Journal of Biomolecular NMR, 10 (1997) 231–243
KLUWER/ESCOM

J-Bio NMR 464

Protein solution structure calculations in solution: Solvated molecular dynamics refinement of calbindin D_{9k}

Johan Kördel[a,*], David A. Pearlman[b] and Walter J. Chazin[c]

[a]*Pharmacia & Upjohn, N62:6, S-112 87 Stockholm, Sweden*
[b]*Vertex Pharmaceuticals Inc., Cambridge, MA 02139-4242, U.S.A.*
[c]*The Scripps Research Institute, La Jolla, CA 92037, U.S.A.*

Received 21 March 1997
Accepted 14 May 1997

Keywords: Calcium-binding proteins; Protein structure; Restrained molecular dynamics; Solvated refinement; Structure refinement

Summary

The three-dimensional solution structures of proteins determined with NMR-derived constraints are almost always calculated in vacuo. The solution structure of $(Ca^{2+})_2$-calbindin D_{9k} has been redetermined by new restrained molecular dynamics (MD) calculations that include Ca^{2+} ions and explicit solvent molecules. Four parallel sets of MD refinements were run to provide accurate comparisons of structures produced in vacuo, in vacuo with Ca^{2+} ions, and with two different protocols in a solvent bath with Ca^{2+} ions. The structural ensembles were analyzed in terms of structural definition, molecular energies, packing density, solvent-accessible surface, hydrogen bonds, and the coordination of calcium ions in the two binding loops. Refinement including Ca^{2+} ions and explicit solvent results in significant improvements in the precision and accuracy of the structure, particularly in the binding loops. These results are consistent with results previously obtained in free MD simulations of proteins in solution and show that the rMD refined NMR-derived solution structures of proteins, especially metalloproteins, can be significantly improved by these strategies.

Introduction

The three-dimensional solution structure of the small EF-hand calcium-binding protein calbindin D_{9k} has been determined in the calcium-bound state (Kördel et al., 1993; entries 2bca and 2bcb in the Brookhaven Protein Data Bank). The ensemble of structures is derived from a large number of experimental constraints (16 per residue on average) and a computational protocol consisting of distance geometry and restrained molecular dynamics (rMD) refinement in vacuo. These calculations produced well-defined high-resolution solution structures with an rmsd of 0.45 Å from the average for the N, C^{α}, and C' atoms in the helices.

Despite their high quality, these structures contain conformational perturbations attributable to the approximations required for rMD refinement in vacuo. The absence of explicit Ca^{2+} ions and water molecules, a distance-dependent dielectric function and reduced net charges of Glu, Asp and Lys side chains (from ±1.0 to ±0.2) are all expected to affect the quality of the structure in specific regions (see e.g. Guenot and Kollman, 1992), in particular in the binding loops and at the protein surface.

One of the primary effects of MD refinement in vacuo is a smaller than expected solvent-exposed surface (e.g. Levitt and Sharon, 1988; Kördel et al., 1993). For example, Norin et al. (1994) observed this phenomenon when comparing unrestrained MD calculations of a lipase in vacuo, in water and in the nonpolar solvent methyl hexanoate. Starting from a crystal structure, they observed that the solvent-exposed surface decreased by 10% after 200 ps of MD in vacuo, increased by 2% in methyl hexanoate, and increased by 14% in water. An analysis of rms fluctuations of all atoms also showed that the protein was more rigid in the in vacuo simulations (0.84 Å) than in the nonpolar solvent (0.93 Å) and in water (1.14 Å).

[*] To whom correspondence should be addressed.

Abbreviations: NOE, nuclear Overhauser enhancement; SAS, solvent-accessible surface; MD, molecular dynamics; EM, energy minimization; rEM (rMD), restrained EM (MD); Rmsd, root-mean-square deviation.

A second problem associated with MD in vacuo is the perturbation of surface side chains (e.g. Guenot and Kollman, 1992). In particular, in rMD refinements, there is a tendency for flexible side chains at the protein surface, which have little or no experimental restraints specifying their conformation, to be effectively pinned by the Lennard-Jones nonbonded attractive forces to the rest of the protein (Kördel et al., 1993). Thus, the side chain may not only occupy an inappropriate conformation, it may be better defined than the NMR data warrant. Although such problems are less likely to occur when using distance geometry or hybrid simulated annealing computational protocols, there are clear advantages to utilizing the full molecular mechanics force field.

So far, few attempts have been made to incorporate empirical potentials to better simulate the effects of solvent or to include explicit solvent in MD refinements of solution structures of proteins. Three examples stand out. Billeter et al. (1993) placed an ensemble of energy minimization (EM) refined distance geometry structures in a 5 Å shell of water molecules and subjected them to 10 ps of rMD at 300 K. Prompers et al. (1995) reported that structures obtained by the simulated annealing method were improved in quality by a subsequent step of rMD at 300 K with the structures imbedded in a 7 Å shell of water molecules. More recently, Berndt et al. (1996) placed a distance geometry structure in a water bath and refined it by 200 ps of rMD at 277 K.

The studies of Smith et al. (1995) are also of interest. These authors used agreement with NOE data to monitor an initial unrestrained MD simulation of lysozyme in water, and found that the structure was perturbed during the course of the simulation. Some of these structural perturbations only became apparent after a relatively long, 500 ps, period of simulation. The specific source of the problem was identified as misadjustment of the parameters in the GROMOS force field. After adjusting the treatment of aromatic hydrogens and modifying the water oxygen–protein carbon interactions, they found that unrestrained MD in water was able to adequately reproduce the experimental NMR data.

Metalloproteins present a particularly difficult problem for MD simulations in vacuo, due to imperfections in the force field parameterization for metal ions and the difficulties in accurately modeling the electrostatic forces so critical to their structural properties. For this reason, no Ca^{2+} ions were included during the in vacuo refinements of $(Ca^{2+})_2$-calbindin D_{9k}. The effects on the binding loops were very noticeable, and result in a much lower precision in the conformational ensemble for these two loops relative to the four helical elements of the protein (Kördel et al., 1993). In contrast, the calcium-binding loops are well defined in the crystalline state (Szebenyi and Moffat, 1986; Svensson et al., 1992). ^{15}N NMR relaxation and backbone amide exchange measurements indicated that the lack of definition in the solution structures is not due to increased mobility, but rather to deficiencies in the data and calculations (Kördel et al., 1992,1993; Skelton et al., 1992). Specific contributions to the lower precision in the binding loops include the increase in the conformational space available to side chains in the loop, due to the absence of calcium ions. The problem is exacerbated by the inaccurate representation of the electrostatic forces and the lack of explicit solvent molecules. The effect on the solution structures refined in this manner (Kördel et al., 1993) is sufficiently severe that in several structures some of the Ca^{2+} ligands actually point *away* from the binding sites and into the vacuum.

In this paper we report the results of initial efforts to improve the accuracy of the solution structures of proteins in general, and of metalloproteins such as $(Ca^{2+})_2$-calbindin D_{9k} in particular, by the inclusion of ions and explicit water molecules in the course of refinement by rMD. We conclude that solvated refinements should not originate from structures previously rMD refined in vacuo, but rather from an earlier stage in the process such as starting structures generated by distance geometry calculations. Our results disclose that significant improvements in the quality of the solution structure of calbindin D_{9k} can be obtained by this treatment.

Materials and Methods

Structure refinement

Four different series of refinements were performed: two in vacuo (VAC, VACION) and two in solvent (SHORTWAT, WAT), as detailed below. All simulations were performed using P43G $(Ca^{2+})_2$-calbindin D_{9k}, the form used for all NMR experiments. The 1002 distance and 174 dihedral angle constraints, as well as the starting structures generated with the distance geometry program DISGEO (Havel and Wüthrich, 1984), were identical to those in our previous study (Kördel et al., 1993). rEM and temperature scaled rMD simulations were performed employing the SANDER module in the AMBER 4.0 suite of programs (Pearlman et al., 1991a,b). Implementation of the restraints and simulation parameters was as described previously (Kördel et al., 1993), except where specifically indicated otherwise.

Refinements in vacuo

The minimization–annealing–minimization protocol previously reported (Kördel et al., 1993) was modified for the new set of rMD refinements in vacuo (VAC) and for the in vacuo refinements which incorporated Ca^{2+} ions and a Ca^{2+} restraining potential (VACION). These refinements were carried out with a 10 ps annealing period and a maximum temperature of 1200 K. The VAC refinement was repeated with a 12 ps annealing period and a maximum temperature of 600 K for direct comparisons with the solvated refinements. If not otherwise stated, the reported analysis of the VAC structures refers to the lower temperature simulation. The protein unit charges were scaled to ±0.2 and a distance-dependent dielectric function was used to partially compensate for the lack of explicit solvent. For the VACION calculations, 200 steps of steepest descent energy minimization had to be used at the outset to alleviate bad contacts between the ions and protein atoms. Approximately 26 min of CPU time on a Cray YMP were needed per structure for 10 ps of rMD in vacuo.

For the VACION simulations, Ca^{2+} ions were placed into the DISGEO structures at the coordinate position corresponding to the center of mass of the C^{α} atoms of the residues that coordinate the ions in the crystal structure (Szebenyi and Moffat, 1986; Svensson et al., 1992). The C^{α} atoms of these residues were utilized rather than the specific carbonyl or carboxyl ligands, because at this early stage of refinement the backbones of the DISGEO structures are more accurate than the side chains. The ligating residues are Ala^{14}, Glu^{17}, Asp^{19}, Gln^{22} and Glu^{27} for the N-terminal site and Asp^{54}, Asn^{56}, Asp^{58}, Glu^{60} and Glu^{65} for the C-terminal site (Table 1).

In the initial attempts at rMD annealing at elevated temperatures, the DISGEO structures loaded with calcium had kinetic energies high enough to allow the ions to escape from the binding sites. This was true not only for the standard annealing cycles heating to 1200 K, but also in simulations heating to 600 K. To retain the calcium ions in the binding loops, two additional restraints were added to the simulation so that each Ca^{2+} ion was constrained to a point defined by the center of mass of the C^{α} atoms of their respective ligating residues. In the crystal structures (Szebenyi and Moffat, 1986; Svensson et al., 1992), the distances between the Ca^{2+} ions and the C^{α} atoms of the ligating residues are all significantly less than 8 Å. Therefore, the bottom of the potential well constructed for the ions was given a radius of 8 Å. A quadratic potential function with a force constant of 3.2 kcal/mol (10-fold lower than for the experimental restraints) was introduced between 8–12 Å from the C^{α} mass center, and the potential function was then switched to a linear form beyond 12 Å. This Ca^{2+} restraining potential is not needed for the refinements that explicitly include solvent molecules and, consequently, was not used for the SHORTWAT or WAT simulations.

Refinements with explicit Ca^{2+} ions and solvent

A series of preliminary experiments were carried out in which the VAC structures were placed in a bath of water and then re-refined. Analysis of these structures suggested that side chains on the surface in the VAC structures were trapped in local energy minima with energy barriers too high to achieve the level of reorganization required to locate their global energy minima in solvent. Consequently, the DISGEO starting structures with Ca^{2+} ions were utilized for the SHORTWAT and WAT refinements. Each structure was immersed in a 56 × 56 × 56 Å cube of Monte Carlo water molecules. After removal of displaced water molecules, the boxes contained approximately 3900 water molecules and approximately 13 000 atoms in total. The system was initially equilibrated with 1000 steps of steepest descent EM, where only the atoms in the water molecules were allowed to change position. This was followed by another 1000 steps of EM, during which all atoms in the system were allowed to move.

A slightly modified rMD annealing protocol with a maximum target temperature of 600 K was used for the SHORTWAT simulations to accommodate the highly mobile water molecules. Initial attempts to heat the system to either 1200 or 800 K were unsuccessful, since the temperature scaling procedure used allowed water molecules to gain too high velocities at these temperatures. An alternative route would have been to reduce the time step from 1 fs to 0.5 fs, which would allow a calculation to be carried out at 1200 K but would make it twice as long. The temperature was held at 0 K for the first ps, increased to 600 K over the following 4 ps with a temperature bath coupling constant (τ) of 0.2, then lowered to 0 K over 7 ps with a τ value of 0.8 between 5 and 10 ps and of 0.05 between 10 and 12 ps. All restraints were imposed from the start, with the force constant increased linearly from 0 to 32 kcal/mol over the first 3 ps. After

TABLE 1
CALCIUM LIGANDS[a] IDENTIFIED IN THE X-RAY STRUCTURE AND IN NMR STRUCTURES[b] OF CALBINDIN D_{9k}

Ca^{2+} site	Atoms	X-ray	VACION	WAT
I	Ala^{14} O	Y	7	10
	Glu^{17} O	Y	3	6
	Glu^{17} $O^{\varepsilon 1}/O^{\varepsilon 2}$	–	9/7	4/3
	Gly^{18} O	–	2	4
	Asp^{19} O	Y	7	10
	Asp^{19} $O^{\delta 1}/O^{\delta 2}$	–	3/3	3/5
	Gln^{22} O	Y	8	10
	Glu^{26} $O^{\varepsilon 1}/O^{\varepsilon 2}$	–	1/1	–
	Glu^{27} $O^{\varepsilon 1}/O^{\varepsilon 2}$	Y/Y	7/6	7/7
	Glu^{60} $O^{\varepsilon 1}/O^{\varepsilon 2}$	–	1/3	–
	Water	1	–	1(8), 2(2)
II	Glu^{51} $O^{\varepsilon 1}/O^{\varepsilon 2}$	–	1/1	–
	Asp^{54} $O^{\delta 1}/O^{\delta 2}$	Y/–	7/10	8/9
	Asn^{56} $O^{\delta 1}$	Y	2	1
	Gly^{57} O	–	1	1
	Asp^{58} $O^{\delta 1}/O^{\delta 2}$	Y/–	8/8	3/2
	Gly^{59} O	–	2	6
	Glu^{60} O	Y	9	9
	Glu^{60} $O^{\varepsilon 1}/O^{\varepsilon 2}$	–	1	–
	Glu^{65} $O^{\varepsilon 1}/O^{\varepsilon 2}$	Y/Y	7/7	5/9
	Water	1	–	2(7), 3(2), 4(1)

[a] Defined as oxygen atoms closer than 3.5 Å to the calcium ion.

[b] The NMR structures were refined in vacuo (VACION) and in water (WAT) with calcium ions present. The numbers reported are the structures, out of the ensembles of 10, in which ligation is identified.

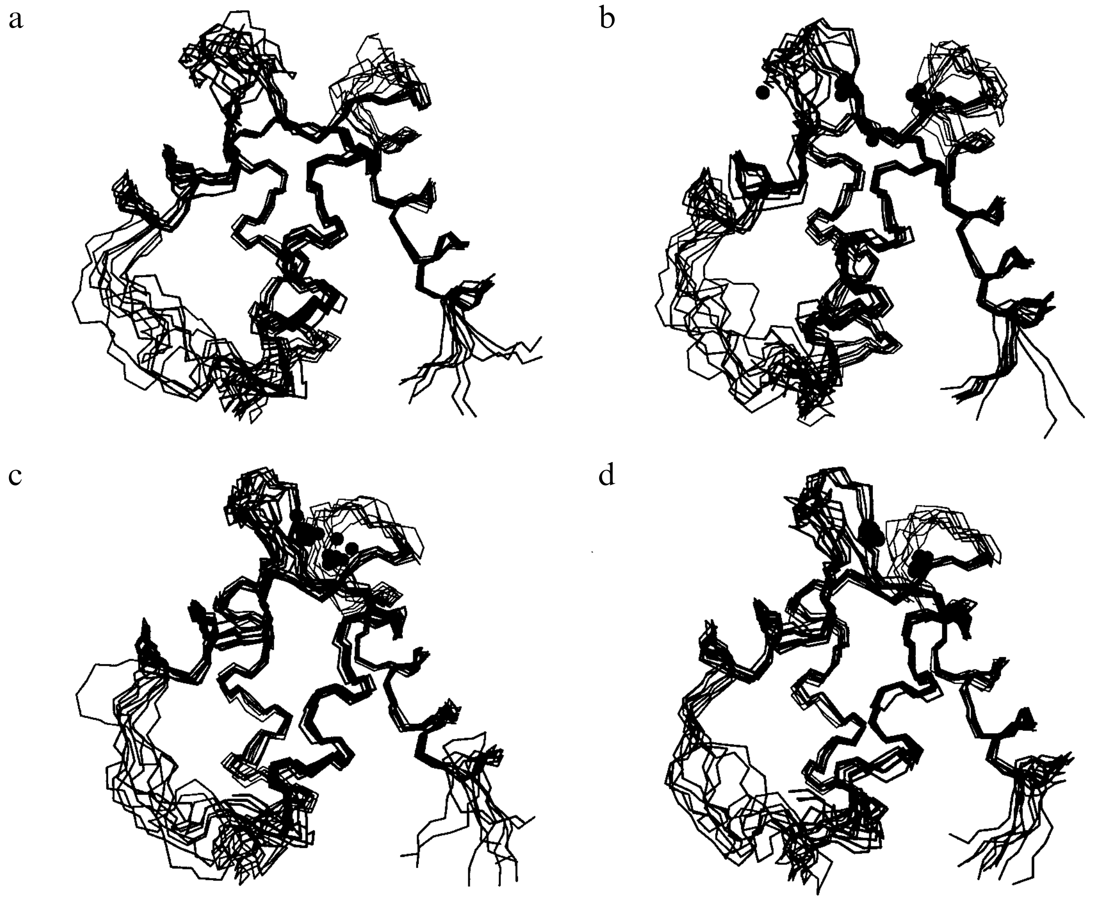

Fig. 1. Comparison of the solution structures of calbindin D_{9k} calculated in vacuo (a, VAC), in vacuo with calcium ions present (b, VACION) and in water with calcium ions with a short (c, SHORTWAT) or long (d, WAT) refinement protocol. The backbones of 10 structures are shown overlaid using the N, C^{α} and C' atoms of the four helices, residues 4–15, 25–35, 46–54 and 63–73. The figure was created with MOLSCRIPT (Kraulis, 1991).

annealing, the system was subjected to a final 1000 steps of steepest descent energy minimization, allowing all atoms to move. In all of these calculations, full partial charges were used as well as a fixed dielectric constant of 1. Each 1000 steps of minimization took 11 min of CPU time and the 12 ps of rMD annealing required 125 CPU min per structure.

For the WAT simulations, the rMD protocol was modified to include periodic boundary conditions throughout the whole refinement. The only other difference from the SHORTWAT simulation was that the period at 600 K was lengthened from 4 to 42 ps, making the total refinement cycle 50 ps rather than 12 ps long. Each 50 ps cycle of rMD annealing took 1045 min of CPU time per structure.

Volume and surface calculations

The molecular surface and solvent-accessible surface (SAS) areas for the protein structures, as well as the volumes confined by these surfaces, were calculated using the ICM 2.3 program (Abagyan et al., 1994; Biosoft, Metuechen, NJ, U.S.A.). The surface packing density was derived by dividing the molecular surface area with the solvent-accessible surface area. The residue-specific solvent-accessible surface areas were calculated using INSIGHT II (v. 95.0; MSI, San Diego, CA, U.S.A.). All SAS computations were carried out using a 1.4 Å probe and the standard algorithm (Lee and Richards, 1971; Shrake and Rupley, 1973). For the residue-specific calculations, hydrogen atoms were ignored to allow an unbiased comparison to the crystal structure. The values reported are fractions of solvent-exposed surface area. The reference values (100% SAS) were calculated as suggested by Billeter et al. (1990). In this protocol, the maximal SAS for each residue is calculated after mutating the two (or one for terminal residues) flanking residues to glycines and removing the rest of the protein, while retaining the original conformation for the Gly-X-Gly tripeptide.

Structural analysis

For each of the four classes of refinement, the 10 structures corresponding to the same 10 DISGEO starting structures were used, with one exception. In the SHORTWAT series, one structure failed to complete the refinement due to rapidly moving water molecules and one additional starting structure had to be included to make a group of 10. The DISGEO structures were randomly selected from those corresponding to the 33 final structures previously published (Kördel et al., 1993). For all comparisons with the crystal structure of calbindin D_{9k}, the *trans*-Pro^{43} form of the structure of Svensson et al. (1992; conformation A of 4icb in the Brookhaven PDB) was used. When explicit hydrogens were required, these were added using INSIGHT II (v. 95.0).

Structures were superimposed using the POLYPOSE program (Diamond, 1992). Rmsd values were calculated using a program written in-house by Dr. R.R. Ketchem. PROCHECK (v. 3.0; Laskowski et al., 1993) was used to assess the quality of the structures and to measure specific dihedral angles for each conformational ensemble, using an arbitrarily selected resolution value of 2.0 Å.

Hydrogen bonds were defined with the HBONDS program (Gippert, 1996). This program computes a hydrogen bond factor ($f = -\cos(\Theta) * (2.0/r)^2$) from the donor–donor hydrogen–acceptor angle (Θ) and the distance (r) between the donor hydrogen and acceptor. Thus, a linear hydrogen bonding arrangement with a donor–acceptor distance of 2.0 Å will have a value of 1.0. As the distance r gets longer or the angle Θ deviates from 180°, f will decrease, and if the distance is shorter than 2.0 Å, f will be greater than 1. The lower threshold for a hydrogen bonding interaction was set to $f \geq 0.5$ in one structure or $\Sigma f \geq 5.0$ for an ensemble of 10 structures.

Results and Discussion

This report describes four different refinement strategies applied to determine the 3D structure of Ca^{2+}-loaded calbindin D_{9k}. Each set of structures has been produced with an identical set of experimental constraints and the same set of starting structures generated by distance geometry calculations. The four ensembles of rMD-refined structures are referred to as: VAC, in vacuo; VACION, in vacuo and including two calcium ions; SHORTWAT, explicit waters and Ca^{2+} ions, 12 ps rMD; WAT, explicit waters and Ca^{2+} ions, 50 ps rMD. An overview of the four ensembles is shown in Fig. 1.

Effect on precision from including Ca^{2+} ions in vacuo

To determine the effect on the precision of the in vacuo refined structures due to the addition of Ca^{2+} ions, the rmsd values for the VAC and VACION structures are compared in Table 2. These data show that there are small overall decreases in both the backbone and all-atom rmsd values, despite a tendency for slight increases in the well-defined helices. It is clear that there are significant effects on the binding loops, particularly loop II (Fig. 2), which more than compensate for the slightly lower precision in the helices. The effect on the binding loops is readily explained by the exclusion of the conformational space occupied by the Ca^{2+} ions. As will be discussed below, despite these large improvements in the precision of the binding loops, the actual representation of the Ca^{2+} coordination sphere clearly remained inaccurate.

Effects of including explicit solvent molecules

Several different protocols were examined in the search for an optimal approach to refinement in the presence of explicit water molecules. Here, we briefly describe one important intermediate experiment, the SHORTWAT simulations, which involved rMD annealing over a period of 12 ps without periodic boundary conditions.

TABLE 2

RMSD VALUES TO THE MEAN STRUCTURE FOR EACH GROUP OF 10 rMD CALBINDIN D_{9k} STRUCTURES REFINED IN VACUO WITH AND WITHOUT CALCIUM IONS PRESENT AND IN WATER WITH CALCIUM IONS PRESENT

Atoms[a]	VAC	VACION	SHORTWAT	WAT
All atoms	1.64 ± 0.16	1.44 ± 0.21	1.53 ± 0.16	1.32 ± 0.13
All bckbn	1.09 ± 0.16	0.93 ± 0.19	1.01 ± 0.18	0.81 ± 0.11
Helix, all	1.11 ± 0.14	1.20 ± 0.17	1.11 ± 0.10	1.01 ± 0.10
Helix, bckbn	0.55 ± 0.11	0.59 ± 0.18	0.59 ± 0.11	0.48 ± 0.07
Helix I+II, all	0.97 ± 0.11	1.04 ± 0.09	1.06 ± 0.12	0.94 ± 0.11
Helix I+II, bckbn	0.53 ± 0.12	0.46 ± 0.15	0.57 ± 0.18	0.41 ± 0.11
Helix III+IV, all	1.20 ± 0.18	1.26 ± 0.22	1.05 ± 0.09	1.02 ± 0.16
Helix III+IV, bckbn	0.46 ± 0.15	0.53 ± 0.20	0.39 ± 0.08	0.40 ± 0.10
Loop I, all	1.10 ± 0.13	1.10 ± 0.13	1.08 ± 0.08	1.04 ± 0.09
Loop I, bckbn	0.63 ± 0.17	0.49 ± 0.14	0.44 ± 0.11	0.43 ± 0.07
Loop II, all	1.73 ± 0.27	1.27 ± 0.19	1.37 ± 0.31	1.27 ± 0.21
Loop II, bckbn	1.20 ± 0.30	0.62 ± 0.17	0.69 ± 0.17	0.72 ± 0.18
Linker, all	1.94 ± 0.58	1.70 ± 0.36	1.70 ± 0.34	1.76 ± 0.32
Linker, bckbn	1.31 ± 0.39	1.18 ± 0.35	1.16 ± 0.30	1.10 ± 0.24
Calcium ions	–	0.72 ± 0.52	0.31 ± 0.13	0.22 ± 0.12

[a] All: all atoms; bckbn: N, C^{α} and C' atoms; helix I: residues 4–15; helix II: residues 25–35; helix III: residues 46–54; helix IV: residues 63–73; loop I: residues 14–27; loop II: residues 54–65; linker: residues 36–45.

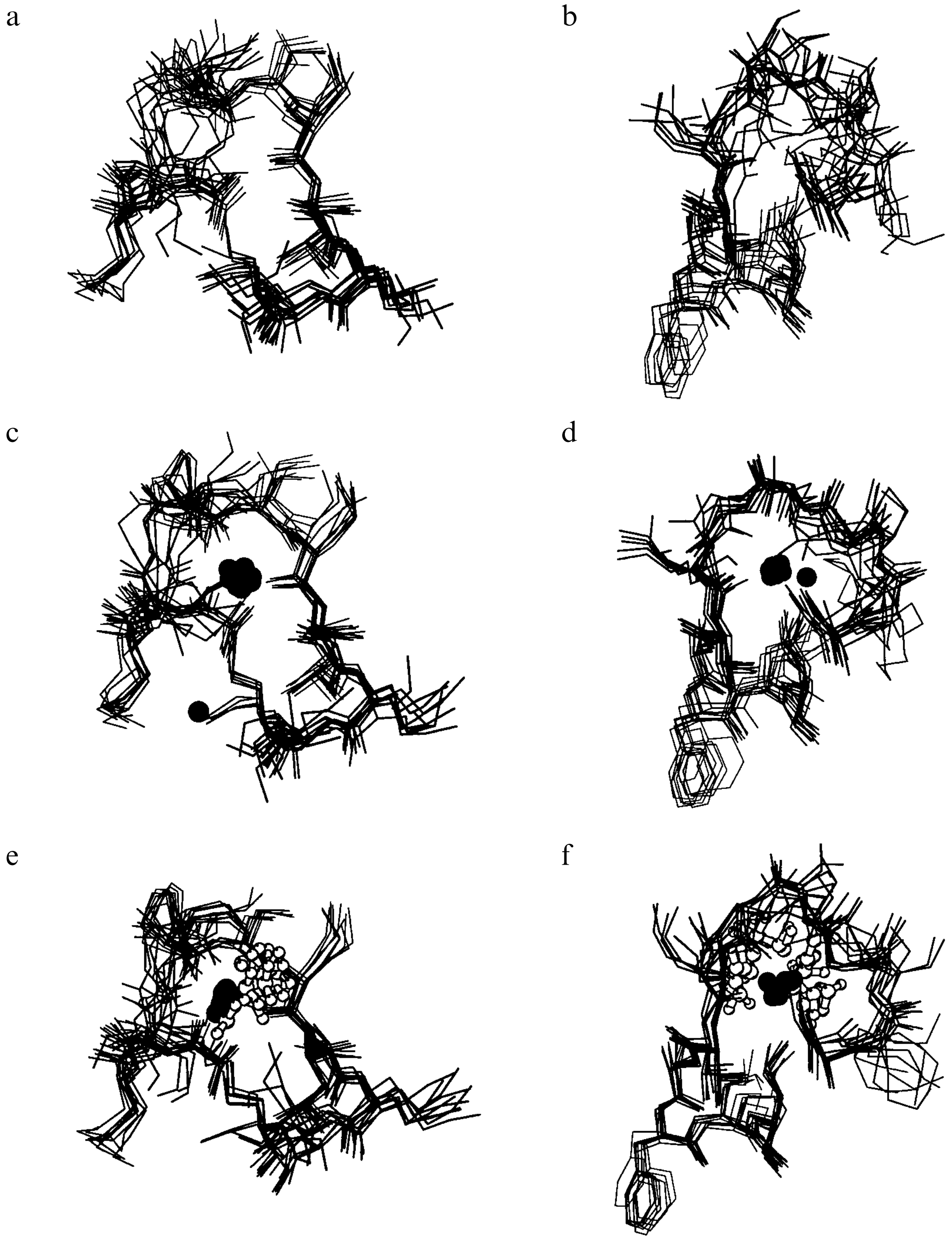

Fig. 2. Comparison of the calcium-binding loops in the solution structures of calbindin D_{9k} calculated in vacuo (a and b, VAC), in vacuo with calcium ions present (c and d, VACION) and in water with calcium ions with a long refinement protocol (e and f, WAT). The backbones of 10 structures are shown overlaid using the N, C^{α} and C' atoms of residues 14–25 (a, c, e; loop I) or 54–65 (b, d, f; loop II). Calcium ions (C–F) are shown as filled spheres and ion liganding water molecules (e, f) in a white ball-and-stick representation. The figure was created with MOLSCRIPT (Kraulis, 1991).

When the system of protein and water molecules is heated to elevated temperatures in the absence of periodic boundary conditions, the water molecules move quite substantially. In a few of the SHORTWAT structures, some water molecules have strayed as much as 1000 Å away from the protein and water cluster. The radial water density outside the protein surface was analyzed by simply binning the water molecules according to the shortest distance between the oxygen atom and any protein atom. Figure 3 clearly shows that the water density is drastically reduced after refinement at elevated temperatures in the absence of boundaries. The first hydration layer has a significant population density, but only 12% of the almost 4000 water molecules can be found inside a 16 Å thick shell. This should be compared to the WAT simulations with periodic boundaries, where 73% of the water molecules can be found inside this shell. Similarly, in the studies of the *Antennapedia* homeodomain–DNA complex, Billeter et al. (1993) observed that 25% of the 1000 surrounding water molecules were displaced more than 10 Å already at 300 K.

The average surface packing density (61.6%), total solvent-accessible surface (4552 ± 3 Å^2) and molecular volume (9829 ± 58 Å^3) of the SHORTWAT structures indicate that the protein surface properties were actually worse than in the VAC structures (60.2%, 4562 ± 124 Å^2 and 9367 ± 43 Å^3, respectively). Clearly, the reduced water density close to the protein surface in this set of refinements is highly undesirable.

Detailed analysis of the SHORTWAT structures showed that this protocol had not generated improved results. Thus, we concluded that the 12 ps refinement in the SHORTWAT protocol is too short to provide a more realistic representation of the protein. Consequently, the final protocol designed involved the application of periodic boundary conditions, and equilibration of the system at 600 K for 42 rather than 4 ps, making the total refinement cycle 50 ps long. All further discussions are based on the comparison of the VAC and WAT structures.

Structural definition and residual restraint energies

To establish a basis for the comparative analysis of the VAC and WAT structures, we compare the precision of these two ensembles, their residual restraint violations, and the various contributions to the molecular energies. As seen from Table 2 and Fig. 2, the inclusion of explicit Ca^{2+} and water molecules leads to some markedly lowered rmsd values, particularly among the side chains. The backbone rmsd values are also lower in general, except for helices I and II which refine to extremely high precision in vacuo. The increase in precision in the WAT structures is most substantial for binding loop II.

The covalent geometry of the two structural ensembles was evaluated using the program PROCHECK (Laskowski et al., 1993). While main-chain bond lengths are very similar in the VAC and WAT structures, a slight increase in distorted main-chain bond angles and group planarity is observed for the WAT structures. When including water in the refinement, the number of main-chain bond angles per structure differing by more than 10.0° from small-molecule values increases from 0.4 to 0.6. Similarly, the number of aromatic rings with distortions from planarity larger than 0.04 Å increases from 3.4 to 4.4 per structure.

Overall, the total residual restraint violations are quite small and very similar in both refinements. The largest NOE violations in the VAC and WAT structures are 0.33 ± 0.22 and 0.47 ± 0.24 Å, respectively, and the sums of all NOE violations are 3.74 ± 0.51 and 4.23 ± 0.57 Å. The total (NOE plus dihedral) residual restraint energies are only 24.0 ± 3.6 and 26.8 ± 3.9 kcal/mol.

Solvent-accessible surface, surface packing density, and solvent distribution

The molecular volumes of the crystal, VAC and WAT structures are 9514, 9367 ± 43 and 9519 ± 75 Å^3, respectively. From these values we can conclude that the exclusion of water causes the protein NMR structures to con-

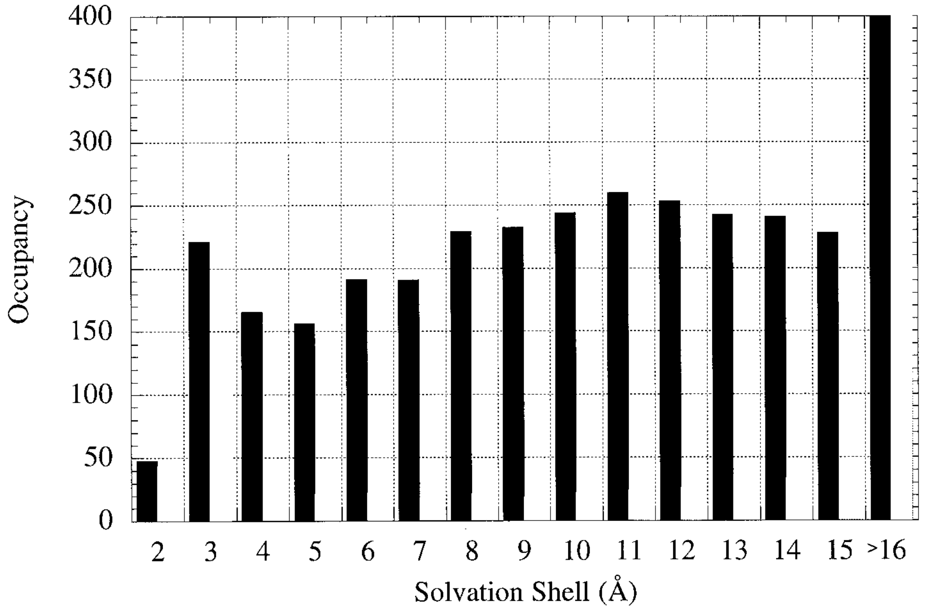

Fig. 3. Density of solvent molecules around the protein in simulations with (black bars; WAT) and without (grey bars; SHORTWAT) periodic boundary conditions after prolonged molecular dynamics simulation at 600 K. Solvation shell grouped values of the shortest distances between the oxygen atom of the water molecules and any protein atom are shown. Average values from 10 structures are shown for each type of refinement; bars for distances greater than 16 Å are truncated.

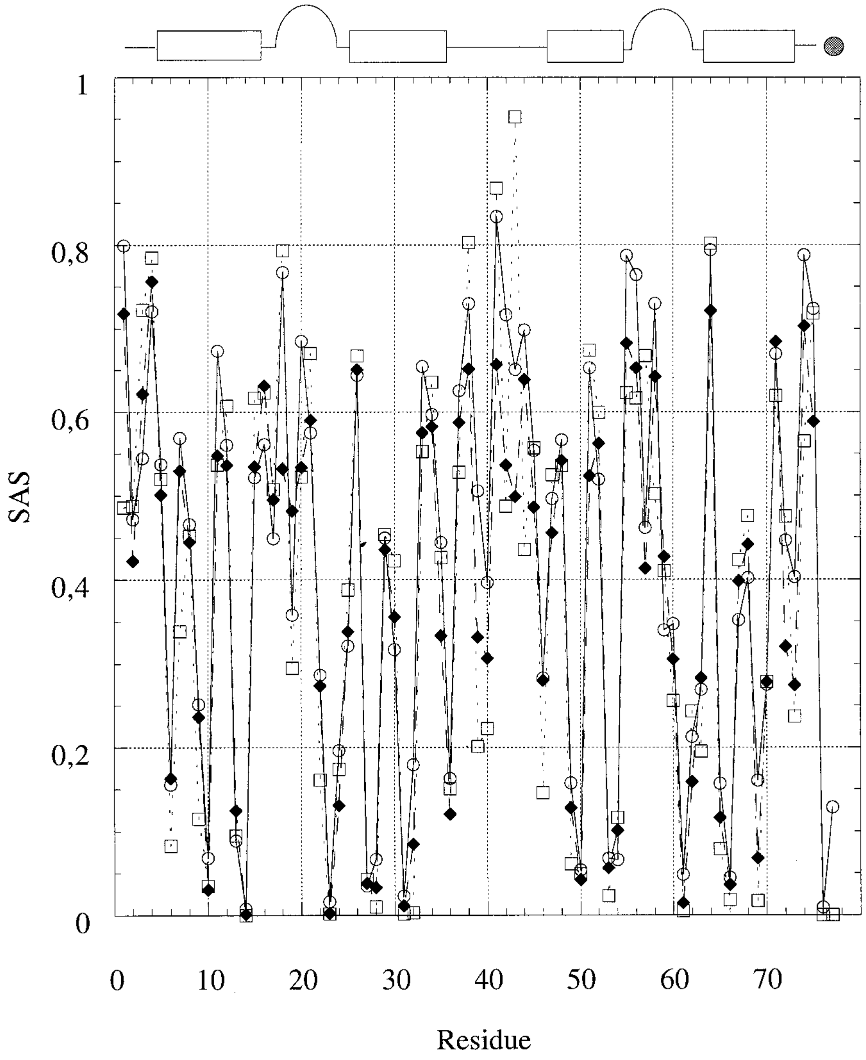

Fig. 4. Residue-specific solvent-accessible surface areas in the solution structures of calbindin D_{9k}. The curves show average values for the structures calculated in vacuo (dashed lines, filled diamonds; VAC) and in water (solid lines, open circles; WAT) and for the crystal structure (dotted lines, open boxes). The secondary structure is outlined at the top, boxes indicating helices and semicircles calcium-binding loops. Note that the values for the two calcium ions are reported as residues 76 and 77, indicated by the filled circle at the top.

tract slightly (1.6%), whereas the WAT structures are comparable to the crystal structure. Comparing the solvent-accessible areas, the differences are markedly larger with overall values of 4757, 4562 ± 124, and 4937 ± 120 Å^3, respectively. This corresponds to a 4.1% smaller SAS for the VAC structures and a 3.8% larger SAS for the WAT structures, relative to the crystal structure. The surface packing densities were computed as 59.8%, 60.2% and 58.8% for the crystal, VAC and WAT structures. The trends observed in these surface parameters closely parallel those observed in comparisons of free MD calculations with different methods for simulating solvent (e.g. Levitt and Sharon, 1988; Guenot and Kollman, 1992).

The SAS is plotted on a residue-by-residue basis in Fig. 4, and a few significant trends are distinguishable. In the four helices, where the rmsd values are lowest and SAS values are mostly small, the VAC and WAT structures tend to agree with each other and with the crystal structure. In the two Ca^{2+} binding loops where rmsd values are somewhat higher and many SAS values are high, there is a tendency towards greater disparity between the VAC and WAT structures, with WAT usually closer to the crystal structure. In the linker between the two EF hands, where rmsd and SAS values are invariably high, significant differences between VAC and WAT structures are observed across the entire polypeptide segment. Thus, we find that the inclusion of explicit water molecules improves the representation of the molecular structure most in regions of the protein that are poorly constrained. In the linker between the two EF hands, where rmsd and SAS values are high, there is a trend toward higher mean rmsd in the VAC structures. However, the structures are so poorly defined (i.e. the standard deviations are so high) that no statistically significant differences between the VAC and WAT structures can be assigned. In the presence of explicit water, the SAS values are improved because of the more correct electrostatic representation, while the rmsd values are reduced due to the higher viscosity in these simulations.

The distribution of the 57 structural waters defined in the crystal structure and the corresponding water molecules in the WAT structures (water oxygen atom less than 2.5 Å from any protein atom) shows an interesting difference. In the X-ray structure most of the structural waters are found closest to a Glu (33%) and then in decreasing order to Asp, Leu, Lys (11% each), Gly (9%), Ser (7%), Thr (5%) and Gln (4%). In the 10 WAT structures a total of 479 water molecules are classified as belonging to the inner hydration shells. As many as 64% of these are found closest to a lysine side chain, followed by Ser (9%), Gln (7%), Asn (4%), Glu (3%) and Gly (3%). These differences in the residue-specific distribution of structural waters cannot be readily reconciled, due to the differences in the manner in which these molecules are identified. The structural waters reported in the crystal structure are only those with high occupancy factors that presumably have long lifetimes. In contrast, the structural waters from the WAT ensemble are identified at the end of the rMD trajectory, and thus represent all distinct hydration sites in 10 snapshots of fully solvated protein molecules, regardless of the residence lifetime of the water molecule.

Hydrogen bonds

When comparing hydrogen bonds in the VAC and WAT structures, one noteworthy result is an increase in the total number of intramolecular protein hydrogen bonds in the water-refined structures. A similar effect has previously been observed by Prompers et al. (1995), but in their case it was much larger in magnitude. The calbindin WAT structures have two more (5%) strong backbone–side chain and seven more (37%) strong side chain–side chain hydrogen bonds than the VAC structures. Interestingly, the intramolecular hydrogen bond energies are greater in the VAC structures. This implies that the intramolecular hydrogen bonds are stronger on average in the VAC structures. Our initial expectations were that in the absence of water molecules to form intermolecular hydrogen bonds, there would be a tendency towards overcompensation and formation of extraneous intramolecular hydrogen bonds. Our results indicate that this is not the case, and in fact, the hydrogen bonding interactions have been underestimated in vacuo.

A few backbone–backbone hydrogen bonds observed in the VAC structures are not found in the WAT structures. All of the latter occur at the end of the helices (Leu^{30} HN-Glu^{26} O, Glu^{52} HN-Glu^{48} O, Gln^{67} HN-Glu^{65} O, Gln^{75} HN-Lys^{72} O) or within the first calcium-binding loop (Gln^{22} HN-Asp^{19} O).

The most remarkable difference between the crystal and solution structures of (Ca^{2+})$_2$-calbindin D$_{9k}$ is the backbone conformation and hydrogen bonding pattern in Helix IV (residues 63–73). In the crystal structure, this helix is primarily α-helical with segments of 3$_{10}$-helix at both ends. In solution, it is irregular regardless of the refinement protocol, with a mixture of i,i+3 and i,i+4 hydrogen bonds. This suggests that the NOE restraints may represent several simultaneously existing conformations, a situation which would be better represented by the use of the time-averaged restraint procedure (Torda et al., 1990).

One hydrogen bond of interest concerns the side chain of Tyr^{13}, which is hydrogen bonded to the side-chain carboxylate of Glu^{35} in the crystal structures (Szebenyi and Moffat, 1986; Svensson et al., 1992). In the previously reported solution structure, the Tyr^{13} hydroxyl was found to hydrogen bond to either the hydroxyl oxygen atom of Thr^{34} or the Glu^{35} carboxylate (Kördel et al., 1993). Multiple hydrogen bond acceptors are also observed in the VAC structures, with hydrogen bond factors of 4.5, 2.3 and 3.3 for the acceptors Thr^{34} O$^{\gamma 1}$, Glu^{35} O$^{\varepsilon 1}$ and Glu^{35} O$^{\varepsilon 2}$, respectively. However, the hydrogen bonding pattern in the WAT structures is the same as that in the crystal structure, with significant hydrogen bond factors only for Glu^{35} (4.1 and 7.3 for the two carboxylate oxygens). Based on the results from the WAT refinement, we conclude that formation of hydrogen bonds to the Thr^{34} hydroxyl is an artifact of the refinement in vacuo.

Calcium ligation

Improvement in the accuracy and precision of the solution structure in the calcium-binding loops is one of the primary objectives of this study. In the VAC structures, it is clear that the binding loop geometries are imprecise and inaccurate, with both side-chain and backbone conformations far from those observed in the crystal structures. Although there is an improvement in the VACION structures (Fig. 2), problems remain. Among these are coordination of Ca^{2+} by the carboxylate side chains of Glu^{26} and Glu^{51} (Table 1). Furthermore, in many of the VACION structures the side-chain carboxylate oxygen atoms of Glu^{17} and Asp^{19} rather than the backbone carbonyl oxygen atoms are in the most favorable positions to chelate Ca^{2+}. Calcium coordination in site II of the VACION structures is more similar to that in the crystal structure than is site I. Interestingly, the opposite is true for the WAT structures (vide infra). The most significant artifact involves the crystal structure ligand Asn^{56}, which is found to participate in calcium ligation in only a few VACION structures.

One extraordinary feature of a few of the VACION structures is coordination of the calcium ion in loop I by the side chain of Glu^{60}, while simultaneously chelating the calcium ion in loop II with the backbone carbonyl oxy-

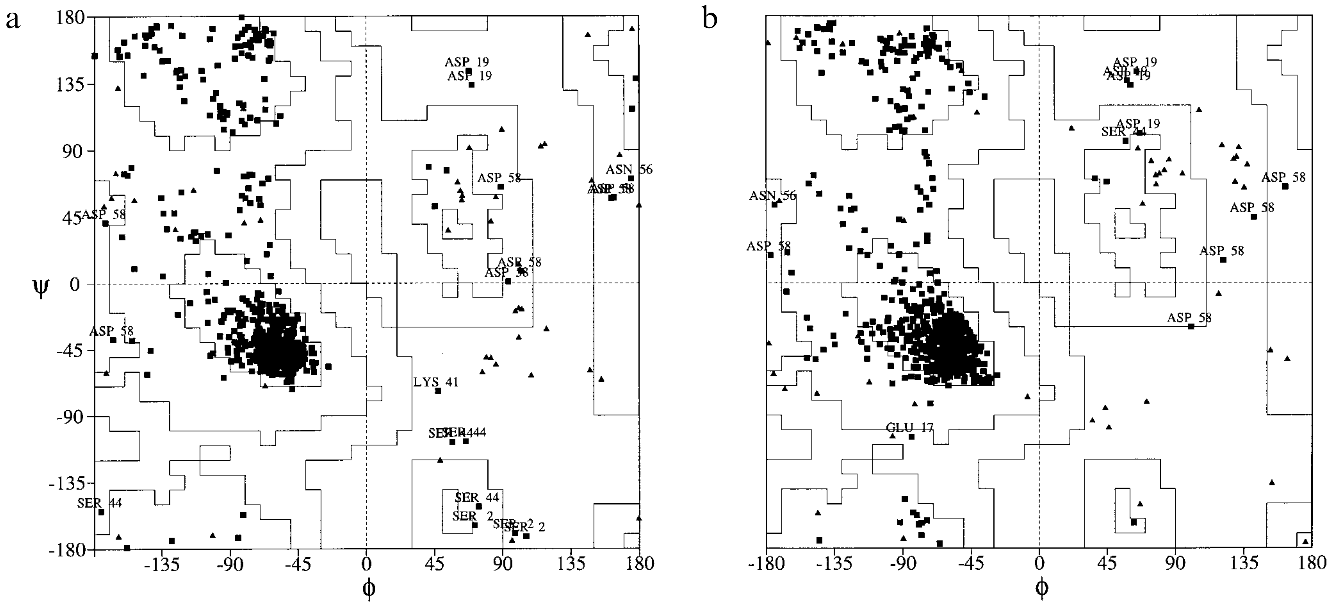

Fig. 5. Ramachandran plots for the ϕ-ψ backbone angles of calbindin D_{9k} from the solution structures refined in vacuo (a; VAC) and with explicit solvent (b; WAT). The figure was created with the structural analysis package PROCHECK (Laskowski et al., 1993).

gen. This dual coordination of the calcium ions has previously been observed in a free MD simulation of calbindin D_{9k} with explicit water (Ahlström et al., 1989). However, this phenomenon is not observed in the WAT structures. Thus, dual coordination by Glu^{60} appears to be an artifact that arises from rMD refinement in vacuo, probably associated with an imbalance in electrostatic forces.

In summary, while there is an improvement in the binding loop geometries by adding calcium ions to the refinement in vacuo, the results from the VACION refinement clearly indicate that the presence of explicit waters is necessary for obtaining a more realistic coordination geometry.

A significant improvement in the calcium coordination geometries is observed in the WAT structures, although comparison to the calcium ligands observed in crystal structures does yield a few differences. The backbone carbonyls of Ala^{14}, Asp^{19} and Gln^{22} ligate the Ca^{2+} ion in site I in all 10 WAT structures, although the Ala^{14} O-Ca^{2+} distance is long (3.03 Å) in one structure. In addition, the fourth backbone carbonyl ligand, Glu^{17}, is present in only 6 of the 10 structures. In the four structures lacking Glu^{17} backbone coordination, the ligation by Glu^{17} appears to be displaced by the side-chain carboxylate of Asp^{19}. The critical bidentate ligation by the side chain of Glu^{27} is observed in 7 of the 10 structures. In the three other structures, the Glu^{27} carboxylate is displaced by the carboxylate side chain of Glu^{17}, which in turn ligates the Ca^{2+} ion in a bidentate manner. A water molecule serves as the seventh Ca^{2+} ligand in the crystal structure and all of the WAT structures have a water ligand. Two structures have two water ligands. Overall, site I is heptacoordinated, as in the crystal structure, in only four of the WAT structures, with the other six having an octahedral geometry due to one additional protein or water ligand.

The Ca^{2+} coordination geometry in site II is more variable and also deviates more significantly from that in the crystal structure. Nonetheless, the three most critical ligands are consistently observed: the side-chain carboxylate at position 1 of the binding loop (Asp^{54}); the backbone carbonyl at position 7 (Glu^{60}); and the side-chain carboxylate at position 12 (Glu^{65}). The side-chain carboxylate coordination at positions 3 (Asn^{56}) and 5 (Asp^{58}) of the binding loop is observed in only 1 and 4 of the 10 structures, respectively. Also, ligation by Glu^{65} is monodentate, not bidentate, in 7 of the 10 structures, although the second oxygen never points away from the Ca^{2+}. In six structures, Asp^{54} ligation is bidentate rather than monodentate.

Many of the defects in the binding site seem to be caused by lack of structural restraints, resulting in poor local geometry and concerted shuffling of the protein ligands. For example, three structures have the side chains of Asn^{56} and Asp^{58} completely turned away from the Ca^{2+} ion, combined with incorrect protein and excess water ligands. Poor local geometry is also associated with Ca^{2+} coordination by the Gly^{57} (in one structure) or Gly^{59} (in six structures) backbone carbonyl. An excess of water molecule ligands in site II is observed in all of the WAT structures, with as many as four in one structure (Fig. 2), as opposed to the one observed in the crystal structure. These excess water molecules displace ligands from the protein and also result in a higher accessibility to bulk solvent than is expected for a typical binding site (Fig. 4). Overall, site II was the expected heptacoordination in 6 of the 10 WAT structures and is hexacoordinated in the four others.

The analysis of Ca^{2+} coordination geometries in the two sites shows that the site I coordination geometry is much better than that of site II. One plausible explanation for this observation could be a greater flexibility in site II. However, the similarity in ^{15}N order parameters and amide proton exchange rates determined for Ca^{2+}-loaded calbindin D_{9k} (Kördel et al., 1992; Skelton et al., 1992) implies that the motional properties of the two sites are similar. The more likely explanation for the poorer coordination geometry is the significantly lower number of experimental restraints available for site II, which has primarily side-chain ligands, versus site I, which has primarily backbone ligands. Thus, the experimental restraints available for site II appear to be insufficient to accurately specify the conformation of the binding site. This finding also implies that the force field alone is not sufficient to generate proper Ca^{2+} coordination geometries.

Dihedral angles and side-chain conformations

The backbone dihedral angles of the VAC and WAT structures are compared in Fig. 5 using Ramachandran plots. In the VAC structures, 1.1% of the residues are found in disallowed regions and 1.7% are found in the generously allowed regions. The corresponding numbers in the WAT structures are 0.9% and 0.9%, respectively. Thus, statistically, the backbone dihedral angles are improved by the inclusion of explicit water in the refinement. However, the effect is small because the VAC structures are already of high quality by this criterion.

The improvement associated with the WAT refinement is more readily apparent in the examination of specific problem areas. For example, the poorly defined residues at the N-terminus (Ser^2) and in the linker region (Lys^{41}, Ser^{44}) have more regular backbone conformations in the WAT structures. Another significant effect is a reduction in the number of cis peptide bonds for residues in poorly defined regions of the structure. There is only one instance of a single peptide bond (Gly^{43}-Ser^{44}) found in a cis conformation in a WAT structure. In contrast, six of the VAC structures exhibit cis peptides involving five different bonds (Lys^1-Ser^2, Gly^{42}-Gly^{43}(III), Gly^{43}-Ser^{44}, Gly^{57}-Asp^{58} and Gly^{59}-Glu^{60}), including one structure with two cis peptides. The sole exceptions to the overall trend are Asp^{19} and Asp^{58}, which consistently occupy strained backbone conformations in the WAT structures. This effect is directly coupled to the instances noted above, where the calcium ion is ligated (incorrectly) by the backbone carbonyl of an adjacent glycine residue (Gly^{18} or Gly^{59}), and hence could be associated with an inaccurate force field parameterization for Ca^{2+} as well as with inconsistencies in the data.

An improvement in the representation of the side chains is also observed in the WAT relative to the VAC structures. In many cases, this corresponds to a *higher* variability in more accurate conformations. The χ^1-χ^2 distribution for the VAC and WAT ensembles is shown in Fig. 6. Comparing the two structural ensembles, there are significantly fewer instances of side chains outside of the preferred χ^1-χ^2 regions in the WAT structures, and also a greater dispersion within these low-energy wells. We attribute this improvement to a reduction in the amount of pinning (Havel, 1991) that had been noted in our previous study (Kördel et al., 1993). There is also better agreement with the side-chain scalar coupling constants. For example, in the VAC structures, the side chains of Lys^{12}, Glu^{26}, Glu^{33}, Leu^{39}, Ser^{44}, Leu^{49}, Glu^{51}, Glu^{52}, Glu^{64},

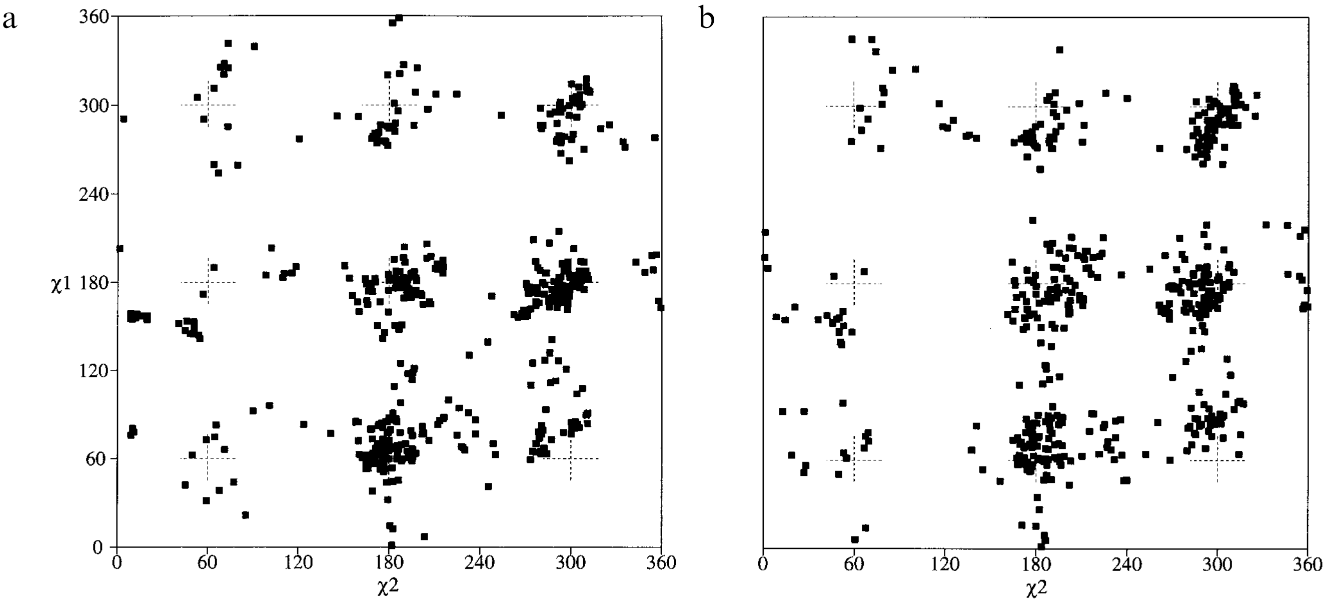

Fig. 6. χ^1-χ^2 Ramachandran plots for the NMR structures of calbindin D_{9k} refined in vacuo (a; VAC) and with explicit solvent (b; WAT). The figure was created with the structural analysis package PROCHECK (Laskowski et al., 1993).

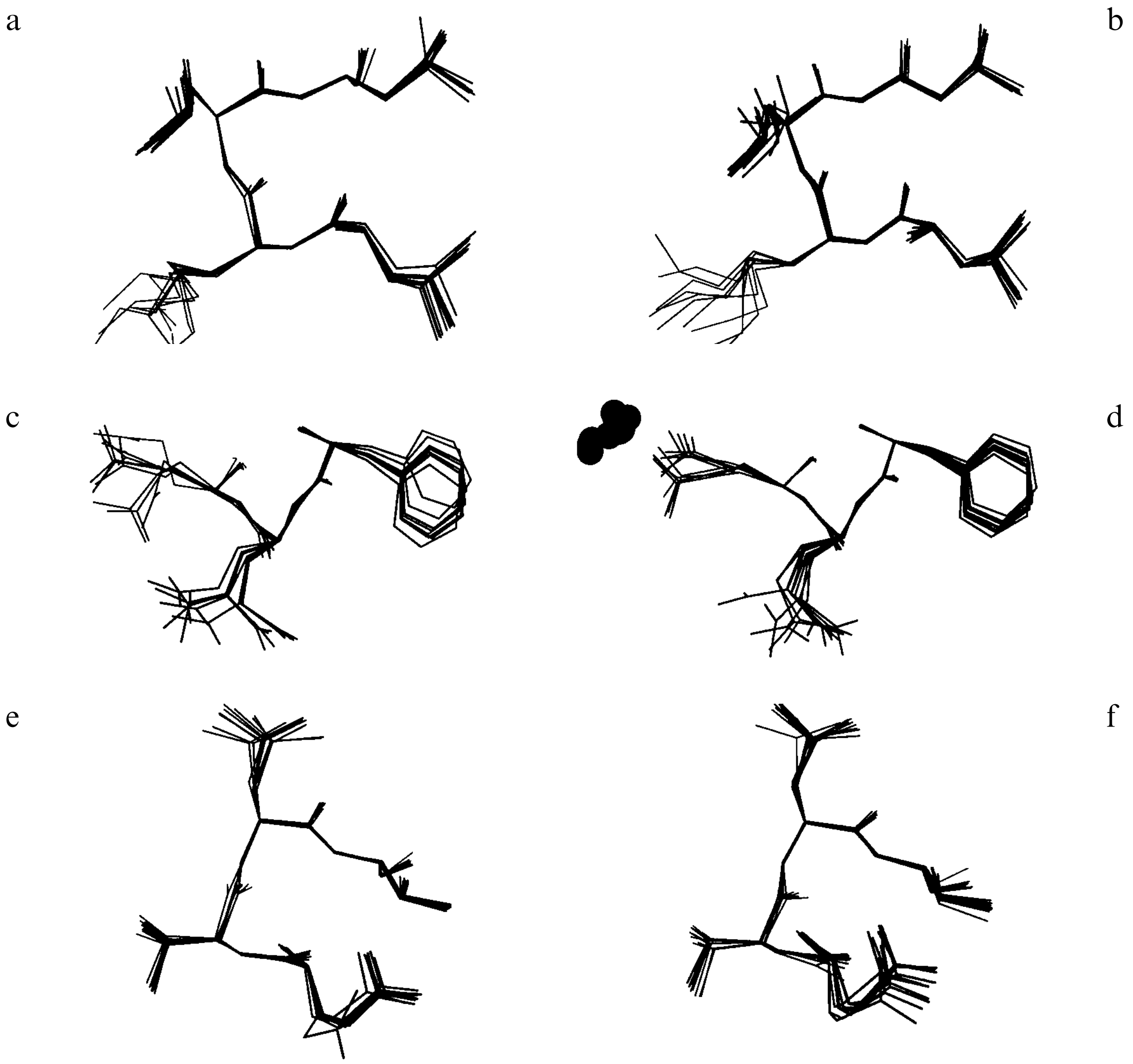

Fig. 7. Comparison of side-chain conformations from the solution structure of calbindin D_{9k} refined in vacuo (a, c, e; VAC) and with explicit solvent (b, d, f; WAT). From top to bottom, Leu^{28} to Leu^{31} (a and b), Phe^{63} to Glu^{65} (c and d) and Gln^{67} to Val^{70} (e and f) are shown. Only the heavy atoms are shown from the ensembles of 10 structures overlaying the N, C^{α} and C' atoms of the fragments in question. The figure was created using MOLSCRIPT (Kraulis, 1991).

and Ile^{73} populate a single χ^1 rotamer, even though $^3J_{\alpha\beta}$ coupling constants indicate conformational averaging. The WAT structures are found to exhibit sampling of multiple side-chain rotamers for all of these residues, except Leu^{49}. Surprisingly, the side chains of Lys^{12}, Gln^{33} and Ser^{44} exhibit a greater range of conformational sampling in the VAC structures than in the WAT structures.

Specific reasons for inaccuracies in the side-chain conformations of certain residues in the original in vacuo refined structures have been discussed previously (Kördel et al., 1993). One of these was the occurrence of well-defined χ^1 values at precisely the mid-point between two classical rotamers for Lys^{16} and Lys^{30}. The same high-energy conformers are observed in the WAT structures, indicating that this effect is specified by the data. Occupation of this distinctly disfavored conformation is caused by the pinning effect of mutually exclusive NOEs from two low-energy conformers. This situation could be resolved by the use of time-averaged restrained MD (Torda et al., 1990).

In our previous study, inaccurate representation of the solvent was used to explain certain side-chain conformations being much better defined than the experimental data warranted, as well as other instances of incorrect positioning of side chains (Kördel et al., 1993). In addition, an increased amount of packing of the aliphatic chains onto the surface of the protein was attributed to the tendency to maximize stabilization from the Lennard-Jones nonbonding attractive forces, since the stabilization of the polar ends of these side chains was significantly reduced by the absence of solvent and scaling of charges. Residues Lys^{29} and Glu^{67} were two specific examples of side-chain representations expected to improve in solvated refinements. The side chain of Lys^{29} is shown in Fig. 7, revealing the expected extension of the side chain into the solvent. However, the side chain of Glu^{67} remains relatively well packed against the protein surface, although a higher sampling of conformational space is observed (see Fig. 7).

Conclusions

Re-refinement of the NMR solution structure of $(Ca^{2+})_2$-calbindin D_{9k} has provided an improved representation of the protein, particularly in the two Ca^{2+}-binding loops. However, the combination of the available experimental data and the current computational protocol were not sufficient to provide a truly accurate representation of the structure in these critical loops. Two likely sources for this problem could be imperfections in the Ca^{2+} ion force field parameterization as well as the limited amount of experimental data available for these loops. The need for more experimental constraints is being addressed by the measurement of heteronuclear scalar coupling constants (B.T. Wimberly, J.C. Madsen and W.J. Chazin, in preparation). Preliminary computational experiments suggest that these additional input restraints will provide a significant increase in the precision and accuracy of the binding loop geometries.

Work in several laboratories on solvent models not requiring explicit solvent molecules is of great general interest for rMD refinement of proteins (C. Brooks, III, personal communication; A. McCammon, personal communication). Improved representation of solvent and electrostatics will be especially important for accurate structural refinement of metalloproteins. The results reported here on $(Ca^{2+})_2$-calbindin D_{9k} show an incremental but clearly significant improvement in the structure, and demonstrate that the quality of rMD-refined NMR-derived solution structures of proteins, especially metalloproteins, can be improved by these strategies.

Acknowledgements

We thank Dr. Garry Gippert for providing his suite of protein structure analysis programs, Dr. Robert Diamond for supplying us with his structure superpositioning program POLYPOSE, Dr. Randall R. Ketchem for assisting with various stages of the analysis of the families of NMR-refined structures, and Beth Larson for assistance with the preparation of the manuscript. We are grateful to Drs. Per Kraulis and Martin Norin for critical reading of the manuscript. This work is supported by an operating grant (GM40120) to W.J.C. from the National Institutes of Health.

References

Abagyan, R.A., Totrov, M.M. and Kuznetsov, D.N. (1994) *J. Comput. Chem.*, **15**, 488–506.

Ahlström, P., Teleman, O., Kördel, J., Forsén, S. and Jönsson, B. (1989) *Biochemistry*, **28**, 3205–3211.

Berndt, K.D., Güntert, P. and Wüthrich, K. (1996) *Proteins Struct. Funct. Genet.*, **24**, 304–313.

Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1990) *J. Mol. Biol.*, **214**, 183–197.

Billeter, M., Qian, Y.Q., Otting, G., Müller, M., Gehring, W. and Wüthrich, K. (1993) *J. Mol. Biol.*, **234**, 1084–1097.

Diamond, R. (1992) *Protein Sci.*, **1**, 1279–1287.

Gippert, G.P. (1996) Ph.D. Thesis, The Scripps Research Institute, La Jolla, CA, U.S.A.

Guenot, J. and Kollman, P.A. (1992) *Protein Sci.*, **1**, 1185–1205.

Havel, T. and Wüthrich, K. (1984) *Bull. Math. Biol.*, **46**, 674–698.

Havel, T. (1991) In *Proteins: Structure, Dynamics and Design* (Eds., Renugopalakrishnan, V., Carey, P.R., Smith, I.C.P., Huang, S.-G. and Storer, A.C.), ESCOM, Leiden, The Netherlands, pp. 110–115.

Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Kördel, J., Skelton, N.J., Akke, M., Palmer, A.G. and Chazin, W.J. (1992) *Biochemistry*, **31**, 4856–4866.

Kördel, J., Skelton, N.J., Akke, M. and Chazin, W.J. (1993) *J. Mol. Biol.*, **231**, 711–734.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Lee, B. and Richards, F.M. (1971) *J. Mol. Biol.*, **55**, 379–400.

Levitt, M. and Sharon, R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 7557–7561.

Norin, M., Haeffner, F., Hult, K. and Edholm, O. (1994) *Biophys. J.*, **67**, 548–559.

Pearlman, D.A., Case, D.A., Caldwell, J.C., Seibel, G.L., Singh, U.C., Weiner, P. and Kollman, P.A. (1991a) AMBER 4.0, University of California, San Francisco, CA, U.S.A.

Pearlman, D.A., Case, D.A. and Yip, P. (1991b) AMBER 4.0, University of California, San Francisco, CA, U.S.A.

Prompers, J.J., Folmer, F.H., Nilges, M., Folkers, P.J., Konings, R.N. and Hilbers, C.W. (1995) *Eur. J. Biochem.*, **232**, 506–514.

Shrake, A. and Rupley, J.A. (1973) *J. Mol. Biol.*, **79**, 351–371.

Skelton, N.J., Kördel, J., Akke, M. and Chazin, W.J. (1992) *J. Mol. Biol.*, **227**, 1100–1117.

Smith, L.J., Mark, A.E., Dobson, C.M. and van Gunsteren, W.F. (1995) *Biochemistry*, **34**, 10918–10931.

Svensson, A., Thulin, E. and Forsén, S. (1992) *J. Mol. Biol.*, **223**, 601–606.

Szebenyi, D.M.E. and Moffat, K. (1986) *J. Biol. Chem.*, **261**, 8761–8777.

Torda, A.E., Scheek, R.M. and van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.