Eur Biophys J (2010) 39:825–838
DOI 10.1007/s00249-009-0561-7

ORIGINAL PAPER

Docking of calcium ions in proteins with flexible side chains and deformable backbones

Ricky C. K. Cheng · Boris S. Zhorov

Received: 3 April 2009 / Revised: 20 October 2009 / Accepted: 23 October 2009 / Published online: 25 November 2009

Abstract A method of docking Ca^{2+} ions in proteins with flexible side chains and deformable backbones is proposed. The energy was calculated with the AMBER force field, implicit solvent, and solvent exposure-dependent and distance-dependent dielectric function. Starting structures were generated with Ca^{2+} coordinates and side-chain torsions sampled in 1000 $Å^3$ cubes centered at the experimental Ca^{2+} positions. The energy was Monte Carlo-minimized. The method was tested on fourteen Ca^{2+}-binding sites. For twelve Ca^{2+}-binding sites the root mean square (RMS) deviation of the apparent global minimum from the experimental structure was below 1.3 and 1.7 Å for Ca^{2+} ions and side-chain heavy atoms, respectively. Energies of multiple local minima correlate with the RMS deviations from the X-ray structures. Two Ca^{2+}-binding sites at the surface of proteinase K were not predicted, because of underestimation of Ca^{2+} hydration energy by the implicit-solvent method.

Keywords Energy minimization · Monte Carlo-minimization · Ca^{2+}-binding proteins

Abbreviations

MCM	Monte Carlo-minimization
RMSD	Root mean square deviation
AGM	Apparent global minimum

R. C. K. Cheng · B. S. Zhorov (✉)
Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON, Canada
e-mail: zhorov@mcmaster.ca

Introduction

Ca^{2+} ions play important physiological roles (Clapham 2007). An extracellular concentration of Ca^{2+} ions of ~1 mM is maintained with the help of calcitonin, parathyroid hormone, and Ca^{2+}-sensing receptors (Huang et al. 2007). Extracellular Ca^{2+} ions are required for activity of many enzymes, e.g., phospholipase A2 (Singh et al. 2007). Ca^{2+} ions protect proteases such as trypsin (Szmola and Sahin-Toth 2007) against autolysis. They also provide thermal stability to extracellular proteins such as thermolysin-like proteases (Veltman et al. 1998) and extracellular domains of transmembrane proteins such as epithelial cadherin (Prasad and Pedigo 2005). Besides the extracellular medium, high Ca^{2+} concentration is also maintained in the endoplasmic reticulum. The cytosolic concentration of Ca^{2+} ions (<1 μM) is much lower than that of the extracellular medium. Ca^{2+} influx through voltage-gated and/or ligand-gated Ca^{2+} channels in the plasma and endoplasmic-reticulum membranes raises Ca^{2+} concentration in the cytosol. This regulates or switches activity of various intracellular proteins. Some proteins, for example troponin, a key component of the skeletal and cardiac muscle-movement machinery, undergo conformational changes upon direct binding of Ca^{2+} ions (Galinska-Rakoczy et al. 2008; Suarez et al. 2008). Other proteins change their conformations in response to binding of Ca^{2+}-sensing proteins such as calmodulin. Important elements of Ca^{2+} homeostasis are Ca^{2+} pumps, which enable uphill movement of Ca^{2+} ions across the cytoplasmic and endoplasmic-reticulum membranes and thus restore low Ca^{2+} concentration in the cytosol.

X-ray structures of Ca^{2+}-binding proteins reveal diverse coordination geometries, coordination numbers, and Ca^{2+}–ligand distances (McPhalen et al. 1991; Nayal

and Di Cera 1994; Pidcock and Moore 2001; Yang et al. 2003). In most complexes Ca^{2+} ions are heptacoordinated by oxygen atoms located at the vertices of a pentagonal bipyramid (Pidcock and Moore 2001) with an average Ca^{2+}–O distance of ~2.4 Å (McPhalen et al. 1991). Data acquired from the X-ray structures have been used in knowledge-based approaches developed to identify Ca^{2+}-binding sites in proteins (Bagley and Altman 1995; Sodhi et al. 2004), to design Ca^{2+}-binding proteins (Yang et al. 2002, 2003, 2005), and to predict Ca^{2+}-binding sites (Deng et al. 2006). Such approaches employ rigid protein geometries and are not suitable for some applications. These include homology modeling, prediction of low-affinity Ca^{2+}-binding sites, simulation of Ca^{2+}-induced conformational changes, and recognition of Ca^{2+}-binding sites in NMR-determined and low-resolution structures. Methods of predicting Ca^{2+} positions and side-chain conformations by optimizing energy calculated with physics-based force fields are desirable for these applications, for understanding atomistic determinants of Ca^{2+}-binding affinity and selectivity, and for elaborating testable mechanisms of Ca^{2+} movement through Ca^{2+} channels and Ca^{2+} pumps.

In this study, we employed the Monte Carlo energy minimization (MCM) method (Li and Scheraga 1987) realized in ZMM software to develop and test a procedure for docking Ca^{2+} ions in proteins with completely flexible side chains and deformable backbones. We first selected the X-ray structures of Ca^{2+}-bound EDTA (Arriortua et al. 1992; Barnett and Uchtman 1979) and D-galactose-binding protein (Vyas et al. 1988) in which the Ca^{2+}–O distances are typical for Ca^{2+}-bound proteins. We then used these structures to parameterize solvent exposure-dependent and distance-dependent dielectric permittivity function (Tikhonov and Zhorov 2008) and optimize the multi-MCM docking procedure (Bruhova and Zhorov 2007; Tikhonov and Zhorov 2007). The procedure involves sampling of tens of thousands of Ca^{2+} positions and side-chain conformations within and around the Ca^{2+} binding region and MC-minimizing the sampled structures. We tested the method on 14 Ca^{2+}-binding sites with known structures. For 12 sites the method correctly predicted Ca^{2+} positions and side-chain conformations.

Methods

Modeling Ca^{2+}-binding proteins

X-ray structures of selected Ca^{2+}-binding proteins (Acharya et al. 1989; Betzel et al. 1988; Chattopadhyaya et al. 1992; Gros et al. 1989; Holland et al. 1992; Meyer et al. 1988; Toyoshima et al. 2000; Vyas et al. 1988) were taken from the Protein Data Bank (Berman et al. 2000). A double-shell model (Zhorov and Lin 2000) was built around the X-ray positions of Ca^{2+} ion(s) to reduce the computational cost of energy minimizations. The flexible inner shell contains residues that have at least one atom within 10 Å of the Ca^{2+} ion(s). The fixed outer shell contains residues that do not belong to the flexible shell and have at least one atom within 20 Å of the Ca^{2+} ion(s). All calculations were performed using ZMM software (Zhorov 1981, 1983) (http://www.zmmsoft.com). Upon importing X-ray structures into ZMM, tautomers of histidines and starting orientations of the hydroxy and carboxamide groups in the fixed-shell amino acids were adjusted to minimize steric clashes and to maximize number of H-bonds. Energy was presented as a sum of van der Waals, electrostatic, torsional, and dehydration components. Interactions between atoms involved in ion–ion and ion–dipole interactions were calculated at all distances, and a cutoff of 8 Å was used for other interactions. Alpha carbon atoms were constrained to their crystallographic positions by pins, unless stated otherwise. A pin is a flat-bottom parabolic penalty function that allows a C^{α} atom of the model to deviate from the corresponding position in the template by up to 1 Å without energy penalty and imposes a penalty of 10 kcal mol^{-1} $Å^{-1}$ for larger deviations. Without the pins, unavoidable steric clashes at early stages of MC-minimizations would significantly disrupt protein folding.

The force field

Ca^{2+}-binding to a protein is governed by many interactions, the strongest among which are favorable electrostatic interaction with the first coordination-shell oxygen atoms and unfavorable dehydration of Ca^{2+} ions and the oxygen atoms. We used the first generation of the AMBER force field (Weiner et al. 1984, 1986) that was developed for simulations with implicit solvent. The later versions of the AMBER force field were parameterized for use with the explicit solvent (Cornell et al. 1995). Because the sampling efficiency of Monte Carlo-minimizations decreases with the number of variables, addition of numerous water molecules would dramatically reduce the chances of solving multiple-minima problems in highly ragged energy hypersurfaces of protein-Ca^{2+} complexes. The available AMBER parameters for Ca^{2+} are similar to those proposed by others (Åqvist 1990; Hori et al. 1988; Yasuhiko and Setsuko 1991). Electrostatic energy was calculated with the solvent exposure-dependent and distance-dependent dielectric function (Tikhonov and Zhorov 2008), which was parameterized as described in a later section. Water molecules, which are seen in X-ray structures, were not included in the models. The dehydration energy was

estimated by the implicit-solvent method, which provides five parameters for each of 17 types of atoms and groups common in proteins (Lazaridis and Karplus 1999). Because no hydration parameters for Ca^{2+} ions are available, we assigned a value of −60 kcal/mol for the reference hydration free energy (ΔG_i^{ref}) and the hydration free energy of the isolated Ca^{2+} ion (ΔG_i^{free}). The atomic volume, the thickness of the hydration shell, and the van der Waals radius of Ca^{2+} were assigned respective values for the NH_3^+ group, the only cationic group parameterized by Lazaridis and Karplus (1999). At first sight, the chosen energy values are inconsistent with the experimental hydration energy of Ca^{2+}, which is −397 kcal/mol (Edsall and McKenzie 1978). The following considerations support our choice. $\Delta G_i^{ref} = \Delta G_i^{free} = -20$ kcal/mol for the NH_3^+ group (Lazaridis and Karplus 1999) are much smaller than experimental hydration energies of NH_4^+ ($\Delta H = -86.8$ kcal/mol) and $EtNH_3^+$ ($\Delta H = -80.5$ kcal/mol) cations (Meot-Ner 1987). However, the values of Lazaridis and Karplus represent the energetics of the first hydration shell, whereas the experimental data represent the entire hydration energy. We varied ΔG_i^{ref} and ΔG_i^{free} from −80 to −20 kcal/mol in test simulations of Ca^{2+} complexes with EDTA and D-galactose binding protein and also tried different combinations of values for solvent exposure-dependent and distance-dependent dielectric function (see below). These simulations predicted complexes with the same Ca^{2+}-coordination patterns, but different Ca^{2+}–O distances. Calculations with larger Ca^{2+} hydration values ($\Delta G_i^{ref} = \Delta G_i^{free} = -80$ kcal/mol) overestimated Ca^{2+}–O distances, whereas calculations with smaller values ($\Delta G_i^{ref} = \Delta G_i^{free} = -40$ kcal/mol) underestimated the distances. On the basis of these results, we chose $\Delta G_i^{ref} = \Delta G_i^{free} = -60$ kcal/mol. These values enable our method to reproduce experimental Ca^{2+}–O distances.

Besides the implicit-solvent component, the energy of Ca^{2+} binding includes strong electrostatic attractions between the ion and its first coordination-shell oxygen atoms and weaker electrostatic interactions with atoms beyond the first coordination shell. Further parameterization of the implicit-solvent energy function along with the electrostatic energy function is a challenging task that would require a stand-alone study. Here we did not attempt to modify the assigned hydration parameters and parameterized only the dielectric function as described below.

Solvent exposure-dependent and distance-dependent dielectric function

Different approaches have been proposed for calculation of electrostatic interactions with Ca^{2+} ions (Åqvist 1990; Hori et al. 1988; Meiler and Baker 2006; Tikhonov and Zhorov 2007). In this study, we calculated electrostatic interactions using Coulomb's law with a solvent exposure-dependent and distance-dependent dielectric function (Tikhonov and Zhorov 2008) $\varepsilon = d\,[\varepsilon_1 - s\,(\varepsilon_1 - \varepsilon_0)]$, where d is the distance between interacting atoms, s is the solvent-exposure factor that ranges from 0 for a pair of water-exposed atoms and 1 for a pair of atoms buried in the protein, and ε_0 and ε_1 are, respectively, the lower and upper limits of dielectric permittivity. ε_0 and ε_1 were optimized to reproduce Ca^{2+}–oxygen distances in the X-ray structures of Ca^{2+} complexes with EDTA (Arriortua et al. 1992; Barnett and Uchtman 1979) and D-galactose-binding protein (Vyas et al. 1988) and root mean square deviation (RMSD) between the predicted and X-ray structures of the Ca^{2+}-binding site in the protein (Table 1). We tested all combinations of $\varepsilon_0 = [1, 2]$ and $\varepsilon_1 = [1, 2, \ldots, 8;]$ and obtained a Ca^{2+}–O distance of ~2.4 Å with [$\varepsilon_0 = 1$, $\varepsilon_1 = 3$] and [$\varepsilon_0 = 1$, $\varepsilon_1 = 4$]. This is in good agreement with the average Ca^{2+}–O distances of 2.42–2.44 Å observed in Ca^{2+}–EDTA complexes (Arriortua et al. 1992; Barnett and Uchtman 1979) and in various Ca^{2+}–protein complexes (McPhalen et al. 1991). The combination [$\varepsilon_0 = 1$, $\varepsilon_1 = 3$] was used in further calculations.

Table 1 Characteristics of two Ca^{2+}-binding complexes predicted with various values of the dielectric-function terms ε_0 and ε_1

ε_0	ε_1	Average Ca^{2+}–O distance (Å)		RMSD[b], Å	
		EDTA	DGBP[a]	Ca^{2+} ion	10 Å shell
1	1	2.21	2.42	0.4	1.3
1	2	2.30	2.43	0.1	1.6
1	3	2.38	2.44	0.4	1.4
1	4	2.43	2.45	0.4	1.1
1	5	2.50	2.48	0.3	1.3
1	6	2.55	2.46	0.3	1.2
1	7	2.59	2.44	0.5	1.2
1	8	2.63	2.48	0.4	1.2
2	2	2.37	2.55	0.2	1.7
2	3	2.42	2.64	0.3	1.3
2	4	2.48	2.66	0.3	1.3
2	5	2.52	2.59	0.2	1.3
2	6	2.56	2.62	0.3	1.2
2	7	2.60	2.60	0.2	1.2
2	8	2.65	2.63	0.4	1.3
X-ray		2.42	2.40	n.a	n.a

[a] D-galactose-binding protein, Ca^{2+}–O distances do not include coordinating bonds with water

[b] Root mean square deviation of the AGM (apparent global minimum) from the X-ray structure of D-galactose-binding protein (1 GBP)

Monte Carlo-minimization

Because the ragged energy landscape limits the convergence radius for a single MCM trajectory, we employed a multi-MCM procedure (Bruhova and Zhorov 2007; Tikhonov and Zhorov 2007). At the first (seeding) stage 20,000 starting structures were randomly generated and briefly MC-minimized to ensure broad sampling of positions of Ca^{2+} ions and side-chain conformations. In the second (refinement) stage the lowest-energy structures collected at the first stage were further optimized in long MCM trajectories. In proteins with two proximal Ca^{2+}-binding sites, fixing one Ca^{2+} ion at its X-ray position would bias the docking of the other Ca^{2+} ion towards its X-ray position. Therefore in such proteins both Ca^{2+} ions and side chains around these ions were sampled simultaneously.

At the seeding stage, x, y, and z-coordinates of a Ca^{2+} ion were randomly displaced from the X-ray coordinates up to 5 Å, which gives a maximum of 8.7 Å displacement. All side-chain torsions of residues that had at least one atom within 5 Å of the X-ray position of the Ca^{2+} ion were assigned random angles. Each starting point was optimized in a brief 20-step MCM trajectory and each energy minimization within the trajectory was terminated after 200 iterations. Preliminary calculations demonstrated that a Ca^{2+} ion could easily move away from its starting position towards acidic residues, whereas the latter were less mobile. To ensure broader sampling of Ca^{2+} ions, their positions during the seeding-stage MC-minimizations were constrained to be within 2 Å of corresponding starting positions. These constraints did not preclude approach of the protein side chains to Ca^{2+}, but precluded movement of Ca^{2+} ions by more than 2 Å from the seeded starting point. Structures up to 200 kcal/mol from the lowest-energy minimum found in all the seeding-stage trajectories were collected and refined by long MCM trajectories. Each refining trajectory was terminated when the last 1,000 consecutive energy minimizations did not improve the lowest energy found in the trajectory. This termination criterion was chosen to ensure thorough optimization of each complex. Doubling this criterion to 2,000 minimizations increased the computational cost with only little effect on results (Fig. 1). The apparent global minimum (AGM) was chosen among minimum-energy complexes collected in all the refining trajectories.

Clustering minimum-energy structures

Both seeding and refining stages produce large number of minimum-energy complexes, which are difficult to collect and analyze. Therefore, during MC-minimizations structures were clustered as follows. When a new minimum

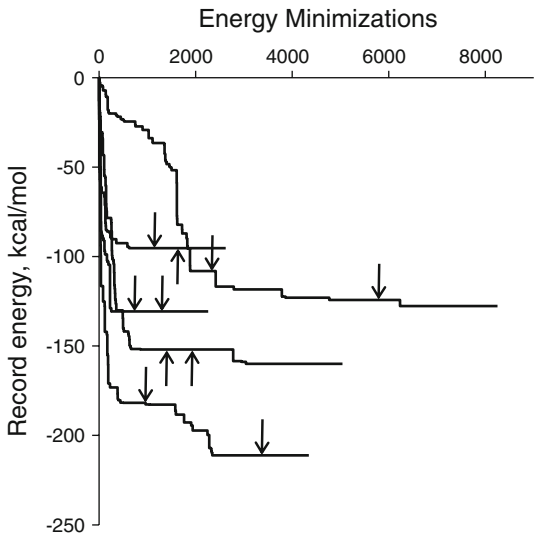

Fig. 1 Convergence of five refining MCM trajectories for Ca^{2+} docking in α-lactalbumin. Each trajectory was terminated when 2,000 consecutive energy minimizations did not improve the record energy (the lowest energy achieved prior to a given energy-minimization step). *Arrows* show steps at which the same trajectories would have been terminated with the standard criterion of convergence (1,000 unproductive steps) and the halved criterion (500 unproductive steps)

with energy below a threshold value from the current lowest-energy structure found at the given stage was generated, it was compared with all the minimum-energy structures collected at that stage. If a match was found with RMSD < 1 Å for a specified set of atoms, the two structures were merged and the lower-energy structure was chosen to represent the cluster. The set of atoms included Ca^{2+} ion(s) and a pair of atoms from each residue which has at least one atom within 5 Å of the Ca^{2+} ions in the X-ray structure. The pair of atoms in a residue is C^{α} and a heavy atom most distal from C^{α} along the side-chain covalent bonds. (When we included all side-chain heavy atoms in the RMSD calculations, computational requirement increased substantially with minimal impact on results.) The above clustering procedure dramatically reduced the number of geometrically similar structures thus increasing the diversity of the collected structures.

Comparing predicted and X-ray structures

The comparison was performed using five RMSD measurements between a minimum-energy and the X-ray structures:

1. Ca^{2+} ion(s) only;
2. heavy atoms within 5 Å of the X-ray position of Ca^{2+} ion(s);
3. side-chain heavy atoms of residues within 5 Å of the X-ray position of Ca^{2+} ion(s);

4. heavy atoms within 10 Å of the X-ray position of Ca^{2+} ion(s); and
5. side-chain heavy atoms of residues within 10 Å of the X-ray position of Ca^{2+} ion(s).

For all RMSD calculations, minimum-energy structures were not re-oriented to X-ray structure. In cases of simultaneous docking of two Ca^{2+} ions, they were considered identical despite the fact they have different "residue" numbers in the PDB files. This precluded incorrect contributions to RMSDs from those structures in which two Ca^{2+} ions exchanged their positions during docking. A prediction was considered a success only if all five RMSD values between the AGM and the X-ray structure were below 2 Å. The average Ca^{2+}–O distance in the X-ray structure (excluding water molecules) was compared with that in the model. We also calculated RMS deviations of Ca^{2+} ligands in the experimental and modeled structures from the ideal octahedron and pentagonal bipyramid vertices that are 2.4 Å distant from the polygon center. The ideal polygon was superimposed with the X-ray and modeled structures by the least squares method. The coordination-geometry type of an X-ray structure was chosen on the basis of the number of oxygen atoms within 2.8 Å of the Ca^{2+} ion. The coordination-geometry type of the modeled structure was assigned according to the best-fitted ideal polygon. Because the modeled structures lack explicit water molecules, the match of the predicted and modeled coordination-geometry types was not used as a criterion for successful prediction.

Results

Because of the large computational cost of Ca^{2+} docking in flexible proteins, we selected for testing only fourteen Ca^{2+}-binding sites in seven proteins. In these proteins Ca^{2+}-coordination number varies from six to eight and the number of water molecules in the Ca^{2+}-binding sites varies from zero to four. Simultaneous docking of a pair of Ca^{2+} ions in proximal binding sites is a more challenging task than docking of a single ion. Three of the selected proteins have pairs of proximal Ca^{2+}-binding sites with inter-calcium distances varying from 3.8 to 12.0 Å. Structural features of each binding site and results of Ca^{2+} docking are described below.

Ca^{2+} ATPase, a member of the P-type ATPases family (Toyoshima et al. 2000), carries out active transport of Ca^{2+} ions. We focused on the site with two Ca^{2+} ions in the transmembrane domain, which comprises ten α-helices. The Ca^{2+} ions are surrounded by four transmembrane helices, two of which unwind to provide optimal coordination geometry (Toyoshima et al. 2000). Figure 2a shows superposition of 100 out of 10,000 randomly generated starting structures, which cover both Ca^{2+}-binding sites seen in the X-ray structure. The multi-MCM procedure yielded a large number of minimum-energy complexes (Fig. 2b) indicating a ragged character of the energy hypersurface.

The AGM matches well the X-ray structure (Fig. 2c, d) with RMSDs of Ca^{2+} ions of 0.7 Å (Table 2). In both 5 Å and 10 Å shells, all heavy atoms and side-chain heavy atoms RMS-deviated from the X-ray structure by 0.7 Å and 0.8 Å, respectively. The second carboxylate oxygen of Glu908 in chain A (Glu^{A908}) coordinated Ca^{2+} in the model to fill the vacancy, which is occupied by a water molecule in the X-ray structure. MC minimizations of the Ca^{2+}−ATPase model without Ca^{2+} ions predicted structures in which side-chain heavy atoms in the 5 and 10 Å shells RMS-deviated from the X-ray structure by 2.1 and 1.6 Å, respectively (not shown). Thus, as expected, side-chain conformations in the Ca^{2+}-binding site depend on interactions with Ca^{2}.

Relative energies of minimum-energy complexes correlate with their RMSDs from the X-ray structure (Fig. 3a). Electrostatic interactions provide the major contribution to the interaction energy (Fig. 3b). Variations of hydration energy among different structures are much smaller than variations of electrostatic energy (Fig. 3c). The dehydration energy does not correlate with RMSD (Fig. 3c). However, the absence of points below the dashed line in Fig. 3c indicates a substantial dehydration of Ca^{2+} ions in low-RMSD complexes. RMSD values calculated over all heavy atoms in the flexible shell within 10 Å of the crystallographic positions of Ca^{2+} ions (Fig. 3a–c) do not exceed 1.2 Å, indicating that overall geometry of the Ca^{2+}-binding sites is rather conserved even in the high-energy local minima. By contrast, RMSD values calculated just over two Ca^{2+} ions may be as high as 8 Å, but the total energy correlates well with this metric (Fig. 3d). The total energy of any minimum-energy complex, in which Ca^{2+} ions deviate less than 1 Å from the X-ray structure, does not exceed 25 kcal/mol (Fig. 3d).

Calmodulin, a sensor of intracellular concentration of Ca^{2+} ions, has four EF hand-type Ca^{2+}-binding sites located in helix–loop–helix motifs. Micromolar affinity of calmodulin for Ca^{2+} decreases dramatically when just one aspartate residue is mutated to glutamate (Wu and Reid 1997) indicating that the affinity is highly sensitive to binding site structure. The X-ray structure at 1.7 Å resolution (Chattopadhyaya et al. 1992) shows two pairs of Ca^{2+}-binding sites. At each site, Ca^{2+}-coordinating oxygen atoms form a pentagonal bipyramid. The oxygen atoms are provided by side chains of residues in relative positions i, $i + 2$, $i + 4$, and $i + 11$ (a bidentate glutamate), the main chain of residue $i + 6$, and a water molecule. We

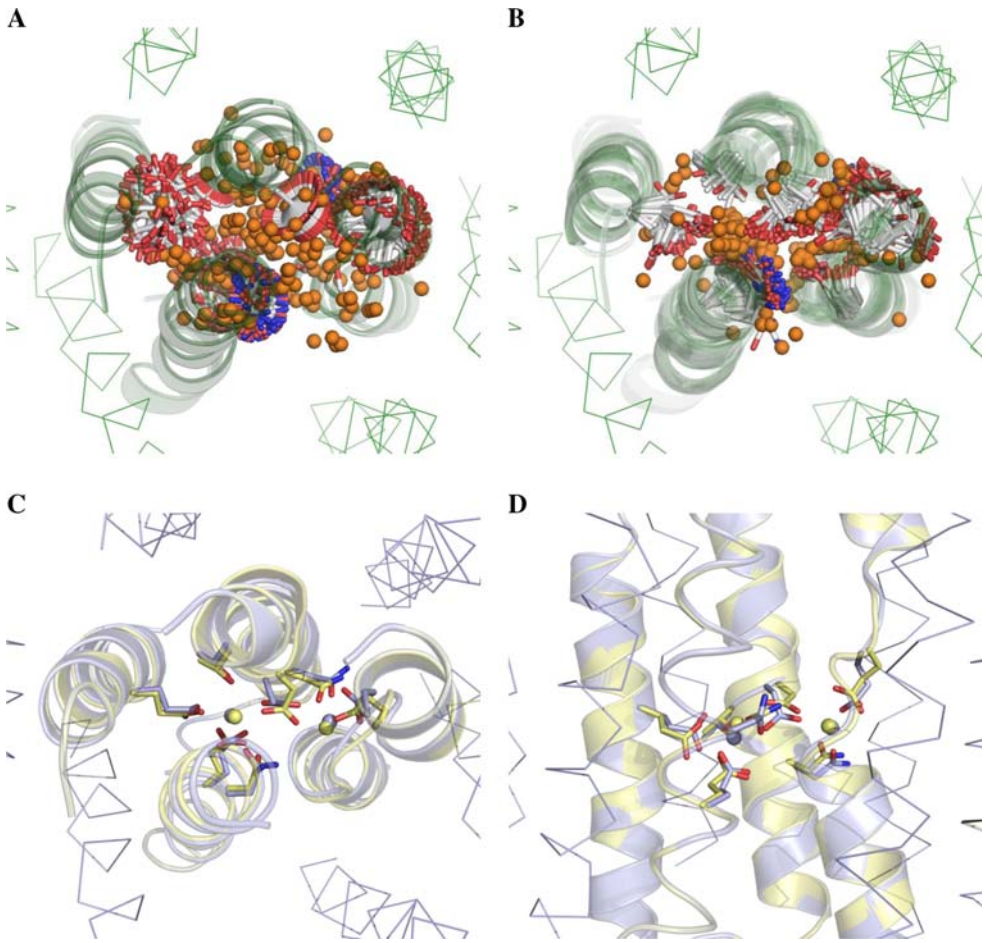

Fig. 2 Starting, minimum-energy, and X-ray structures of Ca^{2+} complexes with Ca^{2+}–ATPase. Helices with Ca^{2+}-binding residues are shown as *ribbons*. For clarity, one of the four ribbons is shown as a rod. C^{α} tracings show helices more remote from the Ca^{2+}-binding site. Ca^{2+} ions are shown as small spheres. **a** Extracellular view at superposition of 100 out of 10,000 starting structures generated at the seeding stage. Side chains that have at least one atom within 5 Å of either of the two Ca^{2+} ions seen in the X-ray structure are shown as *sticks* with *red* oxygen atoms, *blue* nitrogen atoms, and *gray* carbon atoms. **b** Superposition of 100 minimum-energy structures obtained at the refinement stage. Side chains involved in Ca^{2+} coordination are shown as *sticks*. Note deviations of backbones of the Ca^{2+}-binding helices between individual structures. **c** and **d** Extracellular and side views at superposition of the AGM and X-ray structures. Helices, carbon atoms, and Ca^{2+} ions in the AGM are shown in *light blue* whereas those in the X-ray structure are shown in *yellow*

simultaneously docked two Ca^{2+} ions to the N-terminal sites. Side chains of Asp^{58} and Asp^{64} provided additional Ca^{2+} ligands at site 2. The AGM complex matches the experimental structure well (Fig. 4; Table 2) and energy of local minima correlates with the RMS deviations of the Ca^{2+} ions from their X-ray positions (Fig. 5).

Elastase is a member of the serine protease family. The X-ray structure of a porcine pancreatic elastase at 1.65 Å resolution (Meyer et al. 1988) has a Ca^{2+}-binding site with the Ca^{2+} ion hexacoordinated by five protein oxygen atoms and a water molecule at vertices of an octahedron. The AGM matches well the X-ray structure (Fig. 4; Table 2) and the energy of local minima correlates with deviation of the Ca^{2+} ion from its X-ray position (Fig. 5). The major difference between the experimental and predicted structure is that Glu^{70} and Glu^{80} coordinate Ca^{2+} monodentately in the X-ray structure but bidentately in the AGM. Ca^{2+}–Glu^{70} interaction in the X-ray structure is considered monodentate only because one of the Ca^{2+}–O distances (2.9 Å) slightly exceeds the chosen criterion of 2.8 Å for a Ca^{2+}–O coordination bond (Harding 2002, 2006). A likely cause of this discrepancy is the absence of explicit water molecules in the model, which would compete with the second oxygen atom of Glu^{70}.

α-Lactalbumin is a small globular protein stabilized by four disulfide bonds. The X-ray structure (Acharya et al. 1989) shows a Ca^{2+}-binding site formed by three unidentate Asp carboxylates, two backbone carbonyls, and two water molecules at the vertices of a pentagonal bipyramid. The Ca^{2+}-binding site contains only four residues preceded

Table 2 Experimental and predicted characteristics of Ca^{2+}-binding sites

Protein	PDB code	Site no.	X-ray structure					Model				RMSD from X-ray, Å[f]				
			Mean Ca^{2+}–O distance, Å[b]	No. of Ca^{2+} ligands		Coordination geometry		Mean Ca^{2+}–O distance, Å[b]	No. of Ca^{2+} ligands[e]	Coordination geometry		Ca^{2+} ions	5 Å shell		10 Å shell	
				Water	Protein	Type[c]	Ligands' RMSD[d]			Type[c]	Ligands' RMSD[d]		All atoms	Side chains	All atoms	Side chains
Ca^{2+}–ATPase	1su4	1	2.43	0	7	PB	0.38	2.45	7	OH	0.61	0.7[j]	0.7[j]	0.8[j]	0.7[j]	0.8[j]
		2	2.42	2	5	PB	0.39	2.46	6	PB	0.68[h]					
Calmodulin[a]	1cll	1	2.46	1	6	PB	0.40	2.40	6	PB	1.08	0.4[j]	1.0[j]	1.3[j]	1.2[j]	1.5[j]
		2	2.38	1	6	PB	0.43	2.49	8	PB	0.48[h, i]					
Thermitase	1tec	1	2.41	0	7	PB	0.36	2.39	7	PB	0.42	0.4	1.0	1.2	0.8	1.0
		2	2.60	0	6	PB[g]	0.67	2.49	6	PB	0.78	1.3	1.5	1.6	1.3	1.5
Thermolysin	8tln	1	2.47	1	7	PB[g]	0.39	2.42	7	PB	0.92[h, i]	1.0[j]	1.4[j]	1.4[j]	0.9[j]	1.1[j]
		2	2.43	2	4	OH	0.20	2.46	4	OH	0.30[h, i]					
		3	2.36	3	4	PB	0.29	2.43	4	OH	0.53	0.4	0.8	0.8	0.8	1.0
		4	2.30	2	5	PB	0.60	2.53	5	OH	0.67[h]	0.9	1.0	1.0	1.0	1.1
α-lactalbumin	1alc	1	2.29	2	5	PB	0.19	2.45	6	PB	0.54	0.6	1.1	1.3	1.2	1.5
Elastase	3est	1	2.48	1	5	OH	0.30	2.52	7	OH	0.75[h]	0.7	1.4	1.7	1.4	1.7
Proteinase K	2prk	1	2.51	4	4	PB[g]	0.39	n.a.	n.a.	n.a.	n.a.	2.6	1.6	2.0	1.7	2.2
		2	2.49	2	3	PB[g]	0.27	n.a.	n.a.	n.a.	n.a.	5.3	1.5	1.5	1.1	1.2

[a] N-terminus

[b] Coordinating bonds with water are not counted

[c] PB, pentagonal bipyramid (heptacoordinated); OH, octahedron (hexacoordinated)

[d] Ligands are superimposed on to vertices of ideal pentagonal bipyramid and octahedron that are 2.4 Å from the center of the polygon

[e] Ca^{2+} ligands are from proteins because the implicit-solvent method was used

[f] Only heavy atoms are included in RMSD calculations

[g] The X-ray structure has six or eight Ca^{2+} ligands, but they fit best to vertices of a pentagonal bipyramid

[h] The pentagonal bipyramid was re-oriented

[i] Involves loss of a Ca^{2+} ligand in the model that is Ca^{2+}-binding in the X-ray structure, or a gain of a Ca^{2+} ligand in the model that is not Ca^{2+}-binding in the X-ray structure

[j] RMSDs are combined because Ca^{2+} ions in sites 1 and 2 were docked simultaneously

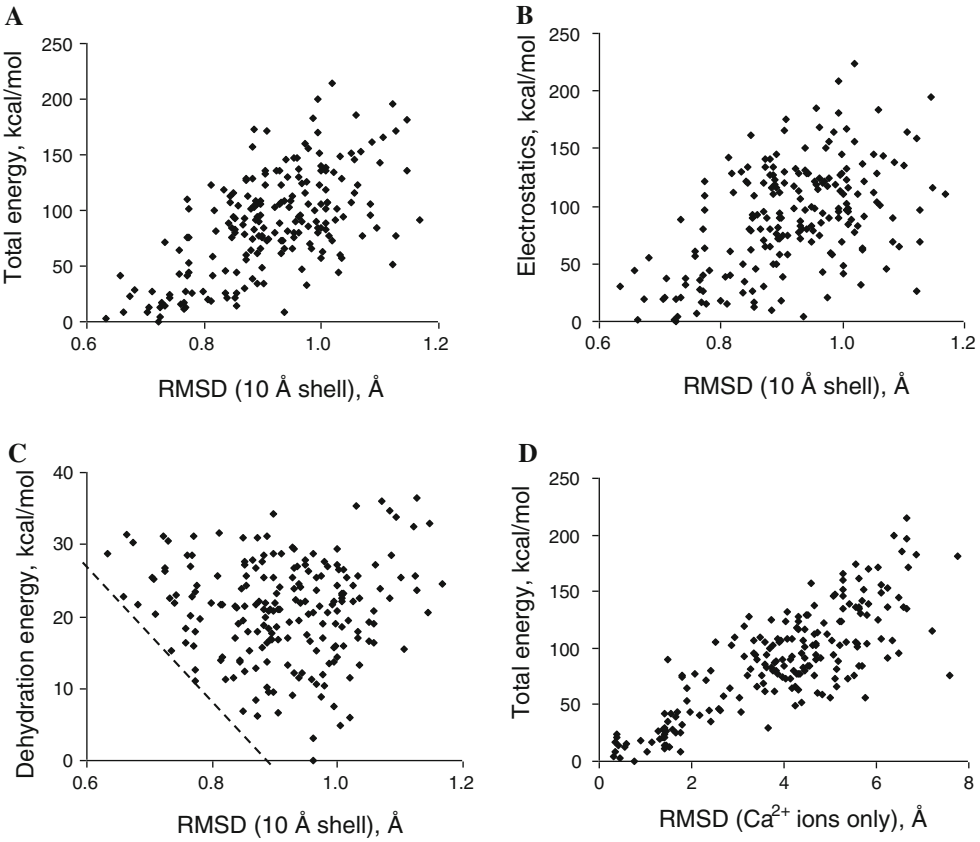

Fig. 3 The relative energy of minimum-energy Ca^{2+} complexes with Ca^{2+}–ATPase plotted against their RMSDs from the X-ray structure. **a** and **b** RMSD of all heavy atoms in the 10 Å shell around Ca^{2+} ions correlates with the total energy (**a**) and with electrostatic contribution to the total energy (**b**). **c**, When all of the minimum-energy complexes are considered, the relative dehydration energy does not correlate with RMSD, but the absence of points below the *dashed line* indicates substantial dehydration of low-RMSD complexes. **d** The total energy versus RMSD of two Ca^{2+} ions. In some minimum-energy complexes, Ca^{2+} ions deviate by up to 8 Å from the X-ray structure, but the energy of high-RMSD structures is high. In contrast, the total energy of any minimum-energy complex, in which Ca^{2+} ions deviate less than 1 Å from the X-ray structure, is below 25 kcal/mol

by a single-turn 3_{10} helix (McPhalen et al. 1991). Therefore the site is considered unrelated to typical EF-hand Ca^{2+}-binding motifs, which are 12 residues long (Acharya et al. 1989). An additional Ca^{2+}-coordinating bond in the AGM complex was provided by the side chain of Asp^{84}. The AGM matches well the X-ray structure (Table 2; Fig. 4) and energy of local minima correlates with their RMSDs from the X-ray structure (Fig. 5).

Thermitase is a thermostable serine protease (McPhalen et al. 1991). The X-ray structure (Gros et al. 1989) shows two Ca^{2+}-binding sites, which contain side-chain nitrogen atoms from Asn^{E85} and Gln^{E66}. Our calculations predict that these side chains contribute oxygen atoms rather than nitrogen atoms to Ca^{2+} binding. One of the Ca^{2+} ions is coordinated by seven protein oxygen atoms at the vertices of a pentagonal bipyramid. Another Ca^{2+} ion is coordinated by six protein oxygen atoms. The coordination geometry at the second site matches a pentagonal bipyramid better than an octahedral pyramid, which is expected for a hexacoordinated Ca^{2+} ion (McPhalen et al. 1991). The vacant axial position is likely to be filled by a water molecule not modeled in the X-ray structure. Indeed, the difference electron-density map ($F_o - F_c$) associated with the X-ray structure suggests a significant electron density not accounted for in the X-ray structure. The AGM matches well with the X-ray structure (Fig. 4; Table 2). The predicted positions of Ca^{2+} ions are 0.4 and 1.3 Å from the respective X-ray positions (Fig. 4).

Thermolysin is a member of the bacterial neutral protease family. The X-ray structure at 1.6 Å resolution (Holland et al. 1992) shows four Ca^{2+}-binding sites. The first site contains eight Ca^{2+}-coordinating oxygen atoms including a water oxygen and the backbone oxygen of Glu^{E187}. The oxygen atoms are at the vertices of a pentagonal bipyramid with a bidentate Glu^{E190} at an axial position. The second Ca^{2+} ion is coordinated by six oxygen atoms including two water oxygen atoms and the backbone oxygen from Asn^{E183}. This site has octahedral geometry. The third Ca^{2+} ion is coordinated by seven oxygen atoms including three water oxygen atoms and the backbone oxygen from Gln^{E61} at the vertices of a pentagonal bipyramid (Fig. 4). The fourth site is formed by seven oxygen atoms, two of which are from water molecules and two from Thr^{E194}. The coordination geometry does not resemble a typical pentagonal bipyramid observed in many sites with heptacoordinated Ca^{2+}. A possible cause is an unusual bidentate coordination of Ca^{2+} by the backbone and side-chain oxygen atoms of Thr^{E194}. Because Ca^{2+} ions in the first two sites are just 3.8 Å apart and share three Ca^{2+}-binding residues, we docked both ions in these sites simultaneously and predicted the AGM with the RMSD of the Ca^{2+} ions 1.0 Å

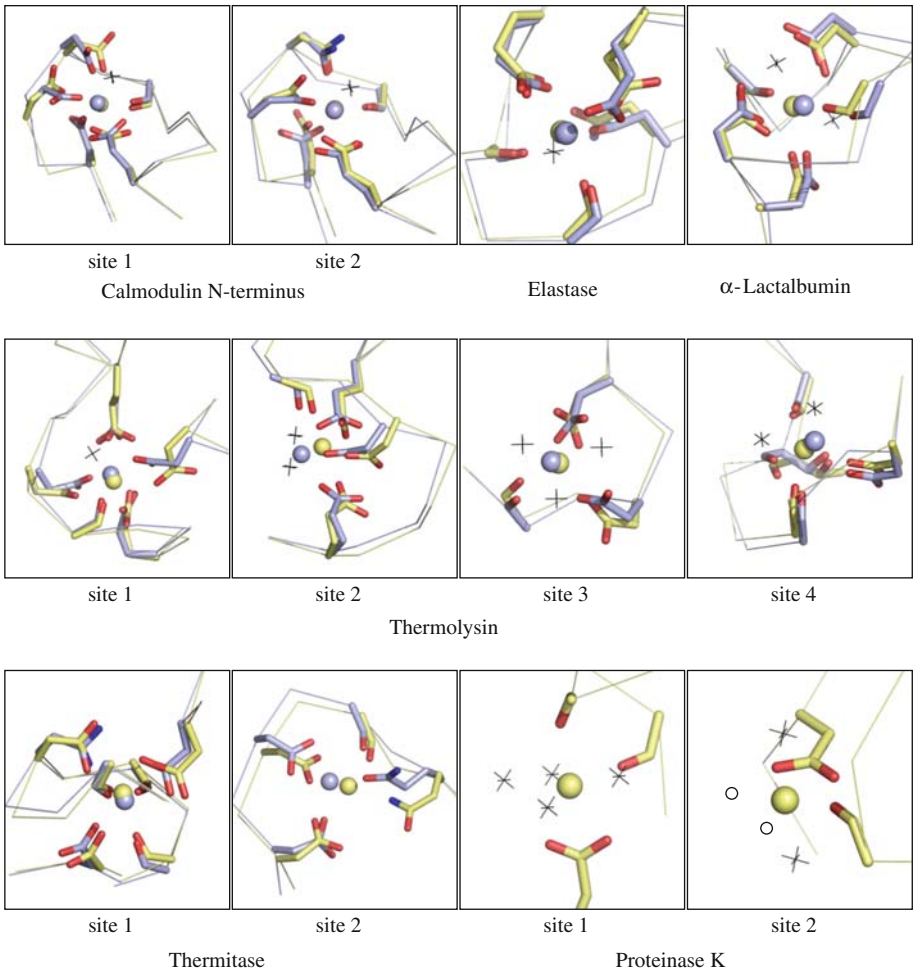

Fig. 4 Predicted and experimental structures of 12 Ca^{2+}-binding sites in six proteins. Oxygen atoms are *red*. Carbon atoms, Ca^{2+} ions, and C^{α} tracings are *light blue* in the predicted structures and *yellow* in the X-ray structures. Ca^{2+}-coordinating water molecules in the X-ray structures are shown as *crosses*. *Low-right corner*, two Ca^{2+}-binding sites of proteinase K seen in the X-ray structure. Four out of eight Ca^{2+}-coordinating oxygen atoms are from water molecules. At the second site, the Ca^{2+} ion is apparently coordinated by five oxygen atoms, including two water molecules. Disposition of Ca^{2+}-coordinating oxygen atoms in this site is approximated by a pentagonal bipyramid (McPhalen et al. 1991), suggesting that additional water molecules (*black open circles*), which are not shown in the X-ray structure, contribute to Ca^{2+} coordination

from the X-ray structure (Fig. 4; Table 2). The major difference between the predicted and experimental structures is that in the AGM the side-chain oxygen atoms from Asn^{E183} and Asp^{E191} coordinate the second-site Ca^{2+} ion, whereas in the X-ray structure respective positions in the Ca^{2+} coordinating sphere are filled by water molecules. Some high-RMSD points seem to approach the abscissa, but the corresponding structures have energy more than 2 kcal/mol above the AGM (Fig. 5), which is difficult to see at the plot scale. For sites 3 and 4, the AGM positions of Ca^{2+} ions are 0.4 and 0.9 Å, respectively, from the X-ray positions (Table 2) and the local-minima energies correlate with RMSDs (Fig. 5). At site 3, three out of seven of Ca^{2+} ligands are water molecules. However, unlike in elastase and α-lactalbumin, there is no free acidic side chain at the Ca^{2+}-binding site. In the absence of explicit water molecules in the model, the protein did not provide additional Ca^{2+} ligands.

Proteinase K is an endolytic serine protease (Betzel et al. 1988). The X-ray structure at 1.5 Å resolution (Betzel et al. 1988) shows two distinct Ca^{2+}-binding sites. At the first site, four of the eight Ca^{2+}-coordinating oxygen atoms are provided by water molecules. The eight oxygen atoms are arranged in a pentagonal bipyramid with a bidentate carboxylate in an axial position. The second Ca^{2+} ion is chelated by only three protein oxygen atoms and apparently by two water molecules. It was suggested that Ca^{2+} in this site is heptacoordinated because the coordination geometry approximates a pentagonal bipyramid in which two unfilled equatorial positions (McPhalen et al. 1991) may be occupied by water molecules not presented in the X-ray structure. Indeed, the difference electron density map ($F_o - F_c$) associated with the X-ray structure has two clusters of electron density at the above-mentioned vacancies, which are not accounted for in the X-ray structure. Thus, up to five water molecules contribute to coordination of the Ca^{2+} ion in the crystal and, apparently, in vitro. It is worthy of note that Ca^{2+} binding at site 2 is so weak that attempts to titrate Ca^{2+}-free proteinase K with Ca^{2+} for this site were unsuccessful (Bajorath et al. 1988). Furthermore, the difference density map ($F_o - F_c$) shows negative density at the position of the Ca^{2+} ion in site 2, indicating that the site may have a low occupancy. In the

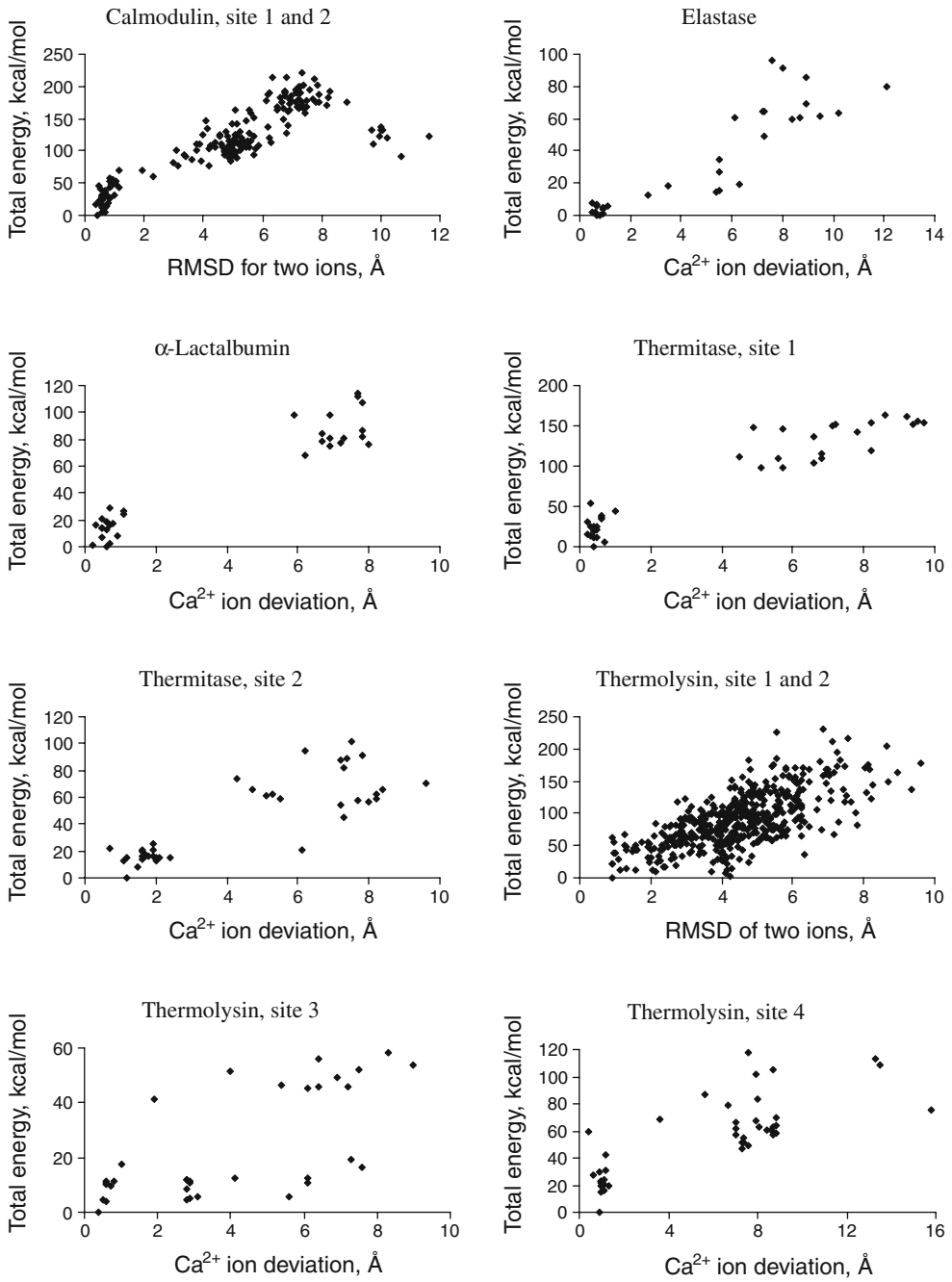

Fig. 5 Relative energy of minimum-energy complexes predicted for Ca^{2+}-binding proteins correlates with RMSD of Ca^{2+} ions from the corresponding X-ray structures. For sites 1 and 2 of thermolysin, a few low-energy but high-RMSD local minima were found. In the X-ray structure, these Ca^{2+}-binding sites are on the protein surface and a nearby acidic residue does not bind the Ca^{2+} ion

absence of explicit water molecules in our models, too few Ca^{2+} ligands were available at the first and second Ca^{2+}-binding sites (Fig. 4). Predicted positions of Ca^{2+} ions are 2.6 and 5.3 Å from respective X-ray positions. The Ca^{2+}-binding sites are at the water-exposed surface of proteinase K where bulk water molecules bind Ca^{2+} ions (Fig. 4). In these sites the implicit-solvent model failed to simulate the Ca^{2+}-coordinating water molecules.

Ca^{2+} coordination patterns in the experimental and modeled structures

Table 2 shows how experimental and modeled Ca^{2+} coordination patterns deviate from the idealized ones, which are, typically, pentagonal bipyramid for heptacoordinated Ca^{2+} and octahedron for hexacoordinated Ca^{2}. Ca^{2+} coordination patterns match the pentagonal

bipyramid in all the X-ray structures except those of elastase and thermolysin site 2, in which they match the octahedron. The modeled and experimental patterns match in nine cases, but mismatch in three cases, including sites 3 and 4 of thermolysin, which contain 2 and 3 water molecules, respectively. In the absence of explicit water molecules in the models, additional oxygen atoms from the protein filled vacancies in the Ca^{2+} coordination shells. In site 1 of Ca^{2+}−ATPase, the predicted positions of Ca^{2+} ligands RMS-deviate from the ideal octahedron and pentagonal-bipyramid by 0.61 and 0.70 Å respectively. These values are comparable with deviations of some X-ray structures from the idealized polygons (Table 2).

Sensitivity of models to backbone deformations

Our motivation for developing the current Ca^{2+} docking procedure was its application for predicting Ca^{2+} binding in homology models of Ca^{2+} channels. Because homology models are less precise than X-ray structures, it is important to know whether the docking procedure remains predictive in models with distorted backbone geometry. Ca^{2+}–ATPase resembles Ca^{2+} channels in some functional and structural aspects and we explored how results of docking two Ca^{2+} ions in the ATPase are sensitive to the backbone deviation from the X-ray structure. We modified the X-ray structure of Ca^{2+}–ATPase by randomly shifting *x*, *y*, and *z*-coordinates of all C^{α} atoms in the entire protein up to 1 Å from the crystallographic values. Then we MC-minimized the entire protein in the absence of Ca^{2+} ions until 1,000 consecutive energy minimizations did not improve the AGM. At the next step, we created a double-shell model of the distorted protein around Ca^{2+} ions using their X-ray coordinates. The RMS deviations of C^{α} atoms in the distorted model from the X-ray structure were 1.1 and 0.8 Å for residues in the fixed and flexible shells, respectively.

The Ca^{2+}-docking procedure was then applied to the distorted model. Importantly, in this experiment, C^{α} atoms were pinned to respective positions in the distorted model rather than to the X-ray structure. The RMSD of Ca^{2+} ions in the distorted-model AGM from the X-ray structure is as small as 0.6 Å. RMSDs of all heavy atoms in the 5 and 10 Å shells are 1.1 and 1.3 Å, respectively. RMSDs of side-chain heavy atoms in the 5 and 10 Å shell are 1.6 and 1.7 Å, respectively. Relative energies of minimum-energy structures still correlate with the RMSDs of Ca^{2+} atoms from their X-ray positions (Fig. 6). However, unlike in the non-distorted Ca^{2+}–ATPase (Fig. 3d), some local minima with large RMSDs of 6–8 Å from the X-ray structure have small relative energies of 5–25 kcal/mol. Attempts to dock Ca^{2+} ions in models in which C^{α} atoms were shifted more than 1.5 Å from the X-ray structures still produced many minimum-energy complexes resembling the X-ray

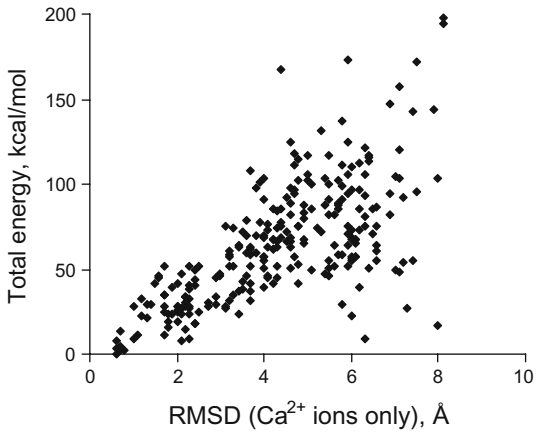

Fig. 6 Docking of Ca^{2+} ions in the Ca^{2+}–ATPase model with C^{α} carbon atoms shifted up to 1.7 Å from the X-ray structure. Energies of minimum-energy structures still correlate with RMSDs of the Ca^{2+} ions from the X-ray structure

structure, but their relative energy did not correlate with RMSD from the X-ray structure. We also attempted to dock Ca^{2+} ions to the N-terminus of apo-calmodulin (Houdusse et al. 2006) in which alpha carbon atoms RMS-deviate from the Ca^{2+}-bound calmodulin as much as 5.7 Å. Pin constraints were not used in these calculations and no constraints were used to impose formation of Ca^{2+} coordination bonds. Unsurprisingly, the unbiased docking of Ca^{2+} ions in the apo-protein was unable to induce large-scale movement of the protein backbone.

Thus, the Ca^{2+} docking procedure successfully predicted twelve Ca^{2+}-binding sites in six proteins, which represent intracellular, extracellular, and transmembrane proteins that bind Ca^{2+} ions with different affinity and different stoichiometry (Figs. 1, 2; Table 2). Furthermore, correct positions of Ca^{2+} ions and side-chain torsions were predicted even for a moderately distorted model of a Ca^{2+}-binding protein. To the best of our knowledge this is the first study in which docking of Ca^{2+} ions in proteins with completely flexible side chains and deformable backbones reproduced X-ray structures. For comparison, predictions of Ca^{2+} positions in rigid atomic-scale structures achieved success rates between 80 and 97% (Deng et al. 2006; Nayal and Di Cera 1994; Schymkowitz et al. 2005; Yamashita et al. 1990).

Discussion

Important roles of Ca^{2+}-binding proteins in biochemistry and cell physiology motivated theoretical studies aimed at developing computational methods of recognition of Ca^{2+}-binding sites in proteins with known 3D structure. Published methods consider the proteins as rigid bodies, with fixed backbones and side chains (Nayal and Di Cera 1994;

Schymkowitz et al. 2005; Wei and Altman 1998; Yamashita et al. 1990). For example, Scheraga and coauthors developed a method to distinguish the EF-hand motif from other Ca^{2+}-binding sites which employs a united-residues force field that ignores side-chain conformations (Khalili et al. 2004). Eisenberg and coworkers observed a number of complexes in which metal ions bind at centers of high hydrophobicity contrast (Yamashita et al. 1990). The authors proposed a "hydrophobicity contrast function" to detect Ca^{2+}-binding sites as a void space bounded by hydrophilic atoms within 3.5 Å of the void center and hydrophobic atoms 3.5–7.0 Å from the center (Yamashita et al. 1990). In an examining set of 14 Ca^{2+}-binding sites, ten sites were predicted as the function maximum within 2.0 Å of the experimental Ca^{2+} position. In a similar approach, a void space with three or more oxygen atoms within 3.4 Å of the void center was sought to correctly predict 58 out of 62 tested Ca^{2+}-binding sites (Nayal and Di Cera 1994). An observation that Asp and Glu side chains and amide and carbonyl groups occur frequently within 7 Å of Ca^{2+}-binding sites was used to predict Ca^{2+}-binding sites within 5 Å from the experimentally observed positions (Wei and Altman 1998). A success rate of 97% in predicting 244 Ca^{2+}-binding sites within 1 Å of X-ray positions was achieved by creating a training-set database of "clouds" representing distribution of Ca^{2+} ions around electronegative atoms and identifying Ca^{2+} ions at the overlapping clouds (Schymkowitz et al. 2005).

Importantly, the above methods rely on high-resolution crystallographic data and use completely rigid proteins. Such methods are less useful for predicting Ca^{2+}-binding sites in homology models and in low-resolution NMR or electron microscopy structures. Even when a high-resolution X-ray structure of a protein is available, experimental determination of Ca^{2+}-binding sites with low affinities can be problematic. For example, Ca^{2+} ions are known to be essential for the function of metabotropic glutamate receptors (Francesconi and Duvoisin 2004), but no Ca^{2+}-binding site is seen in the X-ray structures of the extracellular domains of these receptors (Kunishima et al. 2000; Tsuchiya et al. 2002). Furthermore, simulations of conformational transitions in response to Ca^{2+} binding should consider proteins with flexible backbones and side chains. Unlike the above-mentioned methods, the method proposed in this study does not assume any prior knowledge of side-chain conformations at the Ca^{2+}-binding sites.

The small size of Ca^{2+} ions, their strong interaction with ligand atoms, and the large structural diversity of Ca^{2+}-binding sites suggest that Ca^{2+} binding is governed by the local environment of Ca^{2+}-binding sites rather than the global properties of the proteins. Furthermore, the first coordination sphere of Ca^{2+} ions in proteins contains predominantly oxygen atoms, implying that favorable electrostatic attractions between Ca^{2+} and oxygen atoms and their unfavorable dehydration provide significant contributions to the energy of ion–protein interactions. Despite the simplicity of this concept, we are not aware of previous attempts to predict Ca^{2+}-binding sites in flexible proteins by searching the lowest-energy structures with the help of physics-based force fields. We see two potential reasons for this.

The first is the absence of a generally accepted force field which would take into consideration, among other aspects, polarizability of electronegative atoms in the presence of divalent cations. We did not attempt to take this challenging task and just used a classical force field with electrostatic and dehydration components tuned to reproduce X-ray structures of Ca^{2+} complexes with EDTA and a Ca^{2+}-binding protein. By no means can force-field parameters tuned using just two X-ray structures be regarded as a universal approach for modeling Ca^{2+}-binding proteins. However, our method reasonably predicts the side-chain geometry of Ca^{2+}-binding sites, which are not exposed to the bulk water, and demonstrates a correlation between the energy of local minima and their RMSDs from the X-ray structure. This suggests that a rather simple force field may be used as a starting point for developing a more universal method of docking divalent cations in flexible proteins.

The other reason is the large computational cost of solving the multi-minima problem of Ca^{2+} docking in flexible proteins. Indeed, the number of local minima in the highly ragged energy landscape is large, and prediction of the correct geometry of a Ca^{2+}–protein complex requires global energy minimization. We addressed this problem with the multi-MCM method. To reduce the computational cost, we docked Ca^{2+} ions in double-shell models of the proteins. This approximation did not significantly affect the accuracy of predictions, because residues within 7 Å of Ca^{2+} ions are most important for Ca^{2+} binding (Wei et al. 1999).

We tested our Ca^{2+}-docking method with 14 Ca^{2+}-binding sites in a selection of intracellular, extracellular, and transmembrane proteins. Our motivation for developing the Ca^{2+}-docking procedure was to use it for theoretical studies of the selectivity-filter region of Ca^{2+} channels. Therefore, we put special emphasis on $Ca^{2+}-$ATPase, a transmembrane protein in which side-by-side disposition of Ca^{2+} ions resembles that in models of Ca^{2+} channels (Zhorov and Ananthanarayanan 1996; Zhorov et al. 2001). We considered Ca^{2+}-binding sites either without water molecules in the first coordination shell or with a small number of water molecules in Ca^{2+}-binding sites according to X-ray structures. Because we used an implicit-solvent method (Lazaridis and Karplus 1999), some Ca^{2+}-coordinating water molecules, which are seen in X-ray structures,

were replaced in the models by additional oxygen atoms from the protein. Modeling proteins with explicit water molecules by the MCM method would have resulted in prohibitive computational cost.

For most of the tested structures, the multi-MCM method predicted a large number of minimum-energy complexes. Most importantly, the energy of these complexes correlates with the deviation of Ca^{2+} ions from the crystallographic positions (Figs. 2d, 4). This correlation indicates that combination of the AMBER force field, implicit solvent, and the solvent exposure-dependent and distance-dependent dielectric function is suitable for predicting Ca^{2+} positions in proteins. The fact that we successfully predicted Ca^{2+} coordination for most of the tested proteins supports the abovementioned simple concept of Ca^{2+} coordination in proteins. The essence of this concept is maximization of electrostatic attractions and minimization of dehydration cost with restrictions on interatomic distances, which are provided by van der Waals repulsion.

Modeling limitations

The objective of this study was to elaborate a computational procedure capable of predicting geometry of Ca^{2+}-binding sites, but not the energy of Ca^{2+}–protein complexes. The latter objective is hardly achievable with classical force fields that do not consider polarization of oxygen atoms, which may be strong in the presence of Ca^{2+} ions. Our results show that energies of Ca^{2+}–protein minimum-energy complexes do correlate with the RMSDs of the complexes from the X-ray structures. However, absolute values of these energies are unrealistic and the energy should be considered just as a scoring function. The scoring function enabled prediction of positions of a given number of Ca^{2+} ions and side-chain conformations in a potential Ca^{2+}-binding site, but neither the number of ions in the binding site nor the selectivity of the site to a particular ion. Also, our method was unable to correctly predict Ca^{2+} coordination in models where backbones were substantially distorted compared with the respective X-ray structures. Despite these limitations, the procedure may be useful for various applications including homology modeling, simulation of local conformational transitions upon Ca^{2+} binding, and predicting Ca^{2+} binding in experimental structures in which Ca^{2+} ions are not seen, either because of poor resolution or because of the inability of the method to detect Ca^{2+} ions.

Acknowledgments We thank Denis B. Tikhonov for helpful discussions. The study was supported by the Canadian Institutes of Health Research (grant MOP-53229 to BSZ). This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: http://www.sharcnet.ca).

References

Acharya KR, Stuart DI, Walker NP, Lewis M, Phillips DC (1989) Refined structure of baboon alpha-lactalbumin at 1.7 a resolution. Comparison with C-type lysozyme. J Mol Biol 208:99–127

Åqvist J (1990) Ion−water interaction potentials derived from free energy perturbation simulations. J Phys Chem 94:8021–8024

Arriortua MI, Insausti M, Urtiaga MK, Via J, Rojo T (1992) Synthesis and structure determination of SrCa(edta).5H_2O. Acta Crystallogr C 48:779–782

Bagley SC, Altman RB (1995) Characterizing the microenvironment surrounding protein sites. Protein Sci 4:622–635

Bajorath J, Hinrichs W, Saenger W (1988) The enzymatic activity of proteinase K is controlled by calcium. Eur J Biochem 176:441–447

Barnett BL, Uchtman VA (1979) Structural investigations of calcium-binding molecules. 4. Calcium binding to aminocarboxylates. Crystal structures of Ca(CaEDTA)·7H_2O and Na(CaNTA). Inorg Chem 18:2674–2678

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Betzel C, Pal GP, Saenger W (1988) Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1.5 A resolution. Acta Crystallogr B 44(Pt 2):163–172

Bruhova I, Zhorov BS (2007) Monte Carlo-energy minimization of correolide in the Kv1. 3 channel: possible role of potassium ion in ligand-receptor interactions. BMC Struct Biol 7:5

Chattopadhyaya R, Meador WE, Means AR, Quiocho FA (1992) Calmodulin structure refined at 1.7 A resolution. J Mol Biol 228:1177–1192

Clapham DE (2007) Calcium signaling. Cell 131:1047–1058

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KMJ, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197

Deng H, Chen G, Yang W, Yang JJ (2006) Predicting calcium-binding sites in proteins - a graph theory and geometry approach. Proteins 64:34–42

Edsall JT, McKenzie HA (1978) Water and proteins. I. The significance and structure of water; its interaction with electrolytes and non-electrolytes. Adv Biophys 10:137–207

Francesconi A, Duvoisin RM (2004) Divalent cations modulate the activity of metabotropic glutamate receptors. J Neurosci Res 75:472–479

Galinska-Rakoczy A, Engel P, Xu C, Jung H, Craig R, Tobacman LS, Lehman W (2008) Structural basis for the regulation of muscle contraction by troponin and tropomyosin. J Mol Biol 379:929–935

Gros P, Fujinaga M, Dijkstra BW, Kalk KH, Hol WG (1989) Crystallographic refinement by incorporation of molecular dynamics: thermostable serine protease thermitase complexed with eglin c. Acta Crystallogr B 45(Pt 5):488–499

Harding MM (2002) Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. Acta Crystallogr D Biol Crystallogr 58:872–874

Harding MM (2006) Small revisions to predicted distances around metal sites in proteins. Acta Crystallogr D Biol Crystallogr 62:678–682

Holland DR, Tronrud DE, Pley HW, Flaherty KM, Stark W, Jansonius JN, McKay DB, Matthews BW (1992) Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. Biochemistry 31:11310–11316

Hori K, Kushick JN, Weinstein H (1988) Structural and energetic parameters of Ca^{2+} binding to peptides and proteins. Biopolymers 27:1865–1886

Houdusse A, Gaucher JF, Krementsova E, Mui S, Trybus KM, Cohen C (2006) Crystal structure of apo-calmodulin bound to the first two IQ motifs of myosin V reveals essential recognition features. Proc Natl Acad Sci USA 103:19326–19331

Huang Y, Zhou Y, Yang W, Butters R, Lee HW, Li S, Castiblanco A, Brown EM, Yang JJ (2007) Identification and dissection of Ca(2+)-binding sites in the extracellular domain of Ca(2+)-sensing receptor. J Biol Chem 282:19000–19010

Khalili M, Saunders JA, Liwo A, Oldziej S, Scheraga HA (2004) A united residue force-field for calcium-protein interactions. Protein Sci 13:2725–2735

Kunishima N, Shimada Y, Tsuji Y, Sato T, Yamamoto M, Kumasaka T, Nakanishi S, Jingami H, Morikawa K (2000) Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. Nature 407:971–977

Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins 35:133–152

Li Z, Scheraga HA (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc Natl Acad Sci USA 84:6611–6615

McPhalen CA, Strynadka NC, James MN (1991) Calcium-binding sites in proteins: a structural perspective. Adv Protein Chem 42:77–144

Meiler J, Baker D (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins 65:538–548

Meot-Ner M (1987) Heats of hydration of organic ions: predictive relations and analysis of solvation factors based on ion clustering. J Phys Chem 91:417–426

Meyer E, Cole G, Radhakrishnan R, Epp O (1988) Structure of native porcine pancreatic elastase at 1.65 A resolution. Acta Crystallogr B 44:26–38

Nayal M, Di Cera E (1994) Predicting Ca(2+)-binding sites in proteins. Proc Natl Acad Sci USA 91:817–821

Pidcock E, Moore GR (2001) Structural characteristics of protein binding sites for calcium and lanthanide ions. J Biol Inorg Chem 6:479–489

Prasad A, Pedigo S (2005) Calcium-dependent stability studies of domains 1 and 2 of epithelial cadherin. Biochemistry 44:13692–13701

Schymkowitz JW, Rousseau F, Martins IC, Ferkinghoff-Borg J, Stricher F, Serrano L (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc Natl Acad Sci USA 102:10147–10152

Singh N, Somvanshi RK, Sharma S, Dey S, Kaur P, Singh TP (2007) Structural elements of ligand recognition site in secretory phospho-lipase A2 and structure-based design of specific inhibitors. Curr Top Med Chem 7:757–764

Sodhi JS, Bryson K, McGuffin LJ, Ward JJ, Wernisch L, Jones DT (2004) Predicting metal-binding site residues in low-resolution structural models. J Mol Biol 342:307–320

Suarez MC, Rocha CB, Sorenson MM, Silva JL, Foguel D (2008) Free-energy linkage between folding and calcium binding in EF-hand proteins. Biophys J 95:4820–4828

Szmola R, Sahin-Toth M (2007) Chymotrypsin C (caldecrin) promotes degradation of human cationic trypsin: identity with Rinderknecht's enzyme Y. Proc Natl Acad Sci USA 104:11227–11232

Tikhonov D, Zhorov BS (2007) Sodium channels: ionic model of slow inactivation and state-dependent drug binding. Biophys J 93:1557–1570

Tikhonov DB, Zhorov BS (2008) Molecular modeling of benzothiazepine binding in the L-type calcium channel. J Biol Chem 283:17594–17604

Toyoshima C, Nakasako M, Nomura H, Ogawa H (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. Nature 405:647–655

Tsuchiya D, Kunishima N, Kamiya N, Jingami H, Morikawa K (2002) Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd^{3+}. Proc Natl Acad Sci USA 99:2660–2665

Veltman OR, Vriend G, Berendsen HJ, Van den Burg B, Venema G, Eijsink VG (1998) A single calcium binding site is crucial for the calcium-dependent thermal stability of thermolysin-like proteases. Biochemistry 37:5312–5319

Vyas NK, Vyas MN, Quiocho FA (1988) Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. Science 242:1290–1295

Wei L, Altman RB (1998) Recognizing protein binding sites using statistical descriptions of their 3D environments. Pac Symp Biocomput 3:497–508

Wei L, Huang ES, Altman RB (1999) Are predicted structures good enough to preserve functional sites? Structure 7:643–650

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 106:765–784

Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252

Wu X, Reid RE (1997) Conservative D133E mutation of calmodulin site IV drastically alters calcium binding and phosphodiesterase regulation. Biochemistry 36:3608–3616

Yamashita MM, Wesson L, Eisenman G, Eisenberg D (1990) Where metal ions bind in proteins. Proc Natl Acad Sci USA 87:5648–5652

Yang W, Lee HW, Hellinga H, Yang JJ (2002) Structural analysis, identification, and design of calcium-binding sites in proteins. Proteins 47:344–356

Yang W, Jones LM, Isley L, Ye Y, Lee HW, Wilkins A, Liu ZR, Hellinga HW, Malchow R, Ghazi M, Yang JJ (2003) Rational design of a calcium-binding protein. J Am Chem Soc 125:6165–6171

Yang W, Wilkins AL, Ye Y, Liu ZR, Li SY, Urbauer JL, Hellinga HW, Kearney A, van der Merwe PA, Yang JJ (2005) Design of a calcium-binding protein with desired structure in a cell adhesion molecule. J Am Chem Soc 127:2085–2093

Yasuhiko S, Setsuko N (1991) Parametrization of calcium binding site in proteins and molecular dynamics simulation on phospholipase A2. J Comput Chem 12:717–730

Zhorov BS (1981) Vector method for calculating derivatives of energy of atom-atom interactions of complex molecules according to generalized coordinates. J Struct Chem 22:4–8

Zhorov BS (1983) Vector method for calculating derivatives of the energy deformation of valence angles and torsion energy of complex molecules according to generalized coordinates. J Struct Chem 23:649–655

Zhorov BS, Ananthanarayanan VS (1996) Structural model of a synthetic Ca^{2+} channel with bound Ca^{2+} ions and dihydropyridine ligand. Biophys J 70:22–37

Zhorov BS, Lin SX (2000) Monte Carlo-minimized energy profile of estradiol in the ligand-binding tunnel of 17 beta-hydroxysteroid dehydrogenase: atomic mechanisms of steroid recognition. Proteins 38:414–427

Zhorov BS, Folkman EV, Ananthanarayanan VS (2001) Homology model of dihydropyridine receptor: implications for L-type Ca(2+) channel modulation by agonists and antagonists. Arch Biochem Biophys 393:22–41