J Comput Aided Mol Des (2011) 25:181–194
DOI 10.1007/s10822-010-9409-6

Structural-dynamical investigation of the ZnuA histidine-rich loop: involvement in zinc management and transport

Mattia Falconi · Francesco Oteri · Francesco Di Palma · Saurabh Pandey · Andrea Battistoni · Alessandro Desideri

Received: 8 September 2010 / Accepted: 30 December 2010 / Published online: 15 January 2011

Abstract Comparative homology modelling techniques have been used to model the protein ZnuA from *Salmonella enterica* serovar Typhimurium using the 3D structure of the homologous protein from *Escherichia coli*. These two-domain proteins bind one Zn^{2+} atom, with high affinity, in the inter-domain cleft and possess a histidine-rich loop in the N-terminal domain. Alternative structures of the ZnuA histidine-rich loop, never resolved by the X-ray diffraction method, have been modelled. A model of the apo form, one with the histidine-rich loop deleted and two alternative structures with a second zinc ion bound to the histidine-rich loop, have been generated. In all the modelled proteins, investigated through molecular dynamics simulation, the histidine-rich loop is highly mobile and its fluctuations are correlated to the ligand stability observed in the zinc sites. Based on the plasticity of the histidine-rich loop and its significant effects on protein mobility a possible role in the capture and/or transfer of the zinc ions has been suggested.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9409-6) contains supplementary material, which is available to authorized users.

M. Falconi (✉) · F. Oteri · F. Di Palma · S. Pandey · A. Battistoni · A. Desideri
Department of Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy
e-mail: falconi@uniroma2.it

M. Falconi · F. Oteri · A. Desideri
CIBB, Center of Biostatistics and Bioinformatics, Via della Ricerca Scientifica, 00133 Rome, Italy

M. Falconi · A. Desideri
Interuniversity Consortium, National Institute Biostructure and Biosystem (INBB), Rome, Italy

Keywords ZnuA · ABC transporter family · Zinc transport · Molecular dynamics simulation · Histidine-rich loop

Introduction

The efficient uptake of transition metals plays an important role in the host-pathogen interaction. Recent studies have established that zinc, which is the most abundantly used transition metal in enzymes and is contained in about 5% of the bacterial proteins [1], is not freely available within the infected host [2]. Gram-negative bacteria respond to Zn deficiency by activating a specific uptake pathway based on the ZnuABC transporter that enables bacteria to recruit Zn^{2+} with high efficiency [3]. This bacterial Zn^{2+} uptake system is unique to microorganisms, is vital for maintaining ion homeostasis in several pathogenic and non-pathogenic bacteria, and plays a pivotal role in determining bacterial resistance to the host defense mechanisms.

The high-affinity zinc transporter ZnuABC, like all ATP-binding cassette-type (ABC-type) protein family [4, 5], is composed of three proteins: ZnuB, the membrane permease, ZnuC, the ATPase component and ZnuA, the soluble periplasmic metal-binding protein which captures Zn^{2+} and delivers it to ZnuB. The structures of ZnuA from *Escherichia coli* [6–8] and *Synechocystis* sp. 6803 [9, 10] have revealed that the Zn^{2+} binding site and its environment are highly conserved among different species including the Gram-negative pathogen *Salmonella enterica* serovar Typhimurium (*Se*-ZnuA). On the basis of ligand identity and sequence homology, ZnuA belongs to cluster 9 of the extended family of periplasmic ligand binding proteins (PLBPs), which includes metal-binding receptors (MBRs) specific for Zn^{2+} or Mn^{2+} [11]. The proteins of

this cluster are characterized by a two-domain structure (N- and C-terminal domain), connected by a "backbone" α-helix, in the Zn^{2+}-specific proteins, and by a flexible hinge region in the Mn^{2+} binding ones.

The overall structure of the ZnuA Zn^{2+}-specific protein consists of two $(\beta/\alpha)_4$ domains with the same topology that are connected by a long backbone α-helix with a metal ion bound in a high affinity binding site (primary zinc site) located at the inter-domain interface [6] (*Se*-ZnuA model in Fig. 1). Each domain consists of a parallel four-stranded β-sheet with strand sequence 2-1-3-4. The central β-sheet is flanked by four α-helices, two on each side, with the α-helices alternating with the β-strands in the primary sequence. The overall structures of both sub-domains are similar but there are significant differences in the loop regions contributing to the metal-binding site, as well as in the C-terminal helices that interact with the backbone helix. The backbone helix α5 connects the C-terminal helix α4 of the first domain with the N-terminal strand β5 of the second domain and stabilizes the relative orientation of the two globular domains by close interactions with the longer C-terminal helices α4 and α9 of both domains [6]. The zinc proteins contain in the N-terminal domain a charged flexible loop, rich in histidines and acidic residues (His-rich loop), whose sequence shows significant species-specific differences [7, 9]. Different studies have established that the His-rich loop contributes to the formation of a secondary zinc binding site, which has been proposed to be important in the acquisition of periplasmic Zn^{2+} [10–13]

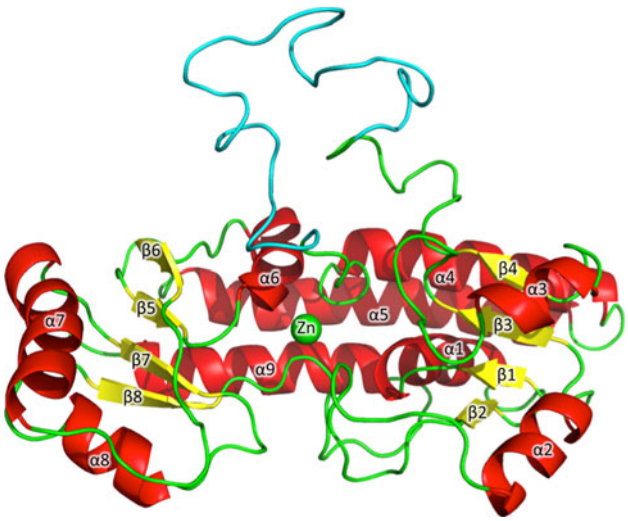

Fig. 1 Cartoon diagram of *Se*-ZnuA Loop1 model: the α-helices are indicated by *red spirals* while the β-strand are represented by *yellow arrows*. Random coil and turns are shown as a *green wire* and the zinc metal is represented by a *green sphere*. The His-rich loop, not detected by the X-ray analysis and added to the model (Met115-Tyr144), is shown as a *cyan wire*. The labels indicate the name of each secondary structure segment as assigned in *Eco*-ZnuA by Yatsunyk and coworkers [8]

for its delivery to ZnuB [7] or for regulation of zinc uptake [9]. Recent studies carried out in *Salmonella enterica* have shed some light on the role of the His-rich loop [14]. In fact, it has been shown that a mutant form of ZnuA lacking this protein region is still able to mediate zinc import. However, the activity of this mutant ZnuA is critically dependent on the presence of an additional periplasmic zinc-binding protein, ZinT, which can form a stable complex with ZnuA and contributes to zinc uptake. At the same time, the presence of ZinT is dispensable for metal import in bacteria expressing wild type ZnuA [14]. Collectively, these results suggest that the His-rich loop of ZnuA and ZinT are two independent structural elements employed to recruit zinc in the periplasmic space under severe metal starvation [14].

Based on the available X-ray structures of ZnuA from *E. coli* [6–8] and *Synechocystis* ZnuA [9, 10] the primary site, binding Zn^{2+} with nanomolar affinity, is located in the inter-domain cleft. The metal ligands are three highly conserved histidines, whereas a fourth ligand is a water molecule or, as in the *E. coli* structure, a glutamic acid residue [8]. A second zinc ion has been identified in a ZnuA structure from *E. coli* [8], located at about 12 Å from the primary binding site, bound on the external surface of the protein by an histidine residue, His228, and other unidentified residues from the His-rich loop [8]. The crystal structures also suggest that upon Zn^{2+} binding the N- and C-terminal domains undergo a ligand dependent conformational change that allows the interaction with the transmembrane transporter ZnuB, but, to date, structural studies on ZnuB and on this specific interaction are lacking [8].

In this paper the structural features of the ZnuA His-rich loop have been investigated building six different homology models (with or without Zn^{2+} and with one or two fully occupied metal binding sites) using as a template the 3D structure of the homologous protein from *E. coli* (*Eco*-ZnuA) [8]. The models have been investigated by molecular dynamics (MD) simulations to analyze the possible conformations/functions of the His-rich loop whose structure has never been detected by X-ray diffraction studies [6–8, 10] and thereby to obtain insights about a possible role of this loop in modulating the stability of the primary zinc site. The results show the plasticity of the zinc sites and suggest the residues that can be involved in the zinc management process.

Methods

Se-ZnuA models

Six different models, based on the *Eco*-ZnuA structure [8], have been generated to take into account different

Table 1 Web servers and programs used to generate the *Se*-ZnuA models

Models	Protein scaffold	His-rich loop	Zinc sites
Loop1	SWISS-MODEL	3D-JIGSAW	Primary site ligands: Glu59, His60, His147, His211 *SYBYL*
Loop2	SWISS-MODEL	CPHmodels	Primary site ligands: Glu59, His60, His147, His211 *SYBYL*
Apo	SWISS-MODEL	3D-JIGSAW	no zinc sites
dLoop	SWISS-MODEL	Asp96–His118 deleted SYBYL	Primary site ligands: Glu59, His60, His147, His211 *SYBYL*
DZn1	SWISS-MODEL	3D-JIGSAW	Primary site ligands: Glu59, His60, His147, His211 *SYBYL* Secondary site ligands: Glu132, Asp136, His141, His228 *PyMol sculpting*
DZn2	SWISS-MODEL	3D-JIGSAW	Primary site ligands: Glu59, His60, His147, His211 *SYBYL* Secondary site ligands: Glu124, Glu133, His139, His228 *PyMol sculpting*

conformational states/conditions of the *Se*-ZnuA protein, i.e. from the Apo to the fully metallated form with different His-rich loop conformations (see Table 1, Fig. 2):

- the Loop1 and Loop2 models, both containing one zinc ion in the primary site bound to Glu59, His60, His147 and His211, representing the zinc site described by the X-ray analysis [8] (Fig. 2), have been modelled with two different His-rich loop conformations;
- the Apo model, was modelled using the Loop1 model without the zinc ion (Fig. 2);
- the dLoop model was modelled using the Loop1 model with the His-rich loop deleted (from Asp119 to His141) (Fig. 2);
- the DZn1 (Double Zn 1) and DZn2 (Double Zn 2) models, each containing two metal ions in two different zinc sites, have been modelled using the Loop1 model as a starting structure. In these models the primary zinc ion maintains the ligands used for Loop1 and Loop2 (Figs. 2, 3), while the secondary zinc ion was bound to residues Glu132, Asp136, His141 and His228 in the DZn1 model and to residues Glu124, Glu133, His139 and His228 in the DZn2 model (Fig. 3, DZn1 and DZn2). In both cases His228 is maintained as a second zinc ligand based on the Yatsunyk and coworkers experimental evidence [8].

Homology modelling procedure

The comparative homology modelling technique implemented in the SWISS-MODEL server [15] has been used to generate the scaffold model, i.e. the *Se*-ZnuA protein without the His-rich loop, using the 3D structure of the homologous protein *Eco*-ZnuA showing a sequence identity of about 84% [8] (PDB access code 2PS3, chain A). Since in the PDB file 2PS3 the His-rich loop residues are missing, the protein segment Met115-Tyr144 has been modelled in two alternative conformations through the 3D-JIGSAW v. 2.0 [16] and CPHmodels v. 3.0 [17] servers and incorporated in the Loop1 and Loop2 models. The His-rich loop is long enough and "topologically" independent from the ZnuA scaffold to be considered as a separate item, or a mobile domain, of undefined structure. The initial conformations of the His-rich loop have been casually chosen to verify if they are able to assume a defined conformation using the conformational exploration capabilities of the isothermal MD simulations. In a recent paper the use of isothermal simulations applied to the models obtained has been indicated as the best method to obtain a meaningful description of loop mobility and the definition of best possible model with the lowest energy loop structures [18]. The Zn cation was included in the primary zinc site (Fig. 3, Loop1) and the conformations of the ligand residues in the primary zinc site (Glu59, His60, His147, His211) were manually adjusted using SYBYL 6.0 (Tripos Inc. 1699, South Hanley Road St. Louis, Missouri, 63144, USA). In the metal-free ZnuA model (Apo) the binding site has been left empty. The secondary zinc sites, i.e. the zinc site partially detected by the X-ray analysis [8], were shaped using the *Sculpting* utility implemented in PyMol [19]. All the structural illustrations (Figs. 1, 2, 3, 7) have been produced using the program PyMol [19].

The structure of the protein with the deleted loop (dLoop) has been obtained deleting residues from Asp119 to His141 and connecting Ala118 to Gly142 in the Loop1 model according to Wei and coworkers [9].

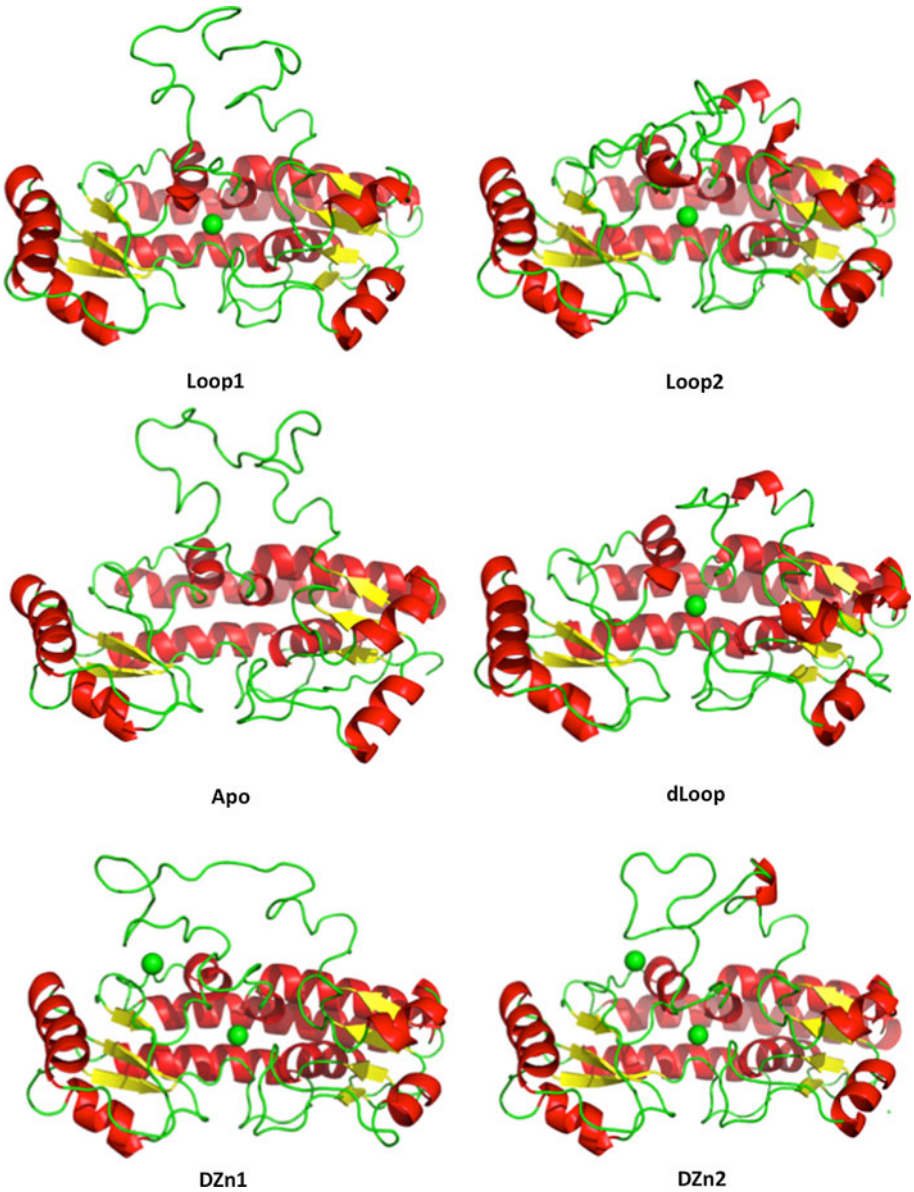

Fig. 2 Cartoon diagram of the six *Se*-ZnuA models: the α-helices are indicated by *red spirals* while the β-strand are represented by *yellow arrows*. Random coil and turns are shown as a *green wire* and the zinc metals are represented by *green spheres*

The conformation of the region near the deletion was adjusted using the *Protein Loop Search* utility implemented in the SYBYL 6.0 Biopolymer module (Tripos Inc. 1699, South Hanley Road St. Louis, Missouri, 63144, USA).

The structures of all the models have been regularized minimizing the models for 500 steps using the AMBER force field with the Powell Minimization Algorithm as implemented in SYBYL 6.0 (Tripos Inc. 1699, South Hanley Road St. Louis, Missouri, 63144, USA).

ZnuA models regularization and validation procedure

The molecular dynamics simulation approach has been used to slowly regularize the ZnuA structures that, due to the structural impositions carried out during the building of the models, show some stereochemical problems. The quality of the ZnuA models has been checked using the PROCHECK program [20] and the MolProbity server [21, 22] (http://molprobity.biochem.duke.edu/). PROCHECK is applied to check the phi and psi torsion angles combinations with respect to available zones of the Ramachandran plot [20], while MolProbity is a structure-validation web service based on the sensitivity provided by optimized hydrogen placement and all-atom contact analysis, complemented by updated versions of covalent-geometry and torsion-angle criteria [21, 22].

Zinc sites parametrization

The protein topology and the coordinates files, used as input for the MD program NAMD 2.7b1 [23], have been obtained with the tLeap module using the AmberTools 1.2

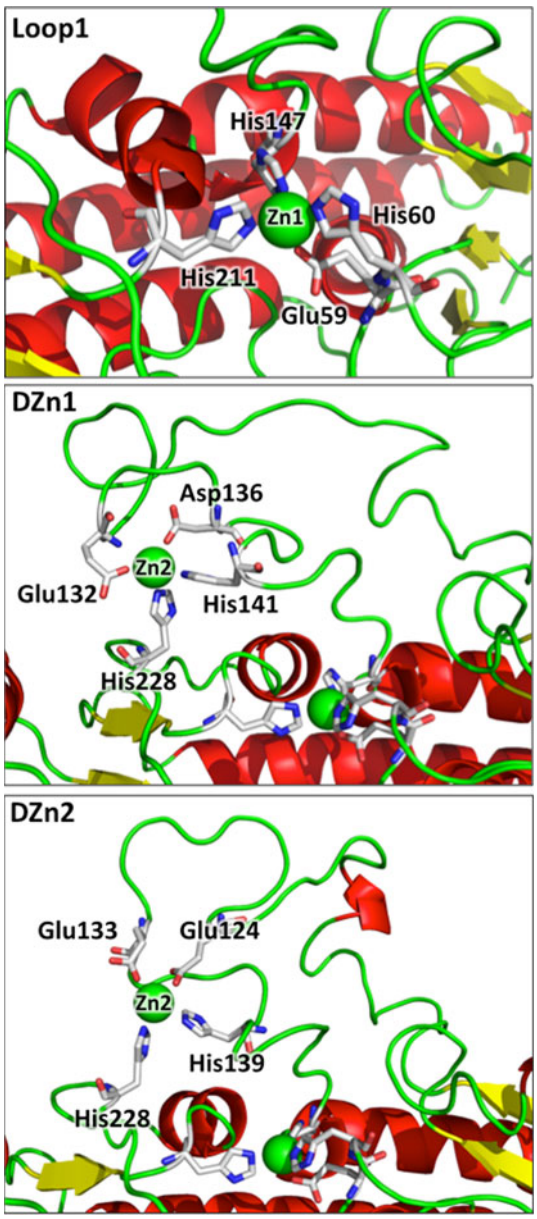

Fig. 3 Close view of the primary zinc site in Loop1 and of the secondary zinc site in DZn1 and DZn2 *Se*-ZnuA models. The α-helices are indicated by *red spirals* while the β-strand are represented by *yellow arrows*. Random coil and turns are shown as a *green wire* and the zinc metals are represented by *green spheres*. The zinc ligands are evidenced by sticks models

program suite [24]. The structures have been modelled using the AMBER03 force field [25].

The potential to treat Zn^{2+} coordinations in proteins can be chosen following various models. The bonded model employs covalent bonds between Zn^{2+} and its ligands to maintain the Zn^{2+} coordination geometry in proteins during simulations [26]. Although bonded Zn-ligand interactions preserve the observed geometry of the metal-binding site, this model is not appropriate for ZnuA where the metal-binding sites are flexible and the Zn^{2+} ligands are believed to exchange during the metal transport [7, 9–13]. A non-bonded model employing electrostatic and van der Waals forces instead of covalent ones has been used to maintain the Zn^{2+} coordination geometry in proteins [27] but, the MD simulations using this model have given zinc geometries different from those of the X-ray structures [28], suggesting that this model doesn't appropriately describe the zinc binding site. A recent approach has been elaborated by Sakharov and coworkers [29, 30] taking into account the charge transfer from the ligands to Zn^{2+} and the electrostatic polarization of Zn^{2+} and its ligands. However this model has been exclusively implemented in the CHARMM22 force field [31] and has been used, up to now, only to describe Zn ions mainly coordinated by cysteines [29, 30].

The CaDA approach has been implemented in the AMBER03 force field [25] and has been applied to proteins containing Zn ions mainly coordinated by histidines and aspartic or glutammic acids [32–34] similar to the zinc coordination in ZnuA, rendering this model appropriate for the ZnuA simulations. In detail the CaDA parametrization replaces the zinc divalent cation (one-atom representation) with the tetrahedron-shaped zinc divalent cation, composed by 4 dummy atoms (each one representing one Zn orbital) attached to the central zinc ion, deprotonating the zinc ligands and using histidinate for histidine. These aspects make the CaDA approach particularly appropriate for the ZnuA simulations.

The binding sites in the different Zn-bound models have been initially minimized without water through the following procedure: a first 2000 steps minimization has been carried out on the tetrahedron-shaped zinc divalent cations with positional restraints of 500 kcal/mole applied to the protein, thereafter the entire binding site, including the ligand residues, has been minimized for others 2000 steps with positional restraints of 500 kcal/mole applied to the rest of the protein and finally energy minimization of 2000 steps without restraints has been carried out for the entire protein.

Classical molecular dynamics simulations

The six protein structures (Table 2) have been immersed in parallelepipedal water boxes (Table 2), using TIP3P water molecules [35], neutralizing the proteins charges with Na^{+} counterions (Table 2), placed in electrostatically preferred positions, imposing a minimal distance between the solute and the box walls of 15 Å. Before running the simulation, the water has been minimized for 2000 steps setting restraints of 500 kcal/mole on protein atoms, then a second minimization run of 2000 steps without restraints has been carried out. The minimized structures have been

Table 2 Simulated systems

Model	Loop1	Loop2	Apo	dLoop	DZn1	DZn2
Total atoms	56,427	54,033	57,070	51,753	53,795	52,172
Protein atoms	4,435	4,435	4,437	4,115	4,434	4,434
Aminoacids	291	291	291	268	291	291
Water molecules	17,326	16,528	17,540	15,879	16,449	15,908
Na^+ ions	14	14	13	5	14	14
Simulation box side X (Å)	101	73	101	100	102	102
Simulation box side Y (Å)	86	88	86	75	88	80
Simulation box side Z (Å)	77	102	77	82	79	77
Saved configurations	60,000	60,000	60,000	60,000	60,000	60,000

thermalized starting at the temperature of 0 K and increasing the temperature of 50 K, every 400 steps, until a temperature of 300 K was reached. At this point an additional 200 ps MD run, with a 1.0 fs time step, has been carried out.

The six optimized and relaxed systems have been simulated for a total of 30 ns, with a 2.0 fs time step, in periodic boundary conditions, using a cut-off radius of 12.0 Å for the non-bonded interactions, that has been smoothed after 8 Å, and updating the neighbour pair lists every 10 steps with a distance between pairs for inclusion in pair lists of 13.5 Å. The electrostatic interactions were calculated each 2 steps (4.0 fs) through the particle-mesh Ewald method [36–38], using a grid spacing of 1 Å.

The simulations were carried out at a constant temperature of 300 K, using Langevin dynamics parameters [39], with a damping coefficient of 5/ps, as coupling coefficient to be applied to all atoms. For the pressure control the Langevin piston Nose–Hoover method, a combination of the Nose–Hoover constant pressure method [40] with piston fluctuation control implemented using Langevin dynamics [41], has been used, specifying the desired pressure at 1.01325 bar (1 atm), the oscillation period at 100 fs, the decay times of the piston with the barostat damping time scale at 50 fs and the temperature of the piston at 300 K. The SHAKE algorithm [42] has been used to restrain the protein hydrogen atoms and the SETTLE algorithm [43] has been applied to restrain the water atoms. The atomic positions were saved every 250 steps (0.5 ps) and the first 3 ns were not used for the analysis.

The systems were simulated at CASPUR research center of Rome, Italy (Inter Universities Consortium for Supercomputing Applications) using 72 cores on the MATRIX cluster composed by 318 nodes, each with two AMD Opteron quadcore CPU, for a total of 2,512 cores.

Simulation analyses

The gyration radius, the distance of the binding site residues to the zinc cations and the root mean square fluctuation (RMSF) analyses, the time evolution of root-mean-square deviation (RMSD), the secondary structures analysis (do_dssp module) and the principal component analysis (PCA) [44, 45], were carried out using the GROMACS MD package version 4.0.7 program [46, 47]. The RMSFs have been calculated over the equilibrated MD trajectories removing the global translations and rotations. The salt bridges, the hydrogen bonds and the solvent accessible surface analyses were carried out using in house modified versions of g_mindist and g_hbond GROMACS modules, respectively, that have been adapted to list the atoms and the residues involved in the interactions. An in house software, using iteratively the program naccess [48], has been applied to calculate the zinc site solvent accessibility. For each ZnuA MD trajectory the NAMDenergy plugin, implemented in the VMD 1.8.7 program [49], has been used to evaluate the averages and standard deviations of the ZnuA models total energy.

Results and discussion

ZnuA models validation

The PROCHECK [20] and MolProbity [21, 22] validation procedures have been applied to the structures at three different stages namely: after the modeling (indicated with I in Table 3A, B), after the initial cycle of minimizations, i.e. before starting the MD simulation, (indicated with II in Table 3A, B) and after the first 3 ns of simulation not included in the analyses, i.e. after the complete MD thermalization procedure (indicated with III in Table 3A, B).

Table 3A reports the percentages of residues located in the most favoured regions, additional allowed regions, generously allowed regions of the Ramachandran plot, obtained through the PROCHECK program [20], for each ZnuA model at each stage of the model regularization. The percentage of residues in the favoured regions of the Ramachandran plot is a good descriptor of the stereochemical quality of a protein structure, considered "ideal" when

Table 3 Validation of ZnuA models

PROCHECK (A)

	% of residues with phi psi in most favoured regions of Ramachandran plot			% of residues with phi psi in additional allowed regions of Ramachandran plot			% of residues with phi psi in generously allowed regions of Ramachandran plot		
ZnuA model	I	II	III	I	II	III	I	II	III
Loop1	83.4	85.4	90.1	14.6	12.3	8.7	0.8	1.6	0.4
Loop2	83.4	89.3	86.6	13.4	9.9	12.3	1.6	0.4	1.2
Apo	83.4	87.7	86.6	14.6	11.5	12.3	0.8	0.4	1.2
dLoop	88.8	88.8	90.1	11.2	11.2	9.9	0	0	0
DZn1	79.1	82.2	81.4	17.8	15.0	16.2	1.2	1.6	2.0
DZn2	81.4	80.2	85.4	15.0	16.6	12.6	2.4	1.2	0.8

MOLPROBITY (B)

ZnuA model	I	II	III
Loop1	**19**	81	72
Loop2	**24**	72	68
Apo	**21**	76	88
dLoop	82	94	89
DZn1	**26**	65	69
DZn2	**28**	69	74

The larger is the % of residues with phi psi in most favoured regions of Ramachandran plot the best is the model while the larger is the % of residues with phi psi in additional allowed and generously allowed regions of Ramachandran plot the worst is the model. Values are rounded to one decimal place

The structures are well modelled if the percentile values are greater than 65%. The bold percentile values in Table 3B indicate structural problems in the models

I after the modeling procedure; *II* after the initial minimizations; *III* after 3 ns of MD simulation

90% of the residues are in the allowed regions [50]. In the first column of the Table 3A, the initial models show some structural problems mainly due to the His-rich loop that disappear after 3 ns of simulation (see column III, Table 3A). In this analysis no residues were found in dis-allowed regions.

The MolProbity absolute quality of the structures, expressed in percentile rank at each stage of the model regularization, is shown in Table 3B. The structures are well modelled if the percentile values are greater than 65% [21, 22]. In the first column of the Table 3B, the initial models show some structural problems mainly due to the His-rich loop. The stereochemical problems evidenced after the modeling procedure (Table 3A, B, stage I) gradually disappear in the structures coming from the minimization and the MD thermalization (Table 3A, B, stages II and III) indicating that all the MD analyses have been carried out on trajectories of reliable ZnuA models.

Root mean square deviations and stability of the systems

All the simulations reach a constant Root Mean Square Deviation (RMSD) value after about 3.0 ns (Fig. 4, black lines) and therefore all the analysis for the six systems have been carried out in the last 27 ns of simulation, i.e. from 3 to 30 ns. The full stability of the system is guaranteed also by the time evolution of the following parameters: total solvent accessible surface area, gyration radius, number of residues in secondary structures and total hydrogen bonds number, reported in Fig. 1s, 2s, 3s and 4s of supplementary material, respectively.

For all systems the values of the entire protein RMSD time dependence (Fig. 4, black lines) are strongly reduced when the His-rich loop is excluded from the calculation (Fig. 4, light grey lines), indicating that it represents the most mobile part of the protein (Fig. 4, dark grey lines). In agreement with this interpretation, in the dLoop trajectory the RMSD values are low due to the absence of this highly mobile region (Fig. 4, Table 4). The His-rich loop represents a highly fluctuating region, not showing preferential conformations, that is stabilized only slightly when three of its residues are used for the modelling of a secondary zinc site. Independently of the presence of the zinc ion, the RMSD values of the protein backbone are very similar, as observed in the Apo simulation (Table 4), suggesting that, at the time scale of the simulation, the protein folding is conserved also in the absence of the metal. To better

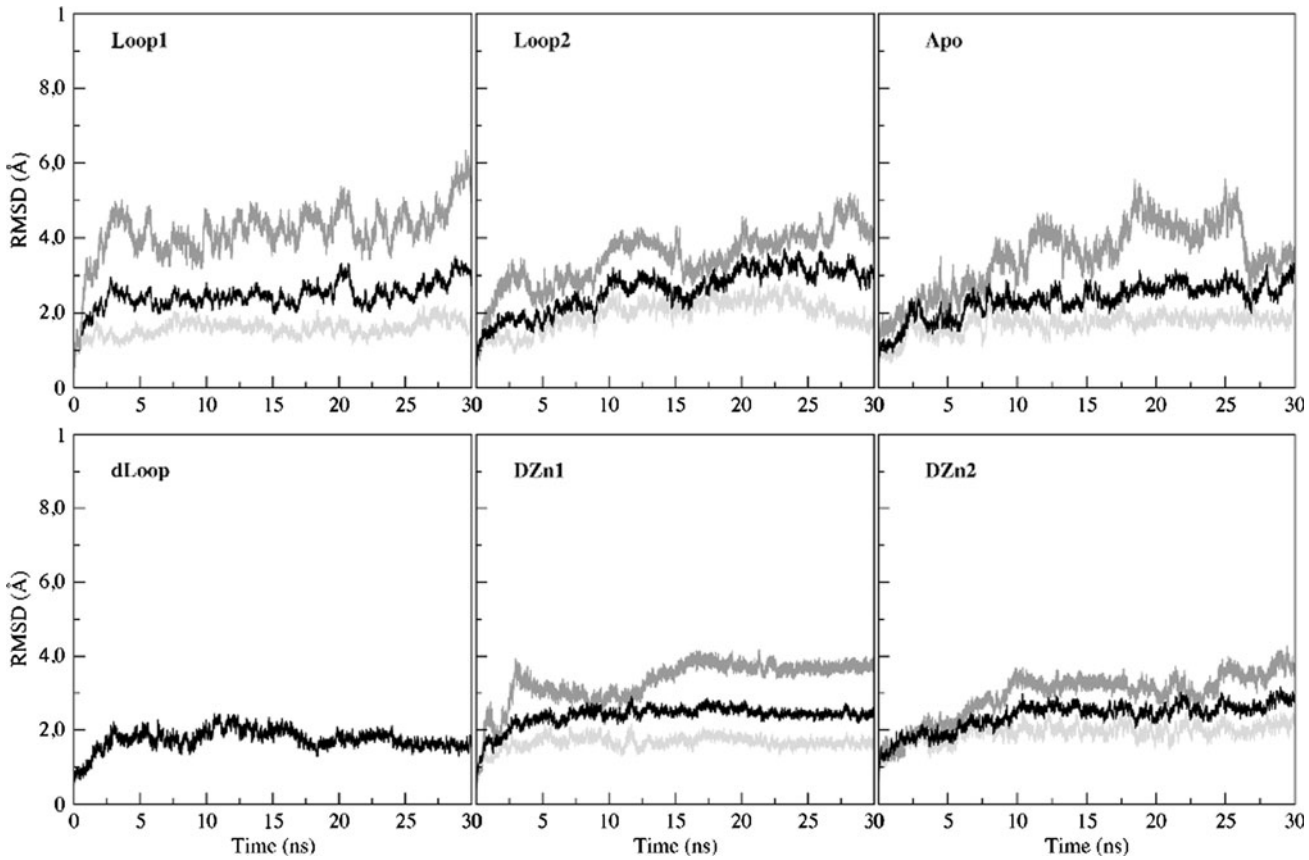

Fig. 4 Backbone RMSD from the trajectory starting structure of Loop1, Loop2, Apo, dLoop, DZn1 and DZn2 *Se*-ZnuA trajectories: the *black line* represents the RMSD of the entire protein, the *light grey line* indicates the RMSD of the protein scaffold (excluding the His-rich loop from the calculation) and the *dark grey line* shows the RMSD of the sole His-rich loop

Table 4 Backbone RMSD average values and standard deviations (Å)

System	RMSD (calculated over the entire backbone)	RMSD (calculated excluding the His-rich loop)
Loop1	2.5 ± 0.2	1.6 ± 0.2
Loop2	2.7 ± 0.5	2.0 ± 0.3
Apo	2.4 ± 0.3	1.8 ± 0.2
dLoop	1.8 ± 0.2	–
DZn1	2.5 ± 0.1	1.7 ± 0.1
DZn2	2.5 ± 0.3	2.0 ± 0.2

Table 5 MD averages of ZnuA models total energies

ZnuA model	Average total energy (kcal/mole) and standard deviation
Loop1	−715 (±119)
Loop2	−539 (±121)
Apo	−203 (±114)
dLoop	−928 (±150)
DZn1	−965 (±124)
DZn2	−1118 (±109)

quantify the energy contribute given by the presence of the zinc ions, the average values and the standard deviations of the total energy have been calculated [49] for each ZnuA simulation. The results indicate that the energies are negative and stable for all the trajectories (Table 5) but the values depends on the number of metals (see Table 5, DZn1, DZn2, Apo) and on the absence or starting conformation of the His-rich loop (see Table 5, dLoop, Loop1, Loop2).

Root mean square fluctuations

Analysis of the residue averaged Root Mean Square Fluctuations (RMSF) plots (Fig. 5), in accordance with the X-ray data [6–8], confirms that the His-rich loop is the most fluctuating region of the *Se*-ZnuA protein structure. This loop is highly mobile in the Loop1, Loop2 and Apo systems indicating that the presence of a single zinc ion in the primary site does not reduce the loop fluctuations. In both the DZn1 and DZn2

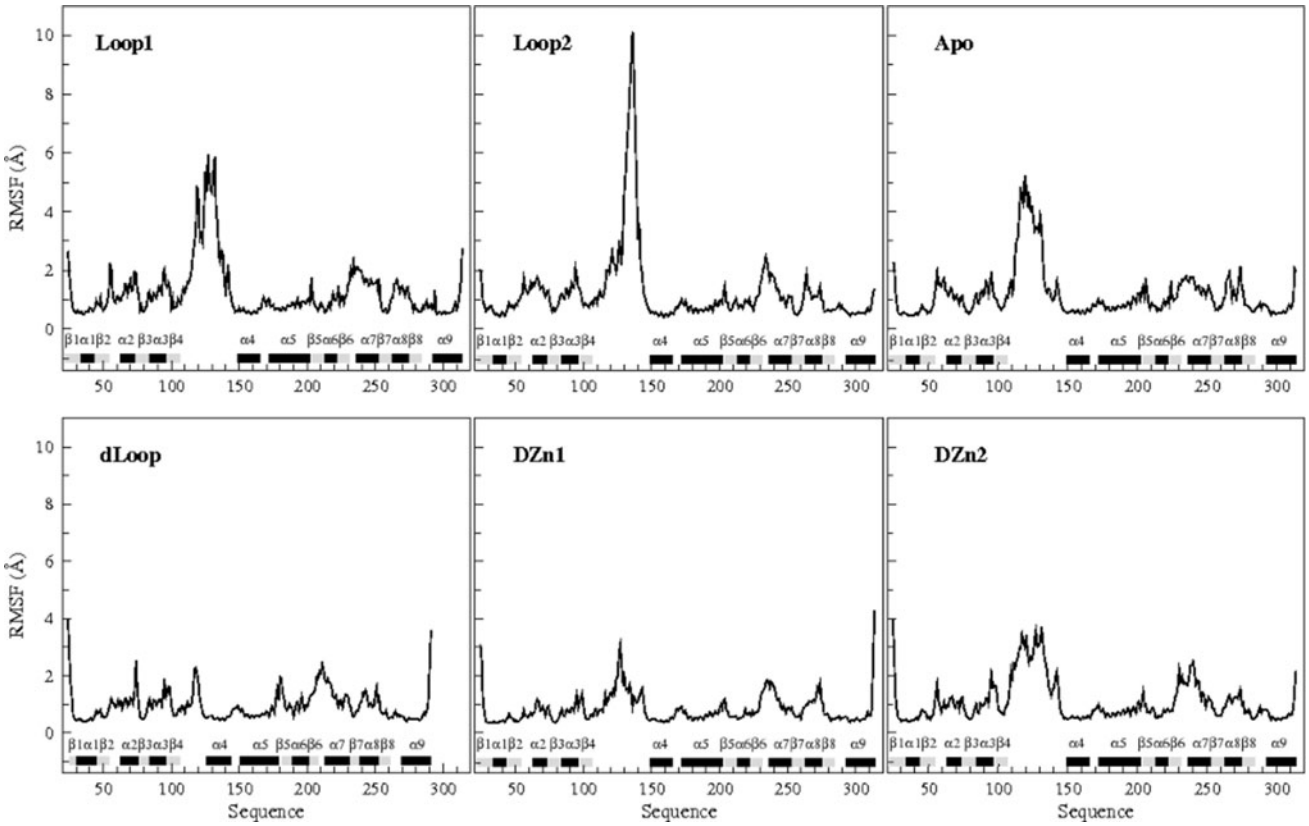

Fig. 5 Average, per residue, backbone RMSF for Loop1, Loop2, Apo, dLoop DZn1 and DZn2 *Se*-ZnuA trajectories. The secondary structure regions in the starting structure are shown by the *black* (α-helix) and *grey* (β-strand) *horizontal bars* on the X-axis

simulations the His-rich loop fluctuations partially decrease likely because of the presence of a second metal ion. In the dLoop trajectory the short part of the loop, that remains after the deletion introduced, maintains a relatively large degree of mobility. For all systems the conformations sampled by the His-rich loop are various and, although wide, do not alter the primary zinc site solvent accessibility, as shown by the time evolution of the zinc site surface area (Fig. 5s of supplementary material), which is stable with a value of about 400 $Å^2$. This result indicates that the loop does not mask the zinc accessibility, but likely modulates the zinc coordination stability (see next paragraphs), producing different structural stress depending on its orientation.

Other high mobility regions include the protein segments containing the two small helices symmetrically located at the border of the protein (helices α2, α3 and α7, α8; as defined in *Eco*-ZnuA) [6] that shield the two short parallel β-sheets (β1, β2, β3, β4 and β5, β6, β7, β8) from solvent (Fig. 1). As the helices move oscillating, along with the loop motion, these fluctuations may have the function to disperse the mechanical stress imposed by the His-rich loop to the protein scaffold.

Salt bridges

Table 6 shows the permanence of salt bridges that are present for at least 60% of the simulation time. For comparison purposes, when a salt bridge is found for at least 60% of the simulation time in one simulation, then it is reported for all the other five systems even if present with small percentages (not bold numbers in Table 6).

The salt bridges present in the protein scaffold form independently of the initial structure and are common for all the investigated systems indicating that they are necessary to provide stability to the protein. The situation is more complex for the His-rich loop. As expected just few salt bridges are observed within this protein region and only one of them, His139–Asp212, is present in three systems, namely Loop1, Apo and DZn1. The salt bridge is present for a high percentage of time (greater than 80%) and involves one residue belonging to the loop (His139) and one residue belonging to the protein scaffold (Asp212), suggesting that Asp212 represents an important structural link between the His-rich loop and the rest of the protein. This is confirmed by the fact that in the DZn2 system, where this salt bridge is not present, Asp212 is involved in

Table 6 Presence of salt bridges (time percentages)

Salt bridge	Loop1 %	Loop2 %	dLoop %	Apo %	DZn1 %	DZn2 %
Asp44–Lys293	**88**	**87**	**87**	**87**	**86**	**84**
Asp55–Lys33	**75**	25	**78**	**60**	**100**	**73**
Asp68–Arg65	**99**	**90**	**100**	**100**	**89**	**78**
Asp68–Arg71	**75**	**78**	**67**	**85**	**70**	55
Asp76–Lys100	**67**	**91**	**60**	**84**	**96**	**60**
Glu90–Arg94	**86**	4	40	**85**	**92**	**78**
Asp98–Lys165	**60**	**99**	34	**100**	42	35
○Glu122–Arg241	–	**62**	–	–	–	–
●Glu132–Lys134	–	24	–	**63**	–	4
○His138–Asp212	–	–	–	–	–	**61**
○His139–Asp212	**83**	–	–	**92**	**83**	15
Glu153–Arg156	**91**	**100**	**100**	**97**	**100**	**63**
His163–Asp178	**68**	**67**	**63**	**67**	**61**	**61**
Asp178–Lys182	**79**	**86**	**93**	**74**	**80**	**83**
Asp183–Lys293	**89**	**88**	**90**	**83**	**89**	**81**
Glu198–Lys312	47	**67**	**63**	47	**66**	**61**
Glu218–Lys219	**91**	54	55	–	12	1
His228–Asp212	30	16	5	6	**70**	20
Glu234–Arg264	**66**	55	**87**	**81**	29	**79**
Glu244–Arg241	**90**	20	**81**	**93**	**68**	**69**
Glu269–Arg273	52	51	**68**	**80**	**72**	**69**
Total number (60%)	15	11	13	16	15	14

● Salt bridges inside the His-rich loop, ○ salt bridges between His-rich loop and protein scaffold

The bold numbers represent salt bridges present for at least 60% of the simulation time

a salt bridge with the nearby His138 for 61% of the simulation time (Table 6).

Zinc ligands

The zinc binding site has been parametrized using the CaDA approach [32–34], that allows bond distance variation of the ligands from the metal ion. This approach is particularly suited for the analysis of a protein like ZnuA since the zinc ion must be transported and cannot be stably bound by a set of given ligands. In contrast, it must have the possibility to switch from one ligand to another in order to be successfully transferred. This procedure allows one to monitor how the conformational stress induced by the protein motion, in particular by the His-rich loop flexibility, can perturb the zinc coordination and permits one to hypothesize possible routes for the zinc transportation.

Analysis of the trajectories indicates that the ligand-zinc distance in the primary site is constant for three coordinating ligands, namely Glu59, His60 and His147, but is fluctuating for His211. The His211 behaviour is different in each system but in all cases it is the ligand that shows the largest distance variation from the zinc ion. In fact, the plot of His211–Zn distance, as a function of time, for the Loop1, Loop2 and dLoop system (Fig. 6), shows that it oscillates from a zinc coordinating to a non zinc coordinating value and the highest oscillations are observed in the Loop2 system, where values as high as 10 Å are reached (Fig. 6). It is interesting to notice that for all the three systems His211 returns several times at a coordination distance after reaching large non coordinating distances, indicating that the detachment is reversible and suggesting that the stretching of the coordination bond is dependent on the orientation/fluctuation of the protein scaffold and likely of the His-rich loop. The most striking behaviour is observed in the DZn2 system where His211 (Fig. 6, last panel, black line) is detached from the zinc after about 2 ns, being replaced in the first coordination sphere by Glu260 (Fig. 6, last panel, light grey line). The two residues actually maintain a common coordination distance from the metal for a short part of the simulation, but afterwards Glu260 becomes the fourth coordinating ligand in the primary binding site. The result is system dependent since such a behaviour is not observed for the DZn1 model, where the secondary site has been differently modelled, but provides evidence for the plasticity of the zinc site, a prerequisite in a molecular system where the zinc ion cannot be tightly bound to the protein, but must have the possibility to be transferred from one ligand to another. The His211 ligand stability observed in DZn1 (Fig. 6), is assisted by a highly permanent salt bridge established between the scaffold residue Asp212 and His139, belonging to the His-rich loop (Table 6, Fig. 7a). Asp212 is involved in the interaction with the His-rich loop also in the

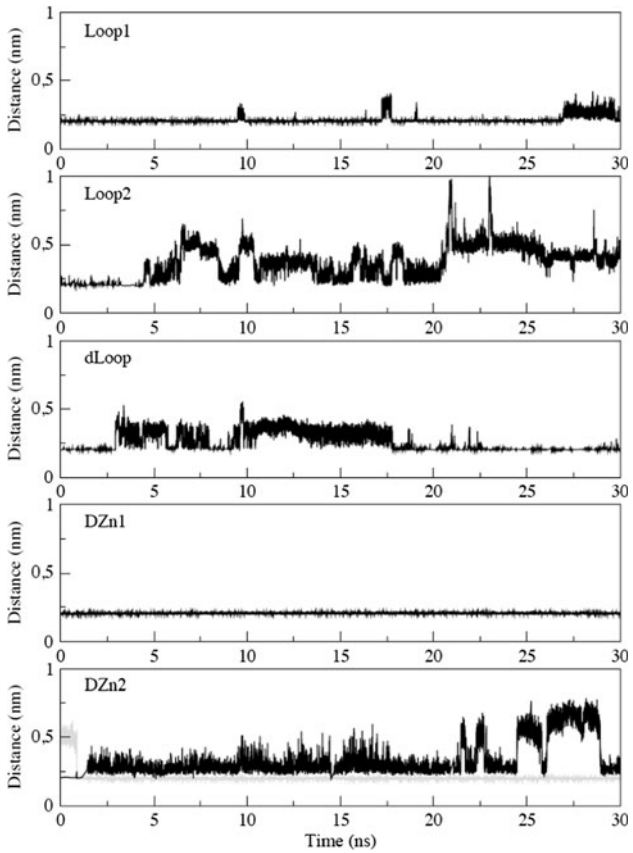

Fig. 6 Panels showing the His211–Zn distances in the primary zinc site of each simulation that includes the metal (*black lines*). The *light grey line* in the DZn2 panel indicates the distance Glu260–Zn

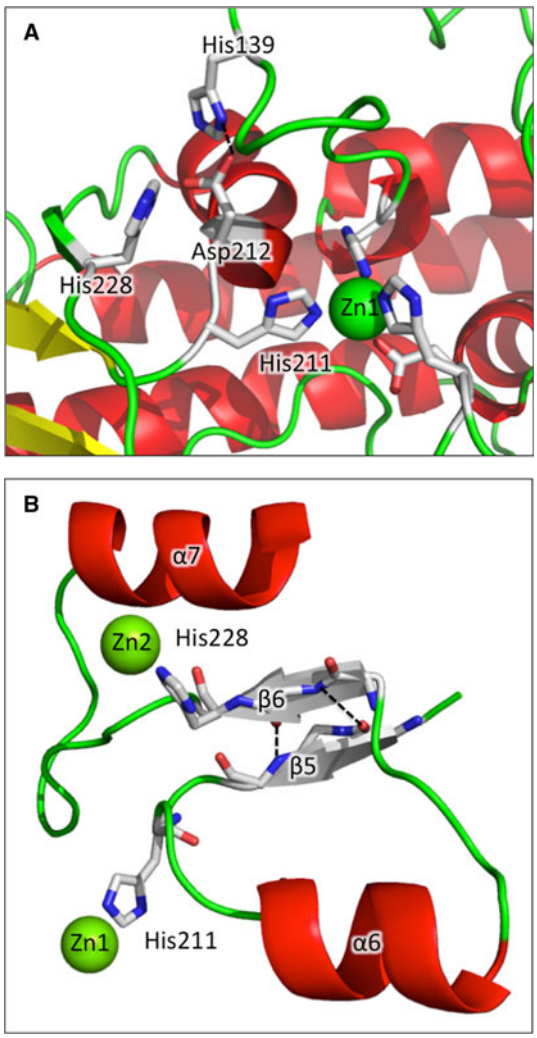

Fig. 7 **a** Picture showing the His139–Asp212 salt bridge (*dashed line*): the α-helices are indicated by *red spirals* while the *β*-strand are represented by *grey arrows*. Random coil and turns are shown as a *green wire* and the zinc metals are represented by *green spheres*. The zinc ligands and the salt bridge components are evidenced by sticks models coloured by atom types. **b** Representation of the backbone connection between the primary and the secondary zinc site: the α-helices are indicated by *red spirals* while the *β*-strand are represented by *grey arrows*. Random coil and turns are shown as a *green wire* and the zinc metals are represented by *green spheres*. The zinc ligands and the backbone of *β*-strand residues forming high permanence hydrogen bonds (*dashed lines*) are evidenced by sticks models coloured by atom types

DZn2 model but in this case its salt link partner is His138 since His139 participates in the coordination of the second zinc site. Based on this scenario we suggest that the scaffold residue Asp212, has an important role in the modulation of the interaction with the His-rich loop.

Interesting structural alterations occur also in the secondary zinc site of the DZn2 model. In fact, His228, the only protein scaffold residue identified by X-ray diffraction as a zinc ligand in the secondary metal-binding site [8], is detached from zinc after about 7.5 ns of simulation. His228 is not replaced by any other ligand, but again this event demonstrates that the zinc ions are loosely bound and that conformational stresses can modify their coordination. In DZn2 the enhanced mobility of the His-rich loop, leading to the detachment of His228 in the secondary zinc site, alters the zinc coordination also in the primary site. As a matter of fact the primary and the secondary zinc site are structurally related by a symmetric motif composed by 2 *β*-strands, 2 loops and 2 α-helices (*β*5-loop-α6-loop-*β*6-loop-α7; from Gly206 to Gln248). His211, located on the *β*5-α6 loop, is coordinating the first zinc site, whilst His228, located on the C-terminal of the *β*6 strand, is coordinating the second zinc (Fig. 7b). Changes occurring on one zinc site are then transmitted to the other site and vice versa through this direct backbone connection that is also strengthen by the presence of two highly permanent hydrogen bonds between residues close to the two coordinating histidines (i.e. Tyr207:O–Leu226:NH and Val209:NH–Gly227:O; permanence above 70% in all the simulations) (Fig. 7b).

Table 7 PCA analysis values

Properties	Loop1	Loop2	Apo	dLoop	DZn1	DZn2
Total eigenvectors number	873	873	873	804	873	873
Eigenvectors for 90% of motion	32	14	37	61	82	44
Percentage of total motion enclosed in first eigenvector	33	62	33	30	27	28

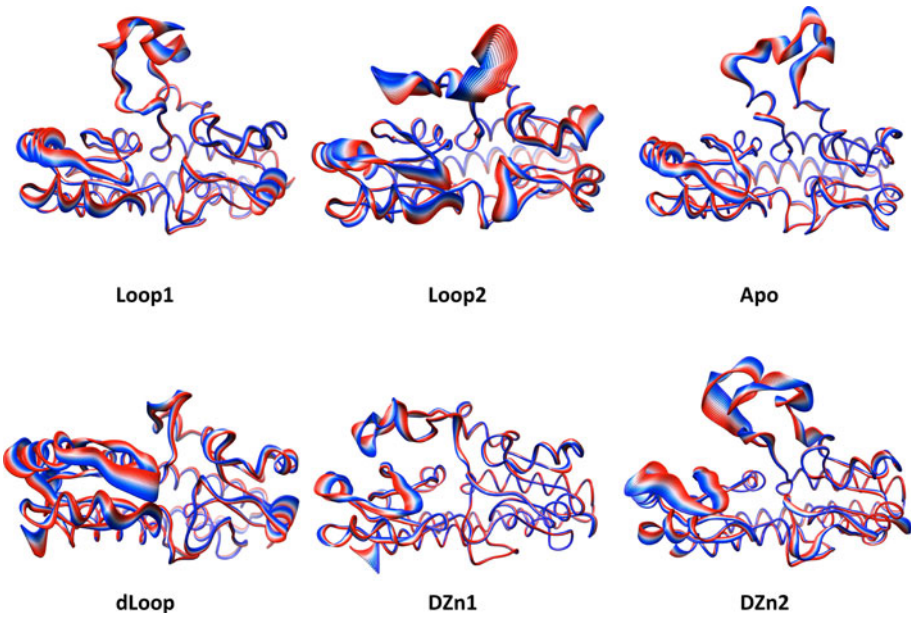

Fig. 8 Tube representation of the motion projections along the first eigenvector for Loop1, Loop2, Apo, dLoop DZn1 and DZn2. The direction of the motion is indicated by the *flanked tubes*, the versus being defined from the *red* to the *blue colour*. The picture was produced using the program VMD [49]

Principal component analysis

The principal component analysis (PCA), or essential dynamics [44, 45], has been also applied to highlight the correlation differences between the six simulations. The analysis, carried out on the Cα atoms of the proteins, is based on the diagonalisation of the covariance matrix built from the atomic fluctuations after the removal of the translational and rotational movement, and permits the identification of the main 3 N directions along which the majority of the protein motion is defined [44, 45]. The total number of eigenvectors in which the motion is dispersed, the number of eigenvectors from which depends about 90% of the motion and the percentage of total motion enclosed in the first eigenvector are reported in Table 7.

The projections of the motions calculated along the first eigenvector are shown in Fig. 8. The width of the ribbon shows the amplitude of the backbone motion whilst the direction goes from the red to the blue colour. The PCA analysis indicates that the His-rich loop is the protein portion displaying the larger amplitude motions while the motions that are associated with the other regions depends on the number of zinc ions included in the model and on the starting arrangement of the His-rich loop, indicating that these two elements have an important role in the modulation of the protein motions. As indicated in the RMSF paragraph, the oscillation of helices α2, α3 and α7, α8 [6], that shield two short parallel β-sheets from solvent (Fig. 1), occur along with the movements of the His-rich loop and their fluctuations may have the role to disperse the mechanical stress caused by this large and unstructured portion of the protein.

Conclusions

The MD simulations here reported provide a comparative structural dynamical picture of the *Se*-ZnuA protein in its apo and metal-bound forms. In all the investigated systems the His-rich loop is the largest fluctuating region independently of the initial conformation. The loop influences the stability of the protein scaffold (Figs. 4, 5, Table 4), probably modulating the interaction of ZnuA with molecular partners, but it never masks the zinc site accessibility (Fig. 5s). The conformational properties of the His-rich loop suggest that the design of new drugs interfering with the metal uptake in the primary zinc site may represent a potential way to combat bacterial infections. The communication between the protein scaffold and the His-rich loop is likely mediated by Asp212, the only residue involved in formation of stable salt bridges with His138 or His139, belonging to the loop (Table 6).

The primary and the secondary zinc site are structurally related by a symmetric motif, composed by 2 β-strands, 2 loops and 2 α-helices, that permits a direct backbone communication between His211 and His228, coordinating the primary and secondary zinc site respectively (Fig. 7b). Changes occurring on one zinc site are then transmitted to the other site and vice versa. In this process Asp212 is hypothesized to have an indirect but crucial role due to its engagement with the His-rich loop. The MD simulations actually show, in the limit of time scale and of the models used, that some ligands can reversibly attach and detach from the metal (Fig. 6, last panel), suggesting a plasticity of the zinc sites, a necessary prerequisite of a molecular system whose function is the transfer of zinc ions. As observed by the PCA analysis the presence of zinc and the conformations assumed by the His-rich loop influence the motions of different portions of the protein scaffold. In conclusion, this study suggests that, in the time limit of simulations, the large fluctuations of the His-rich loop may have functional significance inducing plasticity to the zinc binding residues, a necessary prerequisite to generate a route for the metal transport process between different sites.

Acknowledgments The authors thank professor E. Chiancone for her interest, the reading of the manuscript and the helpful comments. This work was partially supported by a grant from the *Fondazione Roma*. O.F. thanks Filas for a fellowship granted under the project "Caratterizzazione di principi attivi".

References

1. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM (2008) Metal ions in biological catalysis: from enzyme databases to general principles. J Biol Inorg Chem 13:1205–1218
2. Ammendola S, Pasquali P, Pistoia C, Petrucci P, Petrarca P, Rotilio G, Battistoni A (2007) High-affinity Zn^{2+} uptake system ZnuABC is required for bacterial zinc homeostasis in intracellular environments and contributes to the virulence of *Salmonella enterica*. Infect Immun 75:5867–5876
3. Hantke K (2005) Bacterial zinc uptake and regulators. Curr Opin Microbiol 8:196–202
4. Locher KP (2009) Structure and mechanism of ATP-binding cassette transporters. Phil Trans R Soc B 364:239–245
5. Rees DC, Johnson E, Lewinson O (2009) ABC transporters: the power to change. Nat Rev Mol Cell Biol 10:218–227
6. Li H, Jogl G (2007) Crystal structure of the zinc-binding transport protein ZnuA from *Escherichia coli* reveals an unexpected variation in metal coordination. J Mol Biol 368:1358–1366
7. Chandra BR, Yogavel M, Sharma A (2007) Structural analysis of ABC-family periplasmic zinc binding protein provides new insights into mechanism of ligand uptake and release. J Mol Biol 367:970–982
8. Yatsunyk LA, Easton JA, Kim LR, Sugarbaker SA, Bennett B, Breece RM, Vorontsov II, Tierney DL, Crowder MW, Rosenzweig AC (2008) Structure and metal binding properties of ZnuA, a periplasmic zinc transporter from *Escherichia coli*. J Biol Inorg Chem 13:271–288
9. Wei B, Randich AM, Bhattacharyya-Pakrasi M, Pakrasi HB, Smith TJ (2007) Possible regulatory role for the histidine-rich loop in the zinc transport protein, ZnuA. Biochemistry 46:8734–8743
10. Banerjee S, Wei B, Bhattacharyya-Pakrasi M, Pakrasi HB, Tj Smith (2003) Structural determinants of metal specificity in the zinc transport protein ZnuA from *Synechocystis* 6803. J Mol Biol 333:1061–1069
11. Claverys JP (2001) A new family of high-affinity ABC manganese and zinc permeases. Res Microbiol 152:231–243
12. Berducci G, Mazzetti AP, Rotilio G, Battistoni A (2004) Periplasmic competition for zinc uptake between the metallochaperone ZnuA and Cu, Zn superoxide dismutase. FEBS Lett 569: 289–292
13. Desrosiers DC, Sun YC, Zaidi AA, Eggers CH, Cox DL, Radolf JD (2007) The general transition metal (Tro) and Zn2(Znu) transporters in *Treponema pallidum*: analysis of metal specificities and expression profiles. Mol Microbiol 65:137–152
14. Petrarca P, Ammendola S, Pasquali P, Battistoni A (2010) The Zur-regulated ZinT protein is an auxiliary component of the high-affinity ZnuABC zinc transporter that facilitates metal recruitment during severe zinc shortage. J Bacteriol 192:1553–1564
15. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201
16. Bates PA, Kelley LA, MacCallum RM, Sternberg MJE (2001) Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins Struct Funct Gen 5:39–46
17. Nielsen M, Lundegaard C, Lund O, Petersen TN (2008) CpH-Models-30 remote homology modeling using structure guided profile sequence alignment and double-sided baseline corrected scoring scheme. CASP8 conference, 3–7 Dec, Cagliari, Sardinia, Italy, p 193
18. Jamroz M, Kolinski A (2010) Modeling of loops in proteins: a multi-method approach. BMC Struct Biol 10:5
19. DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific. San Carlos, CA, USA. (http://wwwpymolorg)
20. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
21. Davis IW, Weston Murray L, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:375–383
22. Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D66:12–21
23. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
24. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
25. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012
26. Falconi M, Parrilli L, Battistoni A, Desideri A (2002) Protein flexibility in monomeric Cu, Zn superoxide dismutase detected by molecular dynamics simulation and limited proteolysis. Proteins Struct Funct Genet 47:513–520

27. Stote R, Karplus M (1995) Zinc binding in proteins and solution: a simple but accurate nonbonded representation. Proteins Struct Funct Genet 23:12–31
28. Liang JH, Lipscomb WN (1990) Binding of substrate CO2 to the active site of human carbonic anhydrase I: a molecular dynamics study. Proc Natl Acad Sci USA 87:3675–3679
29. Sakharov DV, Lim C (2005) Zn protein simulations including charge transfer and local polarization effects. J Am Chem Soc 127:4921–4929
30. Sakharov DV, Lim C (2009) Force fields including charge transfer and local polarization effects: application to proteins containing multi/heavy metal ions. J Comput Chem 30:191–202
31. MacKerell JAD, Bashford D, Bellott M, Dunbrack R, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WEI, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
32. Pang YP (1999) Novel zinc protein molecular dynamics simulations: steps toward antiangiogenesis for cancer treatment. J Mol Model 5:196–202
33. Pang YP, Xu K, El Yazal J, Prendergast FG (2000) Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. Protein Sci 9:1857–1865
34. Pang YP (2001) Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. Proteins 45:183–189
35. Jorgensen WL, Mahoney MW (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys 112:8910–8922
36. Darden T, York D, Pedersen L (1993) Particle mesh Ewald an N_log(n) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
37. Cheatham TE, Miller JL, Fox T, Darden TA, Kollman PA (1995) Molecular dynamics simulation on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. J Am Chem Soc 117:4193–4194
38. Toukmaji A, Sagui C, Board J, Darden T (2000) Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. J Chem Phys 113:10913–10927
39. Brünger A, Brooks CL III, Karplus M (1984) Stochastic boundary conditions for molecular dynamics simulations of ST2 water. Chem Phys Lett 105:495–500
40. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. J Chem Phys 101:4177–4189
41. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. J Chem Phys 103:4613–4621
42. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341
43. Miyamoto S, Kollman PA (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comp Chem 13:952–962
44. Garcia AE (1992) Large-amplitude nonlinear motions in proteins. Phys Rev Lett 68:2696–2699
45. Amadei A, Linssen AB, Berendsen HJ (1993) Essential dynamics of proteins. Proteins 17:412–425
46. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4:435–447
47. Berendsen HJC, van der Spool D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 95:43–56
48. Hubbard SJ, Thornton JM (1993) NACCESS, computer program. Department of Biochemistry and Molecular Biology, University College, London
49. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
50. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins Struct Funct Genet 12:345–364