggest, although they agree quite well with the simulation results. Interestingly, the ensemble suggests the largest degree of flexibility for the backbone of Val57 even though the side chain of this residue is involved in a strong NOE contact with Phe69. The simulated protein isotropic rotational correlation time was determined to be 2.2 ns for the MDCRY simulation using standard techniques,[50] which is to be compared to the experimental value of 4.8 ns.[5]

The order parameters derived from the simulations displayed good agreement with experiment for the regular secondary structure regions but only qualitative agreement for the loop and turns. This appears to be a systematic problem when investigating protein dynamics, as similar differences have been reported in other simulations.[51−58] The low simulated order parameters could reflect insufficient sampling or a systematic problem with current force fields. Our data were reproducible toward the end of the simulations (see Figure 8) and so we consider our sampling to be sufficient to capture motion over the nanosecond time scale. Also, no correlation between low order parameters and amino acid type was observed, in particular for Gly (residues 29, 54 and 83) and Pro (residues 25, 44, 52 and 80), suggesting that this is not a residue specific effect. Previous studies have indicated a sensitivity of the calculated order parameters to the initial structure used in the simulation.[54] This also occurred to some extent in the current simulations with the MDNMR simulation producing lower reactive loop order parameters than MDCRY, even though their final structures were in excellent overall agreement with each other. More interestingly, the restrained water simulation, which retained the overall structure of the reactive loop, still displayed low order parameters for the loop residues, indicating that similar but different structures can display essentially the same dynamics. Furthermore, this suggests the differences between the calculated and observed order parameters may not be solely due to the different final structures involved.

Loop flexibility can also be quantified by the root-mean-square fluctuations (RMSF) in atomic positions. Figure 9 displays the C^α atomic fluctuations in terms of B factors, which can be compared with the experimental data displayed in Figure 3. The B factors support the flexibility trends exhibited by the calculated order parameters. The NMR ensemble displayed the

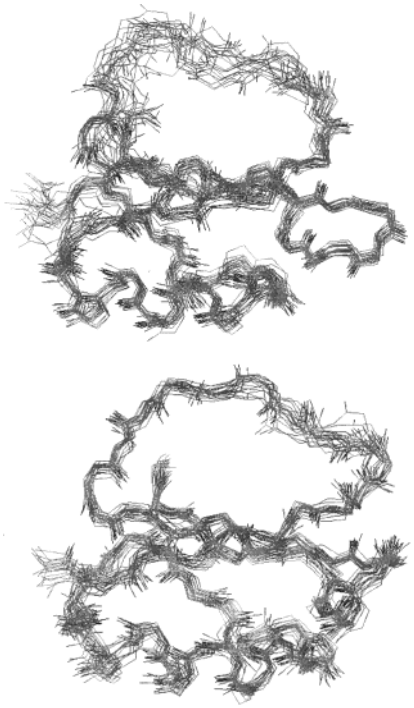

Figure 10. Superposition of the main chain atoms of the 20 NMR structures (top). Superposition of the main chain atoms of the 20 structures from the last nanosecond of the MDCRY simulation (bottom).

highest B factors, especially for the reactive loop. The reactive loop B factors varied significantly between the simulations, with the MDCRY producing the best agreement with experiment, whereas MDNMR produced the best agreement with the NMR ensemble. The restrained water simulation produced intermediate estimates of flexibility.

The flexibility of the reactive loop is thought to play an important role in the rate of hydrolysis of these inhibitors.[33] An absence of sufficient NOE data leads to a high degree of uncertainty for the NMR models in the reactive loop region, as mentioned previously. However, NMR order parameters suggest the loop is nearly as rigid as the secondary structural elements. A comparison of the NMR structures with conformations obtained from the final nanosecond of the MDCRY simulation is presented in Figure 10. The simulations provided a different view of the reactive loop motion. The motion of the loop residues close to the arginine cluster (residues 59−63) was very similar to the motion observed in other regions of the protein, in particular, the secondary structure elements. In this respect, one would expect the simulated order parameters to be similar. The residues surrounding the hydrophobic cluster (residues 54−58) displayed a higher degree of flexibility in agreement with the calculated order parameters. Clearly, the order parameter differences observed in Figure 8 reflect sensitive changes in protein backbone dynamics.

The escape of the four internal waters resulted in some disagreement between the experimental and simulation data. Several additional simulations were performed to determine if the water instability was an artifact of the simulation conditions used. The additional simulations followed the steps described in the Methods except for some small differences. The changes included (i) a larger short-range nonbonded cutoff (0.14 nm/0.14 nm), (ii) a more frequent nonbonded pair list update (every 5 steps), (iii) a longer initial water constraining time of 100 ps, (iv) slow heating of the system from 100 to 300 K over 50 ps before removing the water constraints, (v) simulating at the experimental NMR pH of 4.2 by using protonated Asp and Glu residues and 10 chloride counterions, (vi) using a more recent parametrization of the hydrocarbon force field,[59] and (vii) the use of a different water model (SPC/E).[60] None of the changes resulted in a stable arrangement for the internal water molecules, the water molecules being displaced within 250 ps after removal of the constraints. We believe that continuation of these simulations would have resulted in the same structural changes in the reactive loop as observed for the MDCRY and MDNMR simulations. As a further test, a short simulation of bovine pancreatic trypsin inhibitor (BPTI), which also contains four internal water molecules,[61,62] was performed using our simulation protocols. The internal waters were stable for this protein over a period of 1 ns, in agreement with other simulation studies.[63,64] Hence, we conclude that the exact simulation conditions did not appear to be a major problem.

As mentioned previously, a 5.3 ns simulation of CI2 has been performed starting from the same crystal structure but using a different force field (ENCAD).[65] The RMSD and RMSF data appear to be very similar to that observed here during the first half of this simulation (data not shown). The previously calculated order parameters were in quantitative agreement with the experimental values. The simulation also satisfied most of the interresidue NOEs. An interesting exception was the loss of contact between Val57 and Phe69, which also accompanied the loss of most of the internal water molecules. The same changes occurred during the current unrestrained simulations. Although it is difficult to compare the final structures, some of the side chain rearrangements observed in the present study were also observed in the previous simulation. Hence, the deviation from experiment is not limited to just the GROMOS96 force field.

Conclusions

The major structural features of CI2 were maintained during the simulations. There were very small deviations in the backbone between the final and initial structures. Secondary structural elements underwent very small deviations and fluctuations during the 10 ns. The important arginine cluster was also stable and well reproduced during the simulations. Two potentially important, and maybe correlated, aspects of the structure were poorly reproduced. The internal water molecules were unstable under the present simulation conditions. This appeared to result in a systematic variation in the intraresidue contacts on one-half of the reactive loop (residues 54−58). Correspondingly, one of the strong NOE derived contacts was lost during the current simulations. This problem was also observed in a previous simulation using a different force field and therefore we feel that, although the results clearly disagree with experiment for this region, it is important to highlight these problems if improved force fields are to be developed in the future.

The internal waters appeared to play an important role in stabilizing the loop structure. This is very interesting as the water molecules are situated away from the loop itself, and yet removal of the internal water results in movement of Phe69 and subsequent destabilization of the hydrophobic cluster, eventually leading to changes in the loop structure. The loop structure does not appear to be involved in the folding pathway of CI2, as previous studies have clearly demonstrated that the presence of internal water molecules is not necessary for generation of the correct folding motif.[19,20,25,26,31,32] The reason for the instability of the water molecules during our simulations is unknown. Of course, the force field may be inaccurate or the simulation conditions could be producing artifacts. Possible artifacts due to the simulation conditions were examined and

found to be negligible. The force field has been recently reparametrized[66] and performs well for many peptide systems.[67,68] The calculated rotational correlation time was faster than experiment, presumably indicating a weak protein−water potential,[69] and this difference could help to explain the instability of the internal water molecules. However, one would expect some improvement on using the more polar SPC/E model, which was not observed. Clearly, as the force field performs well for BPTI, this does not appear to be a general problem and is suggestive of more subtle effects controlling the stability of irregular sheet structures, which appear to collapse to more regular sheets using the current force field. Consequently, CI2 represents a very sensitive test for modern force fields and their ability to model, albeit subtle, structural features of proteins.

The restrained water simulation maintained the correct arrangement of the hydrophobic cluster supporting the reactive loop. The restrained simulation was not continued further as the calculated order parameters were still in poor agreement with experiment, the internal waters were lost after removing the restraints, and it is not clear how to interpret ensemble averages from restrained simulations. Furthermore, although the final structure was in good agreement with experiment, there still appeared to be too much backbone flexibility in the loop region. Hence, a complete understanding of the dynamics of this important region is still not possible. In addition, it is also clear that the NMR ensemble of structures display too much flexibility for the reactive loop in comparison with the NMR derived order parameters and are actually in better agreement with the simulations than the experimental data.

Acknowledgment. This project was supported by the Kansas Agricultural Experimental Station (Publication 02-42-J) and the NSF.

References and Notes

(1) Neurath, H. *Science* **1984**, *224*, 350−357.
(2) Laskowski, M. J.; Kato, I. *Annu. Rev. Biochem.* **1980**, *49*, 593−626.
(3) McPhalen, C. A.; James, M. N. *Biochemistry* **1987**, *26*, 261−269.
(4) Itzhaki, L. S.; Neira, J. L.; Fersht, A. R. *J. Mol. Biol.* **1997**, *270*, 89−98.
(5) Shaw, G. L.; Davis, B.; Keeler, J.; Fersht, A. R. *Biochemistry* **1995**, *34*, 2225−2233.
(6) Nanzer, A. P.; Poulsen, F. M.; van Gunsteren, W. F.; Torda, A. E. *Biochemistry* **1994**, *33*, 14503−14511.
(7) Melacini, G.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1999**, *15*, 189−201.
(8) Neira, J. L.; Itzhaki, L. S.; Otzen, D. E.; Davis, B.; Fersht, A. R. *J. Mol. Biol.* **1997**, *270*, 99−110.
(9) Ludvigsen, S.; Shen, H. Y.; Kjaer, M.; Madsen, J. C.; Poulsen, F. M. *J. Mol. Biol.* **1991**, *222*, 621−635.
(10) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1994**, *33*, 13880−13887.
(11) Otzen, D. E.; Rheinnecker, M.; Fersht, A. R. *Biochemistry* **1995**, *34*, 13051−13058.
(12) Otzen, D. E.; Fersht, A. R. *Protein Eng.* **1999**, *12*, 41−45.
(13) Roesler, K. R.; Rao, A. G. *Protein Eng.* **1999**, *12*, 967−973.
(14) Roesler, K. R.; Rao, A. G. *Protein Sci.* **2000**, *9*, 1642−1650.
(15) Silow, M.; Tan, Y. J.; Fersht, A. R.; Oliveberg, M. *Biochemistry* **1999**, *38*, 13006−13012.
(16) Otzen, D. E.; Fersht, A. R. *Biochemistry* **1998**, *37*, 8139−8146.
(17) de Prat, G.; Ruiz-Sanz, J.; Neira, J. L.; Itzhaki, L. S.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3683−3686.
(18) Li, A.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 10430−10434.
(19) Li, A.; Daggett, V. *Protein Eng.* **1995**, *8*, 1117−1128.
(20) Li, A.; Daggett, V. *J. Mol. Biol.* **1996**, *257*, 412−429.
(21) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.
(22) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins* **2000**, *39*, 252−260.
(23) Kurt, N.; Haliloglu, T. *Proteins* **1999**, *37*, 454−464.
(24) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8473−8478.
(25) Daggett, V.; Li, A.; Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1996**, *257*, 430−440.
(26) Pan, Y.; Daggett, V. *Biochemistry* **2001**, *40*, 2723−2731.
(27) Li, L.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13014−13018.
(28) Baysal, C.; Atilgan, A. R. *Proteins* **2001**, *45*, 62−70.
(29) Kurt, N.; Haliloglu, T. *J. Biomol. Struct. Dyn.* **2001**, *18*, 713−731.
(30) Li, L.; Shakhnovich, E. I. *J. Mol. Biol.* **2001**, *306*, 121−132.
(31) De Jong, D.; Riley, R.; Alonso, D. O.; Daggett, V. *J. Mol. Biol.* **2002**, *319*, 229−242.
(32) Fersht, A. R.; Daggett, V. *Cell* **2002**, *108*, 573−582.
(33) Cai, M.; Huang, Y.; Prakash, O.; Wen, L.; Dunkelbarger, S. P.; Huang, J. K. *Biochemistry* **1996**, *35*, 4784−4794.
(34) McPhalen, C. A.; James, M. N. *Biochemistry* **1988**, *27*, 6582−6598.
(35) Longstaff, C.; Campbell, A. F.; Fersht, A. R. *Biochemistry* **1990**, *29*, 7339−7347.
(36) Krishnamoorthi, R.; Gong, Y.; Richardson, M. *FEBS Lett.* **1990**, *273*, 163−167.
(37) Stott, K.; Blackburn, J. M.; Butler, P. J.; Perutz, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6509−6513.
(38) Bryant, J.; Green, T. R.; Gurusaddaiah, T.; Ryan, C. A. *Biochemistry* **1976**, *15*, 3418−3424.
(39) Clore, G. M.; Gronenborn, A. M.; Kjaer, M.; Poulsen, F. M. *Protein Eng.* **1987**, *1*, 305−311.
(40) Clore, G. M.; Gronenborn, A. M.; James, M. N.; Kjaer, M.; McPhalen, C. A.; Poulsen, F. M. *Protein Eng.* **1987**, *1*, 313−318.
(41) Scott, W.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596−3607.
(42) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(43) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451−5459.
(44) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 3169−3174.
(45) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992−1023.
(46) Henry, E. R.; Szabo, A. *J. Chem. Phys.* **1985**, *82*, 4753−4761.
(47) Smith, P. E.; van Schaik, R. C.; Szyperski, T.; Wuethrich, K.; Van Gunsteren, W. F. *J. Mol. Biol.* **1995**, *246*, 356−365.
(48) Chandrasekhar, I.; Clore, G. M.; Szabo, A. *J. Mol. Biol.* **1992**, *226*, 239−250.
(49) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.
(50) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(51) Eriksson, M. A.; Berglund, H.; Hard, T.; Nilsson, L. *Proteins* **1993**, *17*, 375−390.
(52) Philippopoulos, M.; Mandel, A. M.; Palmer, A. G.; Lim, C. *Proteins* **1997**, *28*, 481−493.
(53) Smith, L. J.; Mark, A. E.; Dobson, C. M.; van Gunsteren, W. F. *Biochemistry* **1995**, *34*, 10918−10931.
(54) Horita, D. A.; Zhang, W.; Smithgall, T. E.; Gmeiner, W. H.; Byrd, R. A. *Protein Sci.* **2000**, *9*, 95−103.
(55) Philippopoulos, M.; Lim, C. *J. Mol. Biol.* **1995**, *254*, 771−792.
(56) Wrabl, J. O.; Shortle, D.; Woolf, T. B. *Proteins* **2000**, *38*, 123−133.
(57) Pfeiffer, S.; Fushman, D.; Cowburn, D. *J. Am. Chem. Soc.* **2001**, *123*, 3021−3036.
(58) Dvorsky, R.; Hornak, V.; Sevcik, J.; Tyrrell, G. P.; Caves, L. S. D.; Verma, C. S. *J. Phys. Chem. B* **2002**, *106*, 6038−6048.
(59) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205−1218.
(60) Berendsen, H. J.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(61) Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr. D* **1996**, *52*, 18−31.
(62) Wlodawer, A.; Walter, J.; Huber, R.; Sjolin, L. *J. Mol. Biol.* **1984**, *180*, 329.
(63) Brunne, R. M.; Liepinsh, E.; Otting, G.; Wuethrich, K.; van Gunsteren, W. F. *J. Mol. Biol.* **1993**, *231*, 1040−1048.
(64) Schiffer, C. A.; Huber, R.; Wuethrich, K.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *241*, 588−599.
(65) Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Comput. Phys. Commun.* **1995**, *91*, 215−231.
(66) Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535−547.
(67) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.
(68) Bonvin, A. M.; Sunnerhagen, M.; Otting, G.; van Gunsteren, W. F. *J. Mol. Biol.* **1998**, *282*, 859−873.
(69) Smith, P. E.; van Gunsteren, W. F. *J. Mol. Biol.* **1994**, *236*, 629−636.