A united residue force-field for calcium–protein interactions

MEY KHALILI,¹ JEFFREY A. SAUNDERS,^{1,3} ADAM LIWO,^{1,2} STANISLAW OŁDZIEJ,^{1,2}
AND HAROLD A. SCHERAGA¹

¹Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, USA

²Faculty of Chemistry, University of Gdansk, 80-952 Gdansk, Poland

(RECEIVED May 24, 2004; FINAL REVISION June 24, 2004; ACCEPTED June 28, 2004)

Abstract

United-residue potentials are derived for interactions of the calcium cation with polypeptide chains in energy-based prediction of protein structure with a united-residue (UNRES) force-field. Specific potentials were derived for the interaction of the calcium cation with the Asp, Glu, Asn, and Gln side chains and the peptide group. The analytical expressions for the interaction energies for each of these amino acids were obtained by averaging the electrostatic interaction energy, expressed by a multipole series over the dihedral angles not considered in the united-residue model, that is, the side-chain dihedral angles χ and the dihedral angles λ for the rotation of peptide groups about the $C^\alpha \cdots C^\alpha$ virtual-bond axes. For the side-chains that do not interact favorably with calcium, simple excluded-volume potentials were introduced. The parameters of the potentials were obtained from ab initio quantum mechanical calculations of model systems at the *Restricted Hartree-Fock* (RHF) level with the 6–31G(d,p) basis set. The energy surfaces of pairs consisting of Ca^{2+} -acetate, Ca^{2+} -propionate, Ca^{2+} -acetamide, Ca^{2+} -propionamide, and Ca^{2+} -N-methylacetamide systems (modeling the Ca^{2+} -Asp[−], Ca^{2+} -Glu[−], Ca^{2+} -Asn, Ca^{2+} -Gln, and Ca^{2+} -peptide group interactions) at different distances and orientations were calculated. For each pair, the *restricted free energy* (RFE) surfaces were calculated by numerical integration over the degrees of freedom lost when switching from the all-atom model to the united-residue model. Finally, the analytical expressions for each pair were fitted to the RFE surfaces. This force-field was able to distinguish the EF-hand motif from all potential binding sites in the crystal structures of bovine α -lactalbumin, whiting parvalbumin, calbindin D9K, and apo-calbindin D9K.

Keywords: protein structure prediction; united-residue force-field; calcium cation binding; EF-hand motif

Supplemental material: see www.proteinscience.org

Calcium-binding proteins play many biological roles. They are involved in regulation, membrane-related events, and muscle movement and act as biological switches (Frausto da Silva et al. 1991; Bertini et al. 1994; Lippard and Berg 1994; Hanna and Doudna 2000). Metal cations, in general, bind to centers of high hydrophilicity and reduce the enthalpy of a system upon binding (Frausto da Silva et al.

1991). All metal ions bind to a shell of polar hydrophilic residues surrounded by a shell of nonpolar residues (Yamashita et al. 1990). However, the binding process probably occurs in a step-wise manner, in which the hydrated metal ion initially positions itself in the binding site and, later, some of its inner-shell water molecules are replaced by negatively charged side-chains (Dudev and Lim 2003).

It is not clear whether metal ions help proteins to fold or whether they are incorporated into the proteins after they have already formed their three-dimensional structure. Although there have been numerous attempts to design calcium-binding sites in proteins (Yang et al. 2002, 2003), there has been very little effort to locate the most probable binding site in a protein. Locating potential binding sites in

Reprint requests to: Harold A. Scheraga, Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA; e-mail: has5@cornell.edu; fax: (607) 254-4700.

³Present address: Schrodinger, Inc., Portland, OR 97201, USA

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.04878904>.

a mean-field model is still more difficult because, in most such models, atomic details of the side-chains are eliminated and the side-chains are represented by spheres or ellipsoids. However, charges and the orientation of the side-chain are very important in the process of metal binding. Here, we present an effort to reproduce the correct charge and geometry distribution of the side-chains in a united-residue model. Such a model can help in incorporating metal ions efficiently in the folding process, especially because most metal-binding proteins are large multichain subunits, and it is currently difficult to simulate their folding by using an all-atom potential. To try to fold these proteins with a united-residue force-field that can include cations in the folding process might shed light on the roles that cations play in folding proteins. We start with the calcium cation because, in contrast to transition-metal cations, it forms complexes of predominantly electrostatic type, which are easier to model compared with covalent bonding interactions as occur with cations such as zinc. Because we treat the system at the mean-field level, we consider only the total energetic effect of replacing water molecules in the coordination sphere of a calcium cation and not the detailed kinetics of this process.

The calcium cation shows a greater tendency to form complexes with negatively charged carboxylate groups than with water molecules (Katz et al. 1996). It binds directly to the side-chains of the hydrophilic residues and to the carbonyl groups of the backbone peptide groups as opposed to indirect binding via a metal-bound water molecule (Dudev and Lim 2003; Dudev et al. 2003). The observed total coordination number of calcium in proteins varies from six to eight, with a coordination radius of 1.00 Å and 1.12 Å, respectively (Jernigan et al. 1994; Dudev and Lim 2003; Dudev et al. 2003).

Calcium binding sites can be classified into three categories (Einspahr and Bugg 1984; McPhalen et al. 1991; Pidcock and Moore 2001; Dudev and Lim 2003). Most calcium binding proteins possess a highly conserved structure, the EF-hand motif, that selectively binds calcium (Forsen and Koerdel 1996). This structure corresponds to a short continuous segment of the protein that provides groups for

binding (the classical EF-hand motif). A second structure is the same as an EF-hand motif except that one of the groups comes from a distant site. A third structure corresponds to binding sites in which the groups come from different segments in the protein. The EF-hand motif is the most common of them all (Kretsinger and Nockolds 1973). The classical EF-hand motif is a 12-residue Ca^{2+} -binding loop between two helices, forming a conserved helix-loop-helix structure (Kretsinger and Nockolds 1973; Dudev and Lim 2003). This motif contains Asp, Glu, Asn, and Gln residues that, along with the peptide groups, provide sites for calcium binding. In many calcium-binding proteins, Ca^{2+} binding induces a conformational change in the EF-hand motif, leading to the activation or inactivation of target proteins. In the S100 superfamily, for example, to which calbindin D9K, calmodulin, and troponin C belong, this conformational change is described as a change from the «closed» conformational state in the absence of Ca^{2+} to the «open» conformational state in its presence (Yap et al. 1999).

Results

The united-residue (UNRES) mean-field energy function, together with its extension to include calcium–polypeptide interactions and the parameterization of these additional terms, are described in the Materials and Methods section (equations 1, 8, 9, 10, and 13), with the parameters being summarized in Table 1 in the Materials and Methods section. In this section, we describe the results of tests of the force-field on bovine α -lactalbumin, whiting parvalbumin, and calbindin D9K.

Bovine α -lactalbumin (Protein Data Bank [PDB] access code 1F6S; Chrysina et al. 2000) is an α + β hexamer, each monomer containing 121 residues. Every monomer has one EF-hand motif starting at Asp78 and ending at Asp88. The binding pocket contains only Asp residues. The position of the calcium cation in the native structure and the position found by a systematic search of the optimum binding site and subsequent minimization of the calcium-chain energy (with the polypeptide chain fixed in the native structure; see

Table 1. Parameters^a of the Ca^{2+} force-field (equations 8 and 9) obtained by quantum mechanical energy calculations

Group	ϵ (kcal/mole)	r^0 (Å)	W_c (kcal/mole · Å)	W_{dip} (kcal/mole · Å ²)	W_{quad1} (kcal/mole · Å ³)	W_{quad2}	δ (Å)
Asp	0.1	3.5	−1381.06	283.2	−203.7	3.0	0.12
Glu	0.1	3.5	−1426.87	−98.92	−2454	3.0	0.02
Asn	0.1	3.5		−70.79	1499	3.0	0.05
Gln	0.1	3.5		−93.13	1975	3.0	0.01
Pep	0.1	6.0		109.32	3.901	3.0	0.00

^a ϵ and r^0 were estimated based on the AMBER force-field, and the rest of the parameters were obtained from RHF/6-31G (d,p) calculations. For Asn, Gln, and the peptide group, w_c does not appear because these sites possess zero net charge. For the peptide group, $\delta = 0$ because it is assumed that the calcium cation interacts with the center of mass of the peptide group.

Materials and Methods) are shown in Figure 1. The root mean square deviation (RMSD) between the force-field position of calcium and its position in the PDB structure is 0.7 Å.

Whiting parvalbumin (PDB access code 1A75; Declercq et al. 1999) is an α -helical dimeric protein. Each of the monomers has 108 residues and contains two EF-hand motifs. One of the EF-hand motifs starts at Asp53 and ends at Glu62. The other EF-hand motif starts at Asp92 and ends at Glu101. The positions of the two calcium cations in the experimental structure and the positions found by a systematic search of the optimum binding sites and subsequent energy minimization (with the polypeptide chain fixed in the native structure) are shown in Figure 2. The RMSD between the force-field positions of the calciums and those of the PDB structure is 3.37 Å in the first loop and 3.34 Å in the second loop, respectively.

Calbindin D9K (PDB access code 4ICB; Svensson et al. 1992) is an α -helical monomeric protein of 75 residues. This protein is very well studied, and there are many good crystal and NMR structures of the calcium-bound form, the apo-form, and structures with other ions such as magnesium and lanthanum in the PDB. It contains two EF-hand motifs. The first one starts at Glu17 and ends at Glu27. The second one starts at Asn54 and ends at Glu65. The second motif has more interacting residues, and therefore, it is energetically favored by calcium and other cations even though the pockets are comparable in size. The positions of the two calcium cations in the experimental structure and the positions found by systematic search of the optimum binding sites and subsequent energy minimization (with the polypeptide chain fixed in the native structure) are shown in Figure 3. The RMSD between the force-field positions of the calciums and those of the PDB structure is 5.33 Å in the first loop and 7.55 Å in the second loop, respectively.

These examples demonstrate that our calcium force-field can correctly locate the EF-hand motif as the best binding

site given the experimental geometry of the polypeptide chain. We carried out two more test calculations on calbindin D9K to determine whether the potential can locate the binding site correctly (1) in its apo-form, in which the binding sites have the closed form, and (2) in the native-like structure predicted with the UNRES force-field, in which the binding-site region is distorted because of imperfections of the force-field. In both tests 1 and 2, the geometry of the system containing the polypeptide chain and the calcium cation(s) was energy-minimized (see Materials and Methods).

To carry out test 1, we used the PDB structure of apo-calbindin D9K (PDB access code 1CLB; Skelton et al. 1995). The apo form examined is a mutant of the wild type, in which Phe43 is replaced by Gly. It appears that this mutation does not affect the potential calcium-binding regions, because the mutated residue is not at any of the two EF-hands and it is not a residue that interacts favorably with calcium. The main difference between calbindin D9K and apo-calbindin D9K is in the EF-hand regions. In apo-calbindin D9K, the EF-hands are positioned toward the center of the protein. This form of the EF-hand motif is also known as the «closed» form. However, when calciums bind to the EF-hands, they push the EF-hands to the outside and the EF-hands adopt the «open» form (Yap et al. 1999). The opening of the EF-hands upon calcium binding is thought to be responsible for the biological activity of the protein (Yap et al. 1999).

The total backbone α -carbon (C^α) RMSD between the PDB structures of calbindin D9K and apo-calbindin D9K (residues 1–75) is 2.11 Å. The C^α RMSDs between the first set of residues (17–27) and the second set of residues (54–65) in the two EF-hand motifs of the experimental structure of calbindin D9K and apo-calbindin D9K are 1.13 Å and 1.79 Å, respectively. The crystal structure of apo-calbindin D9K had to be relaxed by minimization of the UNRES energy because it had some side-chain clashes when it was

Figure 1. Stereo view of the placement of the calcium cation in the native structure (black ball) and the optimal location found by the calcium force-field (gray ball). This protein contains one calcium-binding site per monomer. The light gray segments in the figure represent the bond between the C^α and side-chain sites in the interacting residues. The backbone geometry was not optimized.

Figure 2. Stereo view of the placement of calcium cations in the native structure (black balls) of a monomer of whitting parvalbumin and the optimal locations found by the calcium force-field (gray balls). This protein contains two calcium-binding EF-hand motifs per monomer. The light gray segments in the figure represent the bond between the C^α and side-chain sites in the interacting residues. The backbone geometry was not optimized.

converted from the PDB geometry to the UNRES geometry. However, the relaxation procedure increased the RMSD between the relaxed structure of apo-calbindin D9K and the structure of calbindin D9K in the crystal. The increase in RMSD was due mainly to the introduction of three residues to the helix in the region of the first EF-hand motif of apo-calbindin D9K. After the relaxation, the overall C^α RMSD between these structures of apo-calbindin D9K and calbindin D9K was 3.70 Å, and the RMSDs between the first and the second EF-hand motifs were 2.70 Å and 1.82 Å, respectively.

After the introduction of calcium cations into the relaxed apo-structure and minimization of the UNRES energy of the chain, including the positions of the cations, the total C^α RMSD of the final structure from the structure of calbindin D9K in the crystal was 3.21 Å as opposed to 3.70 Å before energy minimization. The RMSDs between the first and the second EF-hand motifs were 2.21 Å and 1.84 Å, respectively, as opposed to 2.70 Å and 1.82 Å of the relaxed structure before calcium placement. The RMSD between the position of calcium in the final calculated structure and in the experimental structure of calbindin D9K in the crystal was 4.39 Å in the first pocket and 7.44 Å in the second pocket. Therefore, introduction of calcium cations and subsequent minimization of the energy of the complex reduced the C^α RMSD between the relaxed apo-form and the native calcium-bound form, especially in the first EF-hand motif, which was distorted by the relaxation procedure. The su-

perposition of the crystal structure of the calcium-bound form (black), the crystal structure of the apo form (red), and the final force-field structure (blue) are shown in Figure 4. It can be seen in the figure that the calcium-bound structure and the final force-field structure are much more similar in the EF-hand motif regions than are the apo structure and the final force-field structure. This shows that our calcium force-field is capable of converting the EF-hand motifs from the «closed» form in the apo structure to the «open» form in the calcium-bound structure.

To carry out test 2, we used the lowest-energy structure of apo-calbindin D9K, which was obtained by a conformational search of apo-calbindin D9K without the calcium force-field using the UNRES force-field and our conformational space annealing search (CSA) method (Lee et al. 1999). We used the latest parameterization of the UNRES force-field obtained by optimizing the free-energy landscapes of four benchmark proteins: 1E0G, 1E0L, 1GAB, and 1IGD (S. Ołdziej, J. Łagiewka, A. Liwo, C. Czaplewski, M. Chinchio, M. Nanas, and H.A. Scheraga, in prep.).

The overall C^α RMSD between this lowest-energy structure and the crystal structure of apo-calbindin D9K was 6.3 Å. The RMSDs between the first and the second EF-hand motifs were 4.17 Å and 3.31 Å, respectively. In this structure, the EF-hand motifs were completely distorted. They did not possess any extended structure, which is the characteristic of the EF-hand motif. Rather, they possessed a significant amount of helical content.

Figure 3. Stereo view of the placement of calcium cations in the native structure (black balls) of calbindin D9K along with the optimal locations found by the calcium force-field (gray balls). The light gray segments in the figure represent the bond between the C^α and side-chain sites in the interacting residues. The backbone geometry was not optimized.

Figure 4. Stereo view of the superposition of calbindin D9K (black), apo-calbindin D9K (red), and the apo-calbindin D9K structure (blue) after the calcium placements and minimization of the backbone and calcium positions with the UNRES-calcium force-field. The backbone geometry of apo-calbindin D9K was optimized before addition of calciums. Although the loops are «open» in the native structure of the apo-calbindin D9K, they become «closed» after the introduction of calcium and represent the geometry of the binding sites in the native structure of calbindin D9K.

To assess whether one calcium cation is capable of correcting one of the distorted EF-hand motifs, one calcium cation was introduced to this structure. The chain, as well as the position of the calcium cation, was energy-minimized. The final structure obtained from the calcium force-field

had an overall C^α RMSD of 5.89 Å from the crystal structure of calbindin D9K. The RMSDs of the first and second EF-hand motif were 3.91 Å and 3.65 Å, respectively. This structure is shown in Figure 5A. It can be seen that although both the overall RMSD, as well as those of the EF-hand motifs, are lower in the final structure than in the initial structure, the calcium cation was not able to correct the distorted EF-hands. It did not enter into any of the EF-hand motifs but stayed in a region between the two distorted regions in the core of the structure.

A closer examination of this preferential site for the calcium cation in Figure 5A shows that it is rich in interacting residues from both of the EF-hand motifs. In an undisturbed EF-hand, most of the interacting residues point toward the center of the EF-hand motif. But here, because of the helical pattern of the distorted EF-hands, the interacting residues alternately point inside and outside of the center of the EF-hands. As a result, some of the interacting residues from both of the EF-hands provide an optimal site between the two EF-hands for the calcium cation to bind.

Initially, we thought that only one calcium cation by itself might not be able to correct the distorted EF-hands. How-

Figure 5. (A) Stereo view of one calcium cation in the lowest energy structure obtained by folding apo-calbindin D9K with CSA, using the force-field of equation 1 and minimizing the total energy of the complex. The binding sites have significant helical content. Because of the distortion of the binding sites, the calcium cation is placed in the center of the protein so that it is close to as many interacting residues as possible. The calcium cation in this placement can interact with acidic/amide containing residues from both binding sites and from the helices. (B) Stereo view of two calcium cations in the structure of A after minimization of the total energy of the complex. The second calcium also ends up in the middle of protein. Further, helices on both sides of the proteins have shifted so that many of the interacting residues are close to the calcium cations. This shows that it is more energetically favorable for this force-field to shift helices than to unwind them in the distorted binding sites.

ever, after introducing the second calcium cation, the total C^α RMSD of the final structure increased to 6.19 Å, and the RMSDs between the first and the second EF-hands became 3.95 Å and 4.45 Å, respectively. Compared with the original structure, this structure is farther from the crystal structure of calbindin D9K. This structure is shown in Figure 5B. The two calcium cations are in the core of the protein, and they are surrounded by the interacting residues of the EF-hands and the two adjacent helices. They interact with Glu17 and Glu27 of the first EF-hand motif and Glu48, Asp54, Asp58, Glu60, and Asn67 of the second EF-hand motif. Further, the helices on the sides of the EF-hands have shifted so that they provide as many interacting residues to the two calcium cations as possible. This shift of the helices is the reason for the increase in the RMSD and the deviation between this structure and the crystal structure of calbindin D9K.

Discussion

The results obtained in this study demonstrate that our calcium force-field is capable of locating the EF-hand motifs in the calcium-bound and apo forms of proteins. In other words, when the binding site has the correct or nearly correct geometry, the force-field locates the EF-hand motif as the most energetically favorable place for the calcium cation to bind after it has explored all the possible sites that interact with the calcium cation. This force-field can, therefore, be implemented to locate the EF-hand motif when other methods are used to predict the fold. Effort is under way in our laboratory to incorporate the calcium cation in simulated folding.

However, the force-field is not able to correct the topology of a distorted binding site, as can be seen in Figure 5, A and B. The most probable reason for this is that the UNRES force-field is not perfect in reproducing loop geometry; it tends to assign some secondary structure to loops. As can be seen in Figure 5B, the helices from both sides of the loops are arranged to surround the calcium cations by the greatest number of interacting residues. This means that it is energetically more favorable for this force-field to reposition an entire helix than to open a helix turn. However, the calcium-binding constant to α -lactalbumin is of the order of 10^7 , so that the free energy of binding is ~ 10 kcal/mole (Permyakov et al. 2001); this is enough to unfold two or three turns of a helix.

The possible reasons for the distortion of the geometry of the calcium coordination sphere (i.e., the location of the calcium with respect to the interacting residues in the binding site) obtained with the force-field derived in this work are: (1) The parameterization was not refined by reproducing the geometry of high-resolution crystal structures of calcium-bound peptides; and (2) the multibody terms to express the calcium-interaction in the potential-energy func-

tion were absent. Both issues are now being investigated in our laboratory.

Materials and methods

The UNRES energy function

In the UNRES model (Liwo et al. 1997a,b, 1998, 1999a,b, 2001, 2002, 2004; Lee et al. 1999), a polypeptide chain is represented as a sequence of α -carbon atoms, (C^α s), linked by virtual bonds with attached united side-chains (SCs), and united peptide groups (ps). Each united peptide group is located in the middle of two consecutive α -carbon (C^α) atoms (Fig. 6).

The interaction sites are the united peptide groups and the united side-chains. The α -carbon atoms are geometric points. All virtual-bond lengths (i.e., $C^\alpha \cdots C^\alpha$ and $C^\alpha \cdots SC$) are fixed; the distance between the neighboring C^α s is 3.8 Å, corresponding to *trans* peptide groups, whereas the side-chain angles (α_{SC} and β_{SC}), the virtual-bond angles (θ s), and dihedral angles (λ s) can vary. In this work, we introduce calcium cations. For side chains that bind the calcium cation specifically (Asp, Glu, Asn, and Gln), the calcium cation interacts with the carboxyl and carbonyl groups in the side-chain and not with the center of mass of the side-chain. However,

Figure 6. The UNRES model of the polypeptide chain with a calcium cation near an interacting residue. Filled circles represent united peptide groups (p); open circles represent the C^α atoms, which serve as geometric points; and ellipsoids represent side-chains. The variables to change the conformation of the polypeptide chain are the virtual-bond angles θ , the virtual-bond dihedral angles γ , and the angles α_{SC} and β_{SC} that define the location of a side-chain with respect to the backbone. For a side-chain with the center of mass at SC_i^c , which is a calcium-binding site (Asp, Glu, Asn, or Gln), the calcium-interaction site, SC_i , is shifted by a distance δ_{SC_i} along the respective $C^\alpha \cdots SC$ axis, whereas for the remaining side-chains as well as for the peptide groups, the interaction site coincides with the center of mass, SC_i^c . The distance $d_{SC_i'}$ of a calcium cation from SC_i is calculated from its distance $d_{Ca^{2+}}$ from C_i^α , the angle $\alpha_{Ca^{2+}}$ between the $C_i^\alpha \cdots SC_i$ axis and the $C_i^\alpha \cdots Ca^{2+}$ vector. $\zeta_{Ca^{2+}}$ is the angle of rotation of Ca^{2+} around the $C_i^\alpha \cdots SC_i$ axis.

UNRES ellipsoids do not contain the details of the side-chains. Therefore, the calcium-interaction site is assumed to lie on the $C^\alpha \cdots SC$ axis and to be shifted from the SC center of mass by the distance δ_{SC} . This new center is called SC' . Therefore, the location of a calcium cation with respect to a favorable amino acid residue i is described by its distance $d_{SC'i}$ from SC' and the $Ca^{2+} \cdots C_i^\alpha \cdots SC_i$ angle (α_{Ca}^{2+}) as shown in Figure 6. The Ca^{2+} cation can rotate about the $C^\alpha \cdots SC$ axis by the angle ζ_{Ca}^{2+} , which, together with $d_{SC'i}$ and α_{Ca}^{2+} , defines its location with respect to the polypeptide chain. It should be noted that the energy of interaction of the SC site with the calcium does not depend on ξ_{Ca}^{2+} , because the corresponding expression has cylindrical symmetry with respect to the $C^\alpha \cdots SC$ axis. For the remaining side-chain types, calcium is assumed to interact with the center of mass of the ellipsoid. For the peptide groups of the backbone, calcium is assumed to interact with the center of mass of the peptide.

UNRES is a physics-based force-field, which is derived as a *Restricted Free Energy* (RFE) function of a polypeptide chain. The RFE is defined as the free energy of a given coarse-grain conformation obtained by integrating the Boltzmann factor of the all-atom (i.e., the polypeptide chain-plus-solvent) energy over the degrees of freedom that are neglected in the united-residue model (Liwo et al. 1998, 1999b, 2001). The complete UNRES potential-energy function, which also includes the protein-calcium cation interaction terms that are introduced in this work, is expressed by equation 1:

$$\begin{aligned}
 U = & \sum_{i < j} \sum_j U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} \sum_j U_{SC_i p_j} \\
 & + w_{el} \sum_{i < j-1} \sum_j U_{p_i p_j} + w_{tor} \sum_i U_{tor}(\gamma_i) \\
 & + w_{tord} \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) \\
 & + w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + \sum_{m=2}^{N_{corr}} w_{corr}^{(m)} U_{corr}^{(m)} \\
 & + w_{SCCa^{2+}} \sum_i \sum_j U_{SC_i Ca_j^{2+}} + w_{AGCa^{2+}} \sum_i \sum_j U_{AG_i Ca_j^{2+}} \\
 & + w_{AGNCa^{2+}} \sum_i \sum_j U_{AGN_i Ca_j^{2+}} + w_{pCa^{2+}} \sum_i \sum_j U_{p_i Ca_j^{2+}} \\
 & + w_{Ca^{2+} Ca^{2+}} \sum_{i < j} \sum_j U_{Ca_i^{2+} Ca_j^{2+}} \quad (1)
 \end{aligned}$$

The terms $U_{SC_i SC_j}$ correspond to the mean free energy of hydrophobic (hydrophilic) interactions between the side-chains. These terms implicitly contain the contributions from the interactions of the side-chain with the solvent. The terms $U_{SC_i p_j}$ correspond to the excluded-volume potential of the side chain–peptide group interactions. The terms $U_{p_i p_j}$ represent the energy of average electrostatic interactions between backbone peptide groups. The terms U_{tor} and U_{tord} are the torsional and the double-torsional potentials, respectively, for rotation about a given virtual bond or two consecutive virtual bonds. The terms U_b and U_{rot} are the virtual-angle–bending and side chain–rotamer potentials. The terms $U_{corr}^{(m)}$ correspond to the correlations (of order m) between peptide-group electrostatic and backbone-local interactions. The last five terms pertain to the calcium–polypeptide chain interactions, which are described in the next section.

The terms $U_{SC_i SC_j}$, U_b , and U_{rot} were parameterized (Liwo et al. 1997a,b) from the distribution and correlation functions determined from the PDB. U_{tor} , U_{tord} , and U_{corr} were based on the cumulant expansion of the RFE of polypeptide chains (Liwo et al. 1998, 1999b, 2001) and parameterized from the RFEs of model systems obtained by high-level MP2/6–31G(d,p) ab initio calcu-

lations of model system (Ołdziej et al. 2003; Liwo et al. 2004). Finally, the w s are weights of the various energy terms, obtained by optimization of the total potential-energy function to obtain a funnel-like energy landscape of benchmark proteins (Liwo et al. 2002).

Analytical expressions for energy of interaction of calcium cations with polypeptide chains

Binding of the calcium cation to the carboxyl and carbonyl groups is mainly electrostatic in nature. Therefore, to derive analytical expressions for the RFE of interaction of the calcium cation with the Asp, Glu, Asn, and Gln side-chains and with the peptide groups, we assumed that the charge distribution of each of these interaction sites (and, consequently, the interaction energy with the calcium cation) is expressed by a multipole series, which provided simplified analytical expressions for the interaction energy. Subsequently, by averaging these expressions over the secondary degrees of freedom (i.e., the rotation angle ζ of the Ca^{2+} cation about the C^α –SC axis shown in Fig. 6), we derived approximate analytical expressions for the mean-field interaction energies of the calcium cation with its interaction sites.

The Asp and Glu side-chains bear a net charge of $-1e$ and also possess a quadrupole moment, which is necessary to describe the angular dependence of the calcium-side-chain interaction energy. For charged systems, choosing the origin of the reference system in the charge center implies that the dipole moment is zero. However, because it is not known a priori what choice of the origin of the reference system provides the analytical expression best fitting the numerically calculated RFE surfaces, even for charged side-chains, we introduced a dipole moment parallel to the rotation axis. The energy of interaction of the calcium cation with the point charge and with the dipole moment does not depend on the rotation angle ζ ; therefore, it is necessary to average only the calcium–quadrupole interaction energy (Fig. 7). In this figure, for clarity we have shown only the quadrupole moment placed at SC' . However, in our derivations, there is also a dipole moment along the Z-axis and a point charge located at SC' .

For an ionized carboxyl group, we assume that a linear quadrupole composed of a charge $2q$ lies on the axis of rotation at the position SC' and two charges of $-q$ are positioned symmetrically with respect to the axis of rotation and separated by a distance d from the central charge of $2q$, as shown in Figure 7.

The magnitude of the quadrupole moment is given by equation 2:

$$Q = 2qd^2 \quad (2)$$

The energy of interaction of the quadrupole placed at SC' with a point charge of magnitude q' positioned at coordinates x , y , and z is given by equation 3

$$E = q'Q \frac{d_{SC'}^2 - 3x^2}{\epsilon d_{SC'}^5} \quad (3)$$

where $d_{SC'}^2 = x^2 + y^2 + z^2$ and is the dielectric constant. Transforming equation 3 from Cartesian coordinates to internal coordinates $d_{SC'}$, α , ζ such that

Figure 7. Illustration of the linear quadrupole representing the asymmetry of the charge distribution of the carboxylate group and of the definition of its local coordinate system. The site SC' corresponds to that of Figure 6. q is any point charge (here +2 corresponding to the charge of Ca^{2+}). $d_{SC'}$ is the distance between the point charge q and the center of interaction SC' , and α corresponds to the angle that $d_{SC'}$ makes with the rotation axis. ζ is the angle of rotation.

$$x = d_{SC'} \sin \alpha \cos \zeta \quad y = d_{SC'} \sin \alpha \sin \zeta \quad z = d_{SC'} \cos \zeta \quad (4)$$

we obtain equation 5.

$$E = q'Q \frac{1 - 3\sin^2 \alpha \cos^2 \zeta}{\epsilon d_{SC'}^3} \quad (5)$$

The energy averaged over the rotation angle ζ is expressed by equation 6:

$$\langle E \rangle = q'Q \frac{1 - (3/2)\sin^2 \alpha}{\epsilon d_{SC'}^3} \quad (6)$$

The energy of interaction of a point dipole (with dipole moment p) with a point charge q' (positioned at a distance $d_{SC'}$ from the dipole and angle α with respect to the dipole axis) is expressed by equation 7:

$$E = \frac{q'p \cos \alpha}{\epsilon d_{SC'}^2} \quad (7)$$

To represent the point charge energies, we use the expression for the Coulombic charge–charge interaction energy and finally a Lennard-Jones term to prevent the collapse of the calcium cation on the respective side-chain. Because the distance δ_{SCi} is much smaller than $d_{SC'i}$ and $d_{Ca^{2+}}$ in Figure 6, we assume that α in Figure 7 is approximately equal to $\alpha_{Ca^{2+}}$ in Figure 6 for simplicity. Taking all this into account, we obtain equation 8 for the energy of interaction of a calcium cation with an aspartate or glutamate side-chain, $U_{AGCa^{2+}}$.

$$U_{AGCa^{2+}}(d_{SC'}, \alpha_{Ca^{2+}}) = \epsilon_{XCa^{2+}} \left[\left(\frac{r_{XCa^{2+}}^0}{d_{SC'}} \right)^{12} - 2 \left(\frac{r_{XCa^{2+}}^0}{d_{SC'}} \right)^6 \right] + \frac{w_{c,X}}{d_{SC'}} + w_{dip,X} \frac{\cos \alpha_{Ca^{2+}}}{d_{SC'}^2} + \frac{\sin^2 \alpha_{Ca^{2+}} - w_{quad2,X}}{d_{SC'}^3} \quad (8)$$

where X denotes the Asp or the Glu side-chain, $\epsilon_{XCa^{2+}}$ and $r_{XCa^{2+}}^0$ are the parameters of the Lennard-Jones potential, $w_{c,X}$ is the parameter of the Coulomb term, $w_{dip,X}$ is the parameter of the dipole term, and $w_{quad1,X}$ and $w_{quad2,X}$ are the parameters of the quadrupole term, respectively.

The expressions for the interaction energy between the Asn and Gln side-chains and calcium ($U_{AGNCa^{2+}}$) or the peptide groups and calcium ($U_{pCa^{2+}}$) are similar. Because these groups are neutral, there is no charge–charge interaction term:

$$U_{AGNCa^{2+}}(d_{SC'}, \alpha_{Ca^{2+}}) = U_{pCa^{2+}}(d_{SC'}, \alpha_{Ca^{2+}}) = \epsilon_{X,Ca^{2+}} \left[\left(\frac{r_{XCa^{2+}}^0}{d_{SC'}} \right)^{12} - \left(\frac{r_{XCa^{2+}}^0}{d_{SC'}} \right)^6 \right] + w_{dip,X} \frac{\cos \alpha_{Ca^{2+}}}{d_{SC'}^2} + \frac{\sin^2 \alpha_{Ca^{2+}} - w_{quad2,X}}{d_{SC'}^3} \quad (9)$$

where X denotes the Asn or the Gln side-chain or the peptide group, and the meaning of the symbols is the same as described by the text under equation 8.

The constants of the Lennard-Jones potential ϵ and r_0 were estimated from the calcium cation parameters of the AMBER force-field (Pearlman et al. 1995). Determination of other parameters of equations 8 and 9 is described in “Parameterization of the Expressions for Calcium–Polypeptide Interaction Energy.”

The terms $U_{SCiCa^{2+}}$ correspond to the interaction between the calcium cation and any side-chains other than Asp, Glu, Asn, and Gln introduced to avoid clashes of the calcium cation with the side-chains. They are expressed by a Lennard-Jones potential, as given by equation 10.

$$U_{SCiCa^{2+}}(d_{SC'}) = \epsilon_{SCiCa^{2+}} \left[\left(\frac{r_{SCiCa^{2+}}^0}{d_{SC'}} \right)^{12} - 2 \left(\frac{r_{SCiCa^{2+}}^0}{d_{SC'}} \right)^6 \right] \quad (10)$$

where

$$r_{SCiCa^{2+}}^0 = \frac{1}{2} \left(\frac{r_{Ca^{2+}}^0}{2} + r_{SCi}^0 \right) \quad (11)$$

$$r_{Ca^{2+}}^0 = \frac{1}{26} \sigma_{SCi} \quad (12)$$

where $r_{Ca^{2+}}^0 = 3.4725 \text{ \AA}$ (Mayo et al. 1990), the values of σ_{SCi} were obtained from (Liwo et al. 1997a), and the values of $\epsilon_{SCiCa^{2+}}$ were all arbitrarily set at 0.2 kcal/mole.

Finally, $U_{Ca^{2+}Ca^{2+}}$ corresponds to the repulsive interaction between two calcium cations, and consists of an excluded-volume Lennard-Jones term and a Coulomb term in equation 13:

$$U_{Ca^{2+}Ca^{2+}}(r_{Ca^{2+}Ca^{2+}}) = \frac{kq_1q_2}{\epsilon r_{Ca^{2+}Ca^{2+}}} + \epsilon_{Ca^{2+}} \left[\left(\frac{r_{Ca^{2+}}^0}{r_{Ca^{2+}Ca^{2+}}} \right)^{12} - 2 \left(\frac{r_{Ca^{2+}}^0}{r_{Ca^{2+}Ca^{2+}}} \right)^6 \right] \quad (13)$$

where $r_{Ca^{2+}}^0$ is defined in the text under equation 11, $\epsilon_{Ca^{2+}} = 0.05$ kcal/mole (Mayo et al. 1990), $k = 332$ is the conversion factor to express the energy in kilocalories per mole if the distance and charge are expressed in angstroms and electron charge units, respectively, $q_{Ca^{2+}} = 2e$ is the charge of the calcium cation, and ϵ is the dielectric constant of water.

Parameterization of the expressions for calcium–polypeptide interaction energy

As in our recent work on the parameterization of the torsional and correlation terms of the UNRES force-field (Oldziej et al. 2003; Liwo et al. 2004), the parameters of the calcium–polypeptide interaction terms were obtained from quantum mechanical calculations on model compounds, followed by fitting the resulting RFE hypersurfaces to the analytical expressions.

The energy landscapes for each of the different interacting residues were obtained by ab initio quantum mechanical calculations at the *Restricted Hartree-Fock* (RHF) level with the 6–31 G(d,p) basis set. The program GAMESS (Schmidt et al. 1993) was used in all calculations. The systems studied were Ca^{2+} – AcO^- , Ca^{2+} – PrO^- , Ca^{2+} – $AcNH_2$, Ca^{2+} – $PrNH_2$, and Ca^{2+} – $AcNHMe$ (where Ac and Pr denote the acetyl and propionyl group, respectively) to model the interactions of the calcium cation with the ionized aspartate and glutamate side-chains, the asparagine and glutamine side-chains, and the peptide group, respectively. The valence geometry of the side-chain sites was not optimized. The bond lengths and angles were taken from ECEPP/3 geometry (Némethy et al. 1992), and the peptide groups were assumed to be planar. The systems are shown in Figure 8, A through E.

For system X, the variables were the distance of the calcium cation from the center of mass of the system (SC in Fig. 8A–E), d , the angle between the C^α –SC axis, $\alpha_{Ca^{2+}}$, the angle of rotation of the Ca^{2+} cation about the C^α –SC axis, $\zeta_{Ca^{2+}}$, and the angle χ is the dihedral angle about a bond as shown in Figure 8, A through D. For Ca^{2+} – AcO^- , Ca^{2+} – PrO^- , Ca^{2+} – $AcNH_2$, and Ca^{2+} – $PrNH_2$, the energy was evaluated on the following grid: $3\text{ \AA} \leq d \leq 7\text{ \AA}$ with a 1 Å increment, $0^\circ \leq \alpha_{Ca^{2+}} \leq 180^\circ$ with a 15° increment, $0^\circ \leq \zeta_{Ca^{2+}} \leq 180^\circ$ with a 30° increment, and $0^\circ \leq \chi \leq 360^\circ$ with a 30° increment. For Ca^{2+} – $AcNHMe$, unlike other systems, AcNHMe is assumed to be rigid and the energy was evaluated on the following grid: $2\text{ \AA} \leq d \leq 6\text{ \AA}$ with a 1 Å increment, $0^\circ \leq \alpha_{Ca^{2+}} \leq 180^\circ$ with a 15° increment, and $0^\circ \leq \zeta_{Ca^{2+}} \leq 180^\circ$ with a 30° increment (Fig. 8E). Subsequently, the RFE surfaces in d and $\alpha_{Ca^{2+}}$ were calculated by integrating out the remaining degrees of freedom, as given by equation 14:

$$F = -\frac{1}{\beta} \ln \left\{ \frac{1}{(2\pi)^{n+1}} \int_0^{2\pi} \dots \int_0^{2\pi} \exp [-\beta E(d_{Ca^{2+}}, \alpha_{Ca^{2+}}, \zeta_{Ca^{2+}}, \chi)] d\zeta_{Ca^{2+}} d\chi \right\} \quad (14)$$

where n is the number of dihedral angles χ in a system, χ denotes the vector of the dihedral angles χ , and $\beta = 0.1$ (or, in other words, scaling down the energies) to account for the screening of the calcium cation by the solvent and, thereby, decreasing the interaction energy calculated in vacuo. Given the dielectric constant of water (78), the effective value of β at 298 K should be $1.69/78.4 \approx 0.02$. However, we consider short distances here between the calcium cation and the interacting residues, and the effective dielectric constant is smaller; therefore, β should take a value between 0.02 and 1.69, selected here as 0.1. Integration was carried out numerically.

The approximate analytical expressions for calcium–side-chain and calcium–peptide interaction energy were fitted to the RFE calculated numerically for each calcium–side-chain and calcium–peptide system by means of a nonlinear least-squares Marquardt method (Marquardt 1963). The resulting parameters are shown in Table 1. To take the effect of hydration into account, all the weights of the charge–charge ($w_{c,x}$), charge–dipole ($w_{dip,x}$), and charge–quadrupole ($w_{quad1,x}$) interaction terms in equations 8 and 9 were divided by the dielectric constant of water (i.e., $\epsilon = 78$).

The weights $w_{AGCa^{2+}}$, $w_{AGNCA^{2+}}$, $w_{SCCa^{2+}}$, $w_{pCa^{2+}}$, and $w_{Ca^{2+}Ca^{2+}}$ of the energy terms in equation 1 were set arbitrarily so that the calcium–polypeptide chain energy term is about one-fourth of the polypeptide chain energy as represented by the UNRES terms of equation 1. This led to the best performance after the force-field of equation 1 was tested on the crystal structure of a number of calcium-binding proteins, and the placement of the calcium in the binding site was monitored. The weights for each of the energy terms of equation 1 are provided in Table 2.

Docking calcium cations to calcium-binding proteins

To find the optimal position of the calcium cation(s) in the crystal structure of calcium-bound proteins (from which the calcium cat-

Table 2. The weights associated with the energy terms of the force-field of equation 1

Weight	Value
w_{SCSC}	1.00000
w_{SCp}	2.07913
w_{el}	0.54841
w_{tor}	1.72954
w_{tord}	1.23544
w_b	2.31367
w_{rot}	0.47054
$w_{loc-el}^{(3)}$	0.87864
$w_{loc-el}^{(4)}$	1.34109
$w_{loc-el}^{(5)}$	0.02730
$w_{loc-el}^{(6)}$	0.00741
$w_{turn}^{(3)}$	1.49008
$w_{turn}^{(4)}$	1.20192
$w_{turn}^{(6)}$	0.02391
w_{SCCa^2}	5.00000
w_{AGCa^2}	5.00000
w_{AGNCA^2}	10.0000
w_{pCa^2}	5.00000
$w_{Ca^2Ca^2}$	10.0000

The subscript for each weight corresponds to the energy term associated with it. For example, w_{SCp} corresponds to the weight of the peptide–side chain interaction energy. The weights $w_{loc-el}^{(3)}$ to $w_{turn}^{(6)}$ are the weights of the $U_{corr}^{(m)}$ terms, where m varies from 2 to 6.

ion(s) were removed), we ranked the different conformations obtained by placing the calcium cation in the neighborhood of a given side chain (i.e., Asp, Glu, Asn, or Gln) by energy based on predetermined distributions of $d_{SC'}$ and $\alpha_{Ca^{2+}}$ (see below). Subsequently, in each case, the energy of the system was minimized with respect to the calcium coordinates (keeping the polypeptide chain fixed). The conformation with the lowest energy was selected as the one with the calcium bound to the optimal calcium-binding site. To determine the distributions of $d_{SC'}$ and $\alpha_{Ca^{2+}}$, we used the crystal structures of calcium-bound proteins from the PDB. The list of the proteins is available in the electronic version

under the supplementary materials. We obtained the following mean and the variance for the distance distribution and angle distribution: $\langle d_{SC'} \rangle = 4.5 \text{ \AA}$ and $[\langle d_{SC'}^2 \rangle - \langle d_{SC'} \rangle^2]^{1/2} = 1.3 \text{ \AA}$, whereas $\langle \alpha_{Ca^{2+}} \rangle = 80^\circ$ and $[\langle \alpha_{Ca^{2+}}^2 \rangle - \langle \alpha_{Ca^{2+}} \rangle^2]^{1/2} = 41^\circ$, respectively.

For the apo- and the UNRES-predicted structure, the procedure for searching the optimal binding site was similar to that described above except that all degrees of freedom (i.e., those of the calcium cation and those of the polypeptide chain) were subjected to energy minimization. This allowed the EF-hand motif to adopt the «open form» conformation upon binding of calcium.

Figure 8. Settings for the calculation of the RFEs of model systems with quantum mechanical calculations to derive the parameters for the force-field. (A) Ca^{2+} -Asp system, modeled as AcO^- . d (equivalent to $d_{SC'}$ in Fig. 6) is varied from 3 Å to 7 Å in 1 Å increments. $\alpha_{Ca^{2+}}$ is changed from 0° to 180° in 15° increments. χ_1 changes from 0° to 360° in 30° increments and $\zeta_{Ca^{2+}}$ changes from 0° to 180° in 30° increments. (B) Ca^{2+} -Glu system, modeled as PrO^- . d (equivalent to $d_{SC'}$ in Fig. 6) is varied from 3 Å to 7 Å in 1 Å increments. $\alpha_{Ca^{2+}}$ is changed from 0° to 180° in 15° increments. χ_1 and χ_2 change from 0° to 360° in 30° increments and $\zeta_{Ca^{2+}}$ changes from 0° to 180° in 30° increments. (C) Ca^{2+} -Asn system, modeled as $AcNH_2$. d (equivalent to $d_{SC'}$ in Fig. 6) is varied from 3 Å to 7 Å in 1 Å increments. $\alpha_{Ca^{2+}}$ is changed from 0° to 360° in 15° increments. χ_1 changes from 0° to 360° in 30° increments and $\zeta_{Ca^{2+}}$ changes from 0° to 360° in 30° increments. (D) Ca^{2+} -Gln system, modeled as $PrNH_2$. d (equivalent to $d_{SC'}$ in Fig. 6) is varied from 3 Å to 7 Å in 1 Å increments. $\alpha_{Ca^{2+}}$ is changed from 0° to 360° in 15° increments. χ_1 and χ_2 change from 0° to 360° in 30° increments and $\zeta_{Ca^{2+}}$ changes from 0° to 360° in 30° increments. (E) Ca^{2+} -peptide system, modeled as $AcNHMe$. P corresponds to the center of mass of the molecule (equivalent to the peptide centers [p] in Fig. 6). Unlike other systems, $AcNHMe$ is rigid and calcium rotates around it. d is varied from 2 Å to 6 Å in 1 Å increments. $\alpha_{Ca^{2+}}$ is changed from 0° to 180° in 15° increments. $\zeta_{Ca^{2+}}$ changes from 0° to 180° in 30° increments.

Electronic supplemental material

The PDB code of the proteins that are used to calculate the probability distribution for the placement of calcium near an acidic residue is available in the electronic version.

Acknowledgments

This work was supported by NSF grant no. MCB00-03722. M.K. was an NIH biophysics trainee.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Bertini, I., Gray, H.B., Lippard, S.J., and Valentine, J.S. 1994. *Bioinorganic chemistry*. University Science Books, Mill Valley, CA.
- Chrysina, E.D., Brew, K., and Acharya, K.R. 2000. Crystal structures of Apo- and Holo-bovine α -lactalbumin at 2.2 Å resolution reveal an effect of calcium on inter-lope interactions. *J. Biol. Chem.* **275**: 37021–37029.
- Declercq, J.P., Evrard, C., Lamzin, V., and Parello, J. 1999. Crystal structure of the EF-hand parvalbumin at atomic resolution (0.91 Å) and at low temperature (100 K). Evidence for conformational multistates within the hydrophobic core. *Protein Sci.* **8**: 2194–2204.
- Dudev, T. and Lim, C. 2003. Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem. Rev.* **103**: 773–787.
- Dudev, T., Lin, Y., Dudev, M., and Lim, C. 2003. First-second shell interaction in metal binding sites in proteins: A PDB survey and DFT/CDM calculations. *J. Am. Chem. Soc.* **125**: 3168–3180.
- Einspahr, H. and Bugg, C.E. 1984. *Metal ions in biological systems*, Vol. 17 (ed. H. Sigel), pp. 51–97. Dekker, Basel, Switzerland.
- Forsen, S. and Koerdel, J. 1996. Biomolecular structure and dynamics: Experiment and theory. *J. Pharmaceut. Biomed.* **14**: 233–246.
- Frausto da Silva, J.J.R. and Williams, R.J.P. 1991. *The biological chemistry of the elements: The inorganic chemistry of life*. Oxford University Press, New York.
- Hanna, R. and Doudna, J. 2000. Metal ions in ribozyme folding and catalysis. *Curr. Opin. Chem. Biol.* **4**: 166–170.
- Jernigan, R., Raghunathan, G., and Bahar, I. 1994. Characterization of interactions and metal-ion binding-sites in proteins. *Curr. Opin. Struct. Biol.* **4**: 256–263.
- Katz, A.K., Glusker, J.P., Beebe, S.A., and Bock, C.W. 1996. Calcium ion coordination: A comparison with that of beryllium, magnesium and zinc. *J. Am. Chem. Soc.* **118**: 5752–5763.
- Kretsinger, R.H. and Nockolds, C.E. 1973. Carp muscle calcium-binding protein, II: Structure determination and general description. *J. Biol. Chem.* **248**: 3313–3326.
- Lee, J., Liwo, A., and Scheraga, H.A. 1999. Energy based de novo protein folding by conformational space annealing and an off-lattice united-residue force-field: Application to the 10–55 fragment of staphylococcal protein A and to apo-calbindin D9K. *Proc. Natl. Acad. Sci.* **96**: 2025–2030.
- Lippard, S.J. and Berg, J.M. 1994. *Principles of bioinorganic chemistry*. University Science Books, Mill Valley, CA.
- Liwo, A., Oldziej, S., Pincus, M.R., Wawak, R.J., Rackovsky, S., and Scheraga, H.A. 1997a. A united-residue force field for off-lattice protein-structure simulations, I: Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **18**: 849–873.
- Liwo, A., Pincus, M.R., Wawak, R.J., Rackovsky, S., Oldziej, S., and Scheraga, H.A. 1997b. A united-residue force field for off-lattice protein-structure simulations, II: Parameterization of short-range interactions and determination of weights of energy terms by Z-score optimization. *J. Comput. Chem.* **18**: 874–887.
- Liwo, A., Kazmierkiewicz, R., Czaplewski, C., Groth, M., Oldziej, S., Wawak, R.J., Rackovsky, S., Pincus, M.R., and Scheraga, H.A. 1998. United-residue force field for off-lattice protein-structure simulations, III: Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. *J. Comput. Chem.* **19**: 259–276.
- Liwo, A., Lee, J., Ripoll, D.R., Pillardy, J., and Scheraga, H.A. 1999a. Protein structure prediction by global optimization of a potential energy function. *Proc. Natl. Acad. Sci.* **96**: 5482–5485.
- Liwo, A., Pillardy, J., Kazmierkiewicz, R., Wawak, R.J., Groth, M., Czaplewski, C., Oldziej, S., and Scheraga, H.A. 1999b. Prediction of protein structure using knowledge-based off-lattice united-residue force field and global optimization methods. *Theor. Chem. Accounts* **101**: 16–20.
- Liwo, A., Czaplewski, C., Pillardy, J., and Scheraga, H.A. 2001. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* **115**: 2323–2347.
- Liwo, A., Arlukowicz, P., Czaplewski, C., Oldziej, S., Pillardy, J., and Scheraga, H.A. 2002. A method for optimizing potential-energy functions by a hierarchical design of the potential-energy landscape: Application to the UNRES force field. *Proc. Natl. Acad. Sci.* **99**: 1937–1942.
- Liwo, A., Oldziej, S., Czaplewski, C., Kozłowska, U., and Scheraga, H.A. 2004. Parameterization of backbone-electrostatic and multibody contributions to the UNRES force field for protein-structure prediction from ab initio energy surfaces of model systems. *J. Phys. Chem. B* **108**: 9421–9438.
- Marquardt, D.W. 1963. An algorithm for least-square estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.* **11**: 431–441.
- Mayo, S.A., Olafson, B.D., and Goddard, W.A. 1990. DREIDING: A generic force-field for molecular simulations. *J. Phys. Chem.* **94**: 8897–8909.
- McPhalen, C.A., Strynadka, N.C., and James, M.N. 1991. Calcium binding site in proteins: A structural perspective. *Adv. Prot. Chem.* **42**: 77–144.
- Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S., and Scheraga, H.A. 1992. Energy parameters in polypeptides, 10: Improved geometrical parameters and nonbonded interactions for use in ECEPP/3 algorithm with application to proline-containing peptides. *J. Phys. Chem.* **96**: 6472–6484.
- Oldziej, S., Kozłowska, U., Liwo, A., and Scheraga, H.A. 2003. Determination of the potentials of mean force for rotation about C α –C α virtual bonds in polypeptides from the ab initio energy surfaces of terminally-blocked glycine, alanine, and proline. *J. Phys. Chem. A* **107**: 8035–8046.
- Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. 1995. «AMBER»: A package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.* **91**: 1–42.
- Permyakov, S.E., Uversky, V.N., Veprintsev, D.B., Cherskaya, A.M., Brooks, C.L., Permyakov, E.A., and Berliner, L.J. 2001. Mutating aspartate in the calcium-binding site of α -lactalbumin: Effects on the protein stability and cation binding. *Prot. Eng.* **14**: 785–789.
- Pidcock, E. and Moore, G.R. 2001. Structural characteristics of protein binding sites for calcium and lanthanide ions. *J. Biol. Inorg. Chem.* **6**: 479–489.
- Schmidt, M.W., Baldrige, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., et al. 1993. General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**: 1347–1363.
- Skelton, N.J., Koerdel, J., and Chazin, W.J. 1995. Determination of the solution structure of apo calbindin D9K by NMR spectroscopy. *J. Mol. Biol.* **249**: 441–462.
- Svensson, L.A., Thulin, E., and Forsen, S. 1992. Proline *cis-trans* isomers in calbindin D9K, observed by x-ray crystallography. *J. Mol. Biol.* **223**: 601–606.
- Yamashita, M.M., Wesson, L., Eisenman, G., and Eisenberg, D. 1990. Where metal ions bind in proteins. *Proc. Natl. Acad. Sci.* **87**: 5648–5652.
- Yang, W., Lee, H.W., Hellinga, H., and Yang, J.J. 2002. Structural analysis, identification, and design of calcium-binding sites in proteins. *Proteins* **47**: 344–356.
- Yang, W., Jones, L.M., Isley, L., Yiming, Y., Lee, H.W., Wilkins, A., Liu, Z.R., Hellinga, H.W., Malchow, R., Ghazi, M., et al. 2003. Rational design of a calcium-binding protein. *J. Am. Chem. Soc.* **125**: 6165–6171.
- Yap, K.L., Ames, J.B., Swindells, M.B., and Ikura, M. 1999. Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins* **37**: 499–507.