Journal of Molecular Graphics and Modelling 38 (2012) 211–219

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Structural analysis of Ca^{2+} dependent and Ca^{2+} independent type II antifreeze proteins: A comparative molecular dynamics simulation study

Sangeeta Kundu [a], Debjani Roy [b,*]

[a] Bioinformatics Centre, Bose Institute, Kolkata, India
[b] Department of Biophysics, Bose Institute, Kolkata, India

ARTICLE INFO

Article history:
Accepted 5 May 2012
Available online 9 June 2012

Keywords:
Antifreeze protein
Molecular dynamics
Principal component analysis
Unfolding
Potential energy map

ABSTRACT

Comparative molecular dynamics simulations of Ca^{2+} dependent psychrophilic type II antifreeze protein (AFP) from herring (*Clupea harengus*) (hAFP) and Ca^{2+} independent type II antifreeze protein from long snout poacher (*Brachyopsis rostratus*) (lpAFP) have been performed for 10 ns each at five different temperatures. We have tried to investigate whether the Ca^{2+} dependent protein obtains any advantage in nature over the independent one. To this end the dynamic properties of these two proteins have been compared in terms of secondary structure content, molecular flexibility, solvent accessibility, intra molecular hydrogen bonds and protein–solvent interactions. At 298 and 373 K the flexibility of the Ca^{2+} independent molecule is higher which indicates that Ca^{2+} could contribute to stabilize the structure. The thermal unfolding pathways of the two proteins have also been monitored. The rate of unfolding is similar up to 373 K, beyond that hAFP shows faster unfolding than lpAFP. The essential subspaces explored by the simulations of hAFP and lpAFP at different temperatures are significantly different as revealed from principal component analysis. Our results may help in understanding the role of Ca^{2+} for hAFP to express antifreeze activity. Furthermore our study may also help in elucidating the molecular basis of thermostability of two structurally similar proteins, which perform the same function in different manner, one in presence of Ca^{2+}, and the other in absence of the same.

1. Introduction

Psychrophilic, or cold-adapted, antifreeze proteins (AFPs) are widespread in organisms living in extreme cold environments. These proteins are generated by several species of plants, microbes and ectothermic animals as a weapon to protect themselves from freezing damage. AFPs depress the freezing point of blood and body fluids below that of the surrounding seawater by binding to and inhibiting the growth of seed ice crystals [1–3]. Antifreeze activity of AFPs has enormous potential, commercial applications. They can be used for protection of economically important fishes and plants against frost or low temperature [4]. Several valuable aquaculture species like Atlantic Salmon and goldfish cannot survive in ice cold sea water, which causes severe limitation in culturing commercially important fishes in colder climates [5,6]. Successful administration and expression of functional AFP genes into these marine fishes generate transgenic fishes which enable them to withstand ice cold sea water. AFPs can also be used as additives to improve the quality and shelf-life of frozen food, as cryoprotective agents for organ and cell cryopreservation, as chemical adjuvants to cancer cryosurgery [7,8]. During the last few decades there has been an increase in interest on AFPs due to their wide industrial applications.

Fish AFPs are classified into five distinct classes based on their structure (type I–type IV and antifreeze glycoprotein) [4]. The focus of this work, type II, AFPs are 14–24 kDa cysteine rich proteins which have been found in herring, Japanese smelt, rainbow smelt, sea raven and long snout poacher [4]. They are largest globular fish AFPs known to date [4]. Type II AFPs are homologous to C-type (Ca^{2+} dependent) lectin like domains (CTLDs) [9]. Homology modeling and NMR studies reveal that these proteins share the same characteristic fold, which include disulphide bridges, α-helices and β-sheets as well as large portion of coil structure. Both herring and two smelt AFPs depend on Ca^{2+} for their antifreeze activity, while the functions of sea raven and long snout poacher AFPs are Ca^{2+} independent. Recently the crystal structures of type II antifreeze protein of herring (*Clupea harengus*) containing 127 residues [4] (2PY2) and long snout poacher (*Brachyopsis rostratus*) containing 130 residues [10] (2ZIB) have been determined at 1.70 Å and 1.34 Å resolutions respectively. These two

* Corresponding author at: Department of Biophysics, Bose Institute, Acharya J.C Bose Centenary Building, P-1/12 C.I.T Scheme – VII M, Kolkata 700054, India. Tel.: +91 33 2569 3213; fax: +91 33 2355 3886.
E-mail addresses: debjani@bic.boseinst.ernet.in, drdebjani@yahoo.com (D. Roy).

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.05.004

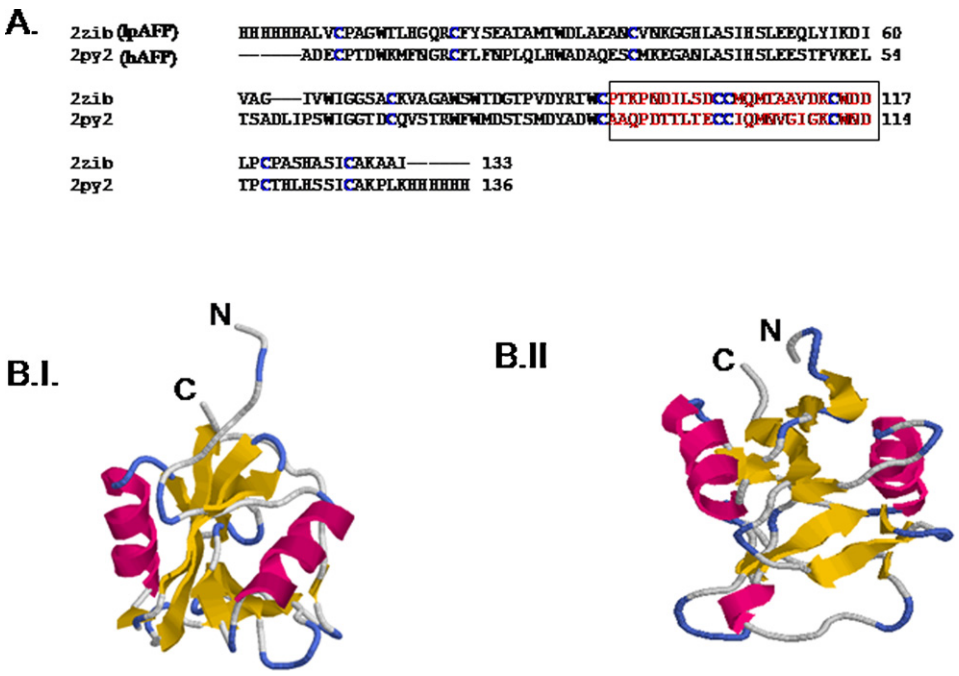

Fig. 1. (A) Pairwise sequence alignment of lpAFP and hAFP. Sequences of Ca^{2+} binding region are shown in box. Conserved cysteine residues of the disulphide bonds are shown in blue (B) The native structure of (I) lpAFP and (II) hAFP. Helices are shown as magenta ribbons, β-strands as yellow arrows and the rest are shown as loops. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

antifreeze proteins are abbreviated as hAFP and lpAFP throughout the text. These two proteins share 39% sequence identity and the root-mean-square deviation of Cα atoms of the superimposed structures is 1.1 Å. hAFP works Ca^{2+} dependently while lpAFP works Ca^{2+} independently. Their overall fold contains two α-helices and eight β-strands (Fig. 1). In addition hAFP has a very small 3_{10} helix. Each monomer of Ca^{2+} dependent hAFP consists of residues from Cys1 to Lys127 having two α-helices (α1, residues 22–30, and α2, 42–50), and eight β-strands (β1, residues 6–8; β2, 11–20; β3, 35–36; β4, 59–64; β5, 73–74; β6, 98–103; β7, 108–112; and β8, 118–125) connected by stretches of irregular secondary structure like turns, bends and β-bridges (Fig. 1). lpAFP on the other hand is composed of residues from His1–Ile130 that also possesses two α-helices (α1, residues 28–37, and α2, 48–57), and eight β-strands (β1, residues 12–14; β2, 17–26; β3, 41–42; β4, 62–67; β5, 76–77; β6, 101–104; β7, 112–115; and β8, 121–128). Presence of disulphide bonds is a unique feature of type II AFPs. In both lp and hAFP all 10 cysteines are paired to form five disulphide bonds. In, hAFP, disulphide bonds are constituted between residues, Cys1–Cys12; Dsb1, Cys29–Cys122; Dsb2, Cys66–Cys97; Dsb3, Cys86–Cys108; Dsb4, and Cys98–Cys114; Dsb5. Three of the five disulphide bonds are conserved in the CTLDs. Additional two disulphide bonds (Cys66–Cys97 and Cys86–Cys108 are located at each end of the Ca^{2+} binding loop (Ala87–Asp111). In lpAFP, intramolecular disulfide bonds exist between residues (Cys7–Cys18; Dsb1, Cys35–Cys125; Dsb2, Cys69–Cys100; Dsb3, Cys89–Cys111; Dsb4, and Cys101–Cys117; Dsb5). In both Ca^{2+} dependent hAFP and independent lpAFP, Dsb2 and Dsb5 connect α1–β8, and β6 to a nonstructural region joining β7 and β8, respectively. Thus these two different proteins display considerable structural similarity. This seems to be quite interesting that two proteins having significant structural similarity can perform the same function in different manner, one Ca^{2+} dependently, and the other Ca^{2+} independently. In the present work, we have performed a comparative study between these two proteins in order to shed light on the molecular features responsible for cold adaptation in these two different psychrophilic proteins. We have tried to investigate whether the Ca^{2+} dependent protein obtains any advantage in nature over the independent one. Molecular dynamics (MD) simulation is an effective tool to evaluate the comparative basis of thermostability, flexibility between two different proteins. So far, no comparative molecular dynamics study dealing with wide range of temperatures has been reported for these two different type II AFPs. In this paper for the first time we have performed molecular dynamics simulation of Ca^{2+} dependent and independent type II AFPs at five different temperatures namely 277 K, 298 K, 373 K, 423 K and 473 K in order to understand the thermodynamic behavior of these two proteins. We also have made an attempt to investigate the role of Ca^{2+} in the structural integrity and activity of type II AFPs.

2. Methodology

2.1. Molecular dynamics simulation

All MD simulations were performed using the GROMACS 4 [11,12] package and GROMOS96 [13] 43a1 force field implemented on LINUX architecture. The starting configuration of Ca^{2+} independent type II AFP containing 130 residues was obtained from long snout poacher, *B. rostratus* while that of Ca^{2+} dependent type II AFP consisting of 127 residues was derived from herring, *C. harengus* (PDB code: 2ZIB [10] and 2PY2 [4] respectively). We abbreviated 2ZIB and 2PY2 as lpAFP and hAFP throughout the text. Pair wise sequence alignment was performed using ClustalW [14] which shows that these two proteins display 39% sequence identity. Crystallographic water molecules and heteroatoms were removed from the systems except the Ca^{2+} ion in case of hAFP. All starting structures were immersed in a triclinic box of SPC water molecules [15]. To neutralize the charge of the system $4Na^{+}$ ions were added to lpAFP and $5Na^{+}$ ions were added to hAFP. The resulting system sizes are listed in Supplementary Table 1. All protein atoms were at a distance equal to 1.0 nm from the box edges. Each system was subjected to energy minimization for 2000 steps by steepest descents. The minimized systems for both lpAFP and hAFP were equilibrated for 50 ps each at five different temperatures namely 277 K, 298 K,

373 K, 423 K and 473 K by position restrained molecular dynamics simulation in order to relax the solvent. The equilibrated systems were then subjected to molecular dynamics simulations for 10 ns each at five different temperatures. In sum, trajectories of 100 ns (50 ns for lpAFP, 50 ns for hAFP) were collected for two proteins investigated. Periodic boundary conditions combined with minimum image convention were used under isothermal, isobaric conditions using Berendsen coupling algorithm [16] with relaxation times of 0.1 ps and 0.5 ps, respectively. The pressure value is 1 bar (1 atm to 0.983 bar). The LINCS algorithm [17] was used to constrain bond lengths using a time step of 2 fs for all calculations. Electrostatic interactions were calculated using the Particle Mesh Ewald (PME) [18,19] summation scheme. van der Waals and Coulomb interactions were truncated at 0.9 nm. The non-bonded pair list was updated every 10 steps and conformations were stored every 2 ps. Secondary structure analysis was performed using the program DSSP [20]. Other analyses were performed using scripts included with the GROMACS [21] distribution. The visual analysis of protein structures was carried out using Rasmol [22].

2.2. Root mean square deviation (RMSD)

The root-mean-square deviation (RMSD) of a selected element with respect to its reference value is defined as

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - r_o)^2}$$

where r_i represents the element position at time i and r_o the reference value. RMSD of certain atoms in a molecule with respect to a reference structure can be calculated with the program g_rms of GROMACS [21] by least-square fitting the structure to the reference structure and subsequently calculating the RMSD.

2.3. Root mean square fluctuation (RMSF)

The root-mean-square fluctuation (RMSF) of a selected element with respect to its average value is defined as

$$\mathrm{RMSF} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i - \langle r\rangle)^2}$$

where r_i represents the element position at time i and $\langle r\rangle$ the average value. g_rmsf tool of GROMACS [21] computes the root mean square fluctuation (RMSF, i.e. standard deviation) of atomic positions after fitting to a reference frame.

2.4. Radius of gyration (Rg)

Radius of gyration (Rg) provides a rough measure for the compactness of a structure. g_gyrate program of GROMACS [21] is capable of calculating Rg which is as follows:

$$\mathrm{Rg} = \sqrt{\frac{\sum_i m_i ||r_i||^2}{\sum_i m_i}}$$

where m_i is the mass of atom i and r_i the position of atom i with respect to the center of mass of the molecule.

2.5. Solvent accessible surface area (SASA)

g_sas tool of GROMACS [21] computes hydrophobic, hydrophilic and total solvent accessible surface area. Solvent accessible surface area (SASA) is defined as the surface traced by the center of a probe solvent molecule as it rolls on the van der Waals surface of a molecule. Hydrophobic and hydrophilic SASA denote the SASA of the hydrophobic and hydrophilic amino acids of the protein respectively which together provides total SASA.

2.6. Essential dynamics

Essential degrees of freedom of both lpAFP and hAFP were extracted from equilibrated portion of the trajectories according to principal component analysis (PCA) or essential dynamics (ED) method [23,24]. The ED method is based on the construction of the covariance matrix (C) of the coordinate fluctuations.

$$\mathrm{C} = cov(x) = \langle (x - \langle x\rangle)(x - \langle x\rangle)^T\rangle$$

where $\langle\rangle$ indicates the average and x is the vector of the atomic positions. T is the transpose of the matrix.

After removal of the translational and rotational degrees of freedom the covariance matrix C is calculated. The covariance matrix is then diagonalized to obtain the eigenvectors and eigenvalues that provide information about correlated motions throughout the protein. The eigenvectors represent the directions of motion, and the eigenvalues represent the amount of motion along each eigenvector. The eigenvectors are then sorted according to their eigenvalues in descending order. Usually, the first 10–20 eigenvectors suffice to describe almost all conformational substates accessible to the protein. For the simulation of both lpAFP and hAFP only Cα atoms were included in the definition of the covariance matrices for the protein. The root-mean-square inner product (RMSIP) [25,26] was calculated to measure the degree of overlap between the conformational spaces of the two proteins explored by the simulation at both the same and different temperatures. The RMSIP is defined as follows:

$$\mathrm{RMSIP} = \sqrt{\frac{1}{10}\sum_{i=1}^{10}\sum_{j=1}^{10}(\eta_i^a \cdot \eta_j^b)^2}$$

Where η_i^a, η_j^b are the ith and jth eigenvectors from two different sets a and b respectively. The RMSIP was computed from the equilibrated portion of the trajectory. For diagonal elements the RMSIP was computed by splitting the equilibrated portion of the trajectory into two equal halves. Sub-matrices of the overall covariance matrices that only involved segments surrounding the Ca^{2+} binding loop region were subjected to PCA and RMSIP analyses.

2.7. Potential energy surface

To draw potential energy surface of the protein we calculated the principal components 1, 2 and potential energy of the system as a function of time for all the simulations of both lp and hAFP. Thus for each combinations of PC1 and PC2 experienced by the protein the potential energy values were assembled. We used ORIGIN program to generate the 3D data grid using correlation method. The correlation method computes a new value for each cell in the matrix from the values of the points in the adjoining cells in the matrix that are included within the search radius. Thus smoothness of the surface is achieved.

3. Results and discussion

3.1. Structural stability

The root-mean-square deviation (RMSD), a crucial parameter to evaluate the equilibration of MD trajectories, is computed for backbone atoms by using the starting structures of the MD simulations as references. Fig. 2A(I) and (II) shows the backbone RMSDs of

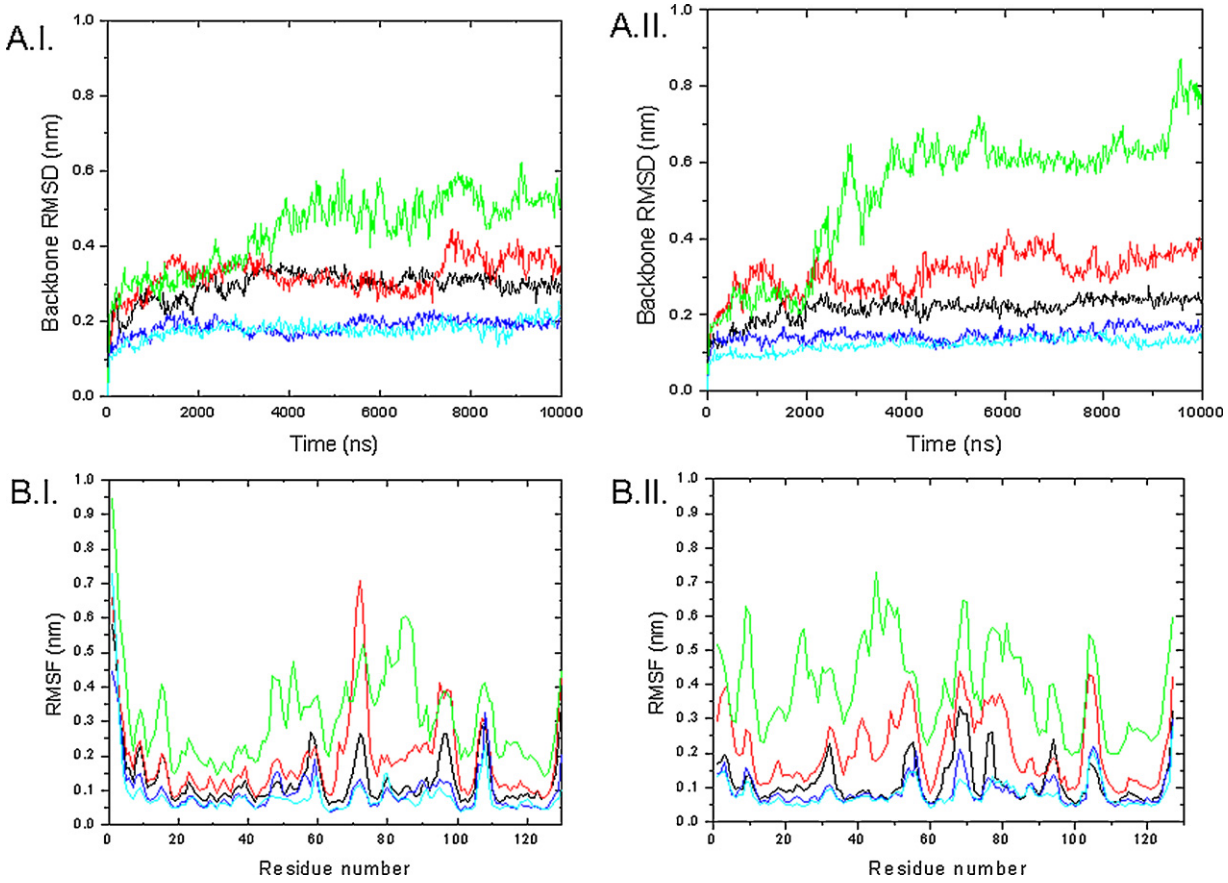

Fig. 2. Time evolution of (A) backbone RMSD of (I) lpAFP and (II) hAFP at different temperatures. (B) RMSFs as a function of residue number of (I) lp and (II) hAFP at different temperatures. The color-coding scheme is as follows: 277 K (blue), 298 K (cyan), 373 K (black), 423 K (red) and 473 K (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

lp and hAFP from the corresponding starting structures as a function of time at different temperatures. No significant difference in RMSD is found at lower temperatures (277 and 298 K), however at 373 K Ca^{2+} independent lpAFP exhibits slightly enhanced RMSD than the corresponding Ca^{2+} dependent hAFP. When the temperature is more elevated (at 473 K) hAFP shows greater RMSD than lpAFP. Thus greater structural distortions are observed in hAFP than lpAFP at higher temperature. This is also evident by the increased value of radius of gyration (Rg) and solvent accessible surface area (SASA) of hAFP than lpAFP at 473 K (Table 1). Rise of SASA is mostly contributed by the hydrophobic residues for both the proteins (Table 1). Up to 373 K, Rg as well as SASA are constant throughout the simulation for both the type II AFPs.

Table 1
Average values of backbone RMSD, Cα RMSF, Rg, total SASA, hydrophobic SASA, hydrophilic SASA, average number of protein–protein (intra molecular) and protein–solvent H-bonds in lpAFP and hAFP at different temperature simulations.

	277 K	298 K	373 K	423 K	473 K
lpAFP					
Backbone RMSD[a] (nm)	0.17	0.18	0.29	0.32	0.44
Cα RMSF[a] (nm)	0.08	0.13	0.15	0.17	0.25
Rg[a] (nm)	1.38	1.38	1.39	1.40	1.42
SASA[a] (nm^2)	87.36	86.65	89.33	93.48	96.64
Hydrophobic SASA (nm^2)	56.49	55.87	56.40	60.63	64.77
Hydrophilic SASA (nm^2)	30.87	30.50	31.09	31.71	31.87
Protein–protein H-bond	88.88	87.30	85.68	71.93	68.29
Protein–solvent H-bond	217.04	206.22	169.40	154.10	139.97
hAFP					
Backbone RMSD[a] (nm)	0.12	0.14	0.21	0.31	0.52
Cα RMSF[a] (nm)	0.07	0.08	0.10	0.21	0.39
Rg[a] (nm)	1.33	1.34	1.33	1.36	1.41
SASA[a] (nm^2)	87.00	88.10	87.27	93.75	100.21
Hydrophobic SASA (nm^2)	54.96	55.71	54.85	59.32	63.64
Hydrophilic SASA (nm^2)	32.03	32.38	32.23	34.42	36.56
Protein–protein H-bond	81.77	80.22	76.20	68.46	65.20
Protein–solvent H-bond	241.55	225.11	188.81	173.92	153.33

[a] RMSD, RMSF, Rg, SASA and H-bonds denote root mean square deviation, root mean square fluctuation, radius of gyration, solvent accessible surface area and hydrogen bonds respectively.

3.2. RMSF flexibility

A detailed picture of the residue mobility of lp and hAFP at various temperatures is illustrated by the plot of root mean square fluctuations (RMSF) of Cα atoms relative to the crystal structure (Fig. 2B(I) and (II)). The regions of greatest flexibility correspond to the termini of the proteins. Secondary structural elements generally exhibit low flexibility in both the proteins; peaks in RMSF profile are located in poorly structured loop region. At 277 K, the RMSFs observed for lp and hAFP exhibit more or less similar distribution and magnitude as seen in Fig. 2B. The RMSFs derived from the simulation agree well with the atomic mobility of the crystal structure as derived from B-factors. At 298 and 373 K, lpAFP exhibits slightly enhanced flexibility than the Ca^{2+} dependent hAFP. The overall flexibility of the two proteins has been calculated by the trace of the covariance matrix of atomic positional fluctuations. For lpAFP, we have obtained 1.718 nm^2 at 298 K and 3.054 nm^2 at 373 K, whereas for hAFP, we have 1.268 nm^2 at 298 K and 2.148 nm^2 at 373 K. These values also support increased flexibility of lpAFP over hAFP at 298 and 373 K. However at 473 K most of the residues of hAFP become highly mobile showing much greater fluctuation since hAFP mostly unfold at this temperature. Surprisingly, Ca^{2+} independent long snout AFP does not show significant enhancement of RMSF at this temperature.

3.3. Secondary structure

To gain further insight into the role of individual residues and secondary structures in these two types of AFPs, time dependent secondary structure fluctuations were analyzed using DSSP for lp and hAFP. Up to 373 K, most of the secondary structural elements from the starting crystal structure remain stable for both lp and hAFP. Supplementary Table 2 reports the average number of residues in a given secondary structure and in the corresponding crystal structure. When the average secondary structure content over time is considered, differences between lpAFP and hAFP are evident from 423 K onwards simulations, secondary structure particularly the β-sheet and α-helix content of hAFP decrease with temperature. This is particularly evident at 473 K. On the contrary, the α-helix content of lpAFP remains quite stable at 473 K up to 6.9 ns. Beyond that α-helix content decreases to some extent. The β-sheet becomes highly distorted from the beginning of the simulation. Therefore, at 473 K extent of denaturation of hAFP is more than that of lpAFp.

3.4. Coordination of Ca^{2+} ion

Sequence analysis suggests that hAFP contains one Ca^{2+} binding site that is present in the CTLDs [4]. The bound Ca^{2+} is coordinated by Gln89 $O^{\varepsilon 1}$, Asp91 $O^{\delta 2}$, Glu96 $O^{\varepsilon 1}$, Asn110 $O^{\delta 1}$ and Asp111 O and $O^{\delta 1}$ (Fig. 3A). Previous mutational results indicate that residues Thr93, Leu94, Thr95 and Thr112 are most critical for antifreeze activity [4]. All these residues lie in the vicinity of the bound Ca^{2+} ion. We have investigated the coordination of Ca^{2+} ions during the simulation time at all temperatures in hAFP (Table 2 and Fig. 3B) since it is very crucial for hAFP activity. We have found that Ca^{2+} coordination is maintained throughout the simulation up to 373 K, beyond that it gets disrupted.

3.5. Calcium binding loop region

Ca^{2+} dependent type II AFPs require Ca^{2+} for their antifreeze activity. It was suggested that ice binding site of hAFP chiefly consists of residues Asp91, Thr93, Thr95 and Glu96 that form a relatively flat surface to interact with ice [4]. These ice binding residues all belong to part of the Ca^{2+} binding loop region. In order to investigate whether presence of this particular ion favors Ca^{2+} dependent type II AFPs, we have critically analyzed the residues present in the vicinity of Ca^{2+}. Calcium binding loop of hAFP consists of residues Ala87–Asp111, while the corresponding region in Ca^{2+} independent lpAFP consists of residues Pro90–Asp114. Both the proteins display comparable flexibility of the calcium binding loop at 277 K. At 298, 373 and 423 K, the ion binding loop region of hAFP shows less flexibility than the corresponding sequence in the lpAFP (Supplementary Fig. 1A and B). When the temperature is elevated at 473 K, both the proteins display comparable flexibility. To analyze correlated variations localized within the Ca^{2+} binding loop region, we have constructed covariance matrices including only this particular region and calculated RMSIPs (Supplementary Table 3). Root-mean-square inner product measures the degree of overlap or similarity of eigenvector sets. RMSIP data in Supplementary Table

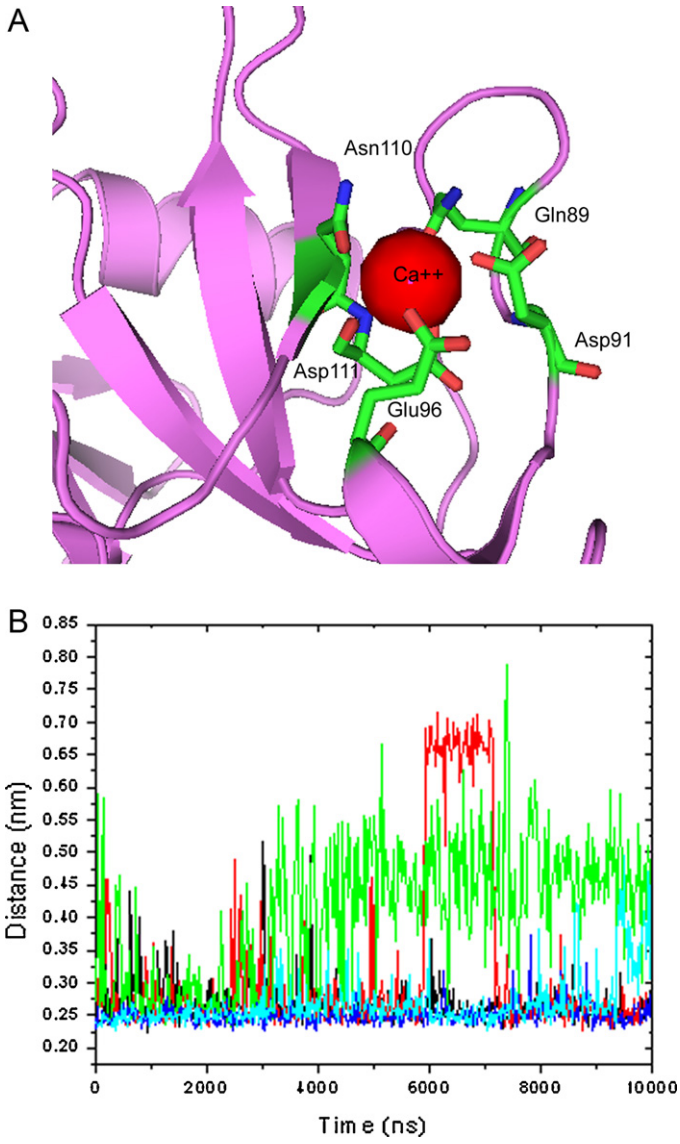

Fig. 3. (A) Ca^{2+} coordination sphere, the Ca^{2+} ion is coordinated with Gln89 $O^{\varepsilon 1}$, Asp91 $O^{\delta 1}$, Glu96 $O^{\varepsilon 1}$, Asn110 $O^{\delta 1}$, Asp111 O and Asp111 $O^{\delta 1}$ of hAFP. (B) Time evolution of coordinated bond formed between Asp 111 O and Ca^{2+} in hAFP at different temperatures. The color-coding scheme is as follows: 277 K (blue), 298 K (cyan), 373 K (black), 423 K (red) and 473 K (green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 2
Average values of distances of the coordinated bond formed between Ca^{2+} ion and the neighboring residues in hAFP at different temperature simulations.

	Crystal*	277 K	298 K	373 K	423 K	473 K
Gln89 $O^{\varepsilon 1}$	0.26	0.27	0.26	0.29	0.63	0.63
Asp91 $O^{\delta 1}$	0.25	0.27	0.25	0.26	0.36	0.41
Glu96 $O^{\varepsilon 1}$	0.28	0.28	0.27	0.26	0.28	0.71
Asn110 $O^{\delta 1}$	0.28	0.28	0.27	0.29	0.56	0.65
Asp111 O	0.24	0.25	0.25	0.26	0.33	0.40
Asp111 $O^{\delta 1}$	0.25	0.26	0.26	0.26	0.29	0.30

* The coordinated bond distances measured in the crystal structure is also shown. Distances are shown in nm.

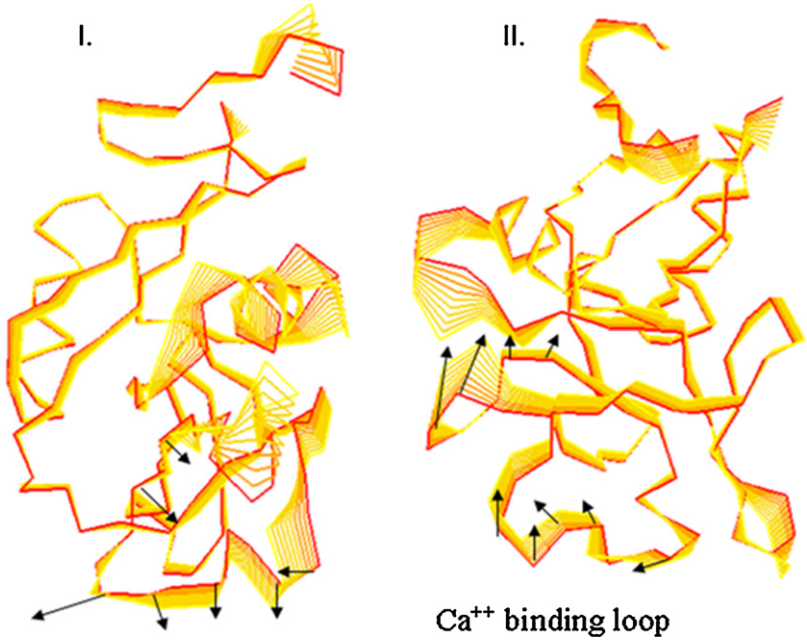

Fig. 4. Superimposition of 10 configurations obtained by projecting the Cα motion onto the first eigenvector at 298 K of (I) lpAFP and (II) hAFP. Arrows are used as qualitative indicators of the motion direction.

3 can be best described by splitting into two groups, those between simulations of the same protein and those between different proteins. The diagonal elements of Supplementary Table 3 represent the overlap within the individual trajectories. As expected the highest degree of overlap occurs within the individual trajectories obtained at different temperatures (diagonal elements of the Table). This Supplementary Table 3 also suggests that the RMSIPs of the Ca^{2+} binding loop region between the same proteins at different temperatures are high. On the contrary, the RMSIPs of different proteins at different temperatures are significantly low. Thus these data indicate that the essential subspaces explored by the simulations of the individual protein (but at different temperatures up to 373 K, higher temperature data not shown, as the coordination of Ca^{2+} ion is

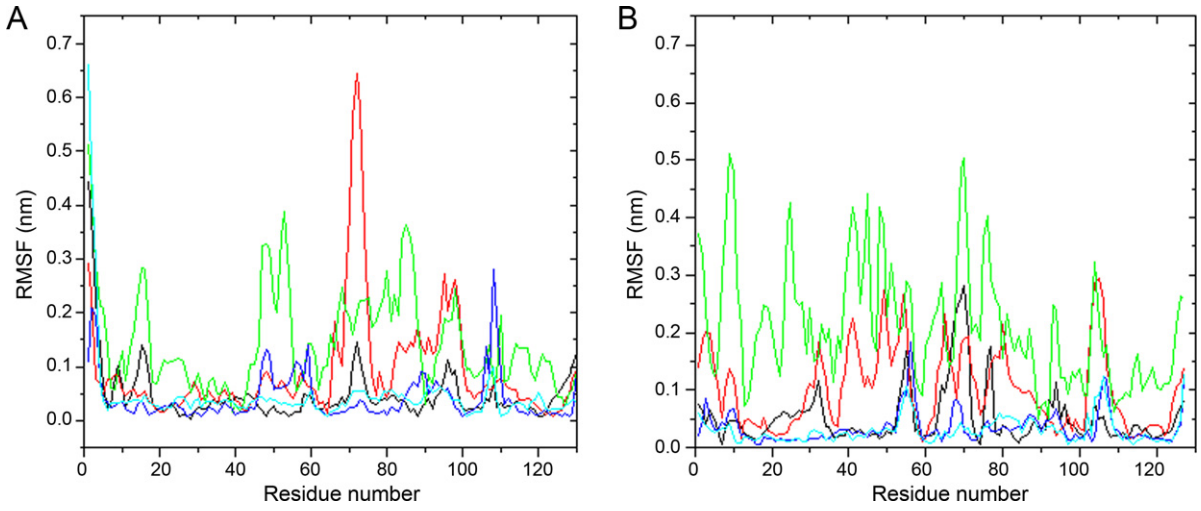

Fig. 5. RMSFs of Cα atoms projected along PC1 of (A) lpAFP and (B) hAFP at different temperature simulations.

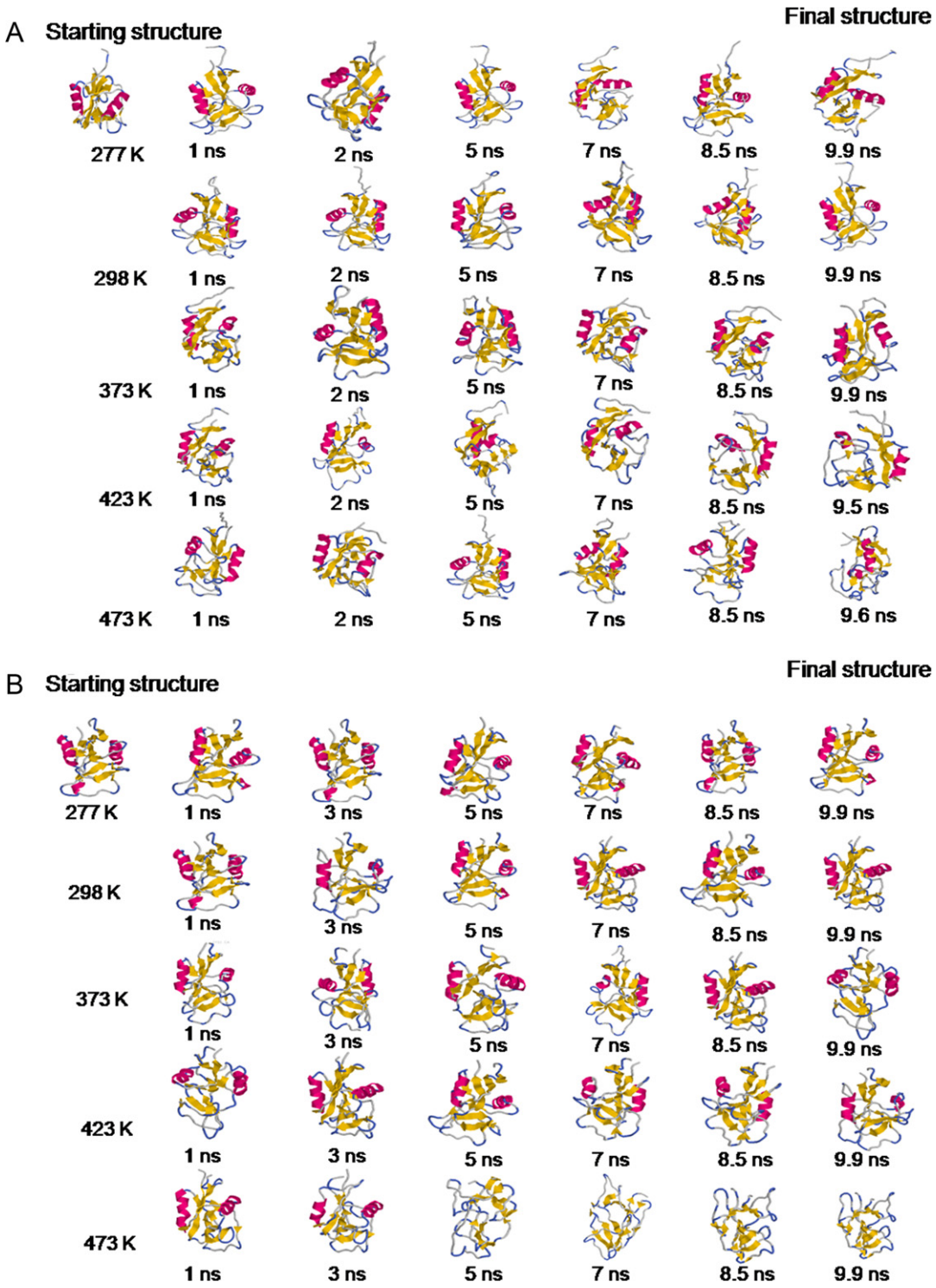

Fig. 6. Snapshots of (A) lpAFP and (B) hAFP evolved at different time points at different temperatures. The structures having similar secondary structures but generated at different temperatures have been put in a same column.

maintained up to 373 K) overlap significantly. The overall flexibility of Ca^{2+} binding loop region of the two proteins has been calculated by the trace of the diagonalized covariance matrix of the atomic positional fluctuations. We have obtained the following values for lpAFP: 0.139 nm^2 at 277 K, 0.291 nm^2 at 298 K, 0.412 nm^2 at 373 K. Similarly for hAFP we have, 0.103 nm^2 at 277 K, 0.165 nm^2 at 298 K and 0.210 nm^2 at 373 K again confirming the overall increased flexibility of lpAFP than hAFP at 298 and 373 K. These results also suggest that each protein has similar direction of motion at different simulated temperatures. However two proteins show different direction of motions at different simulated temperatures. The internal motion of the Ca^{2+} binding loop region has also been studied. Fig. 4 illustrates the motion along the first principal component (PC1) of lp and hAFP at 298 K, we have found that the direction of internal motion of Ca^{2+} binding loop region is quite different in the two proteins.

3.6. Correlated structure variation

The overall concerted motions of the two proteins at different temperatures are investigated by the essential dynamics analysis. The fluctuations of Cα atoms of lpAFP and hAFP projected on first principal component at different temperatures are shown in Fig. 5. The primary internal motions of both the proteins are similar to corresponding RMSF plots of Cα atoms shown in Fig. 2B(I) and (II). The dynamics of two proteins is best achieved via characterization of its phase space behavior. The projection of trajectories obtained at 298 K and 373 K onto the first two principal components (PC1 and PC2) shows the motion of two proteins in phase space (Supplementary Fig. 2). On these projections, we see clusters of stable states. lpAFP covers a larger region of phase space particularly along PC1 and PC2 plane compared to that covered by hAFP. Our observation thus supports the idea of higher flexibility of lpAFP than hAFP at 298 and 373 K. Two distinct clusters are seen in the phase space of hAFP at 373 K, the structures are well separated based on the presence of β1, the structures present in the right cluster contain β1 while the structures present in the left clusters are devoid of β1. This structural demarcation is not observed for lpAFP.

3.7. Unfolding profile

The sequence of unfolding events has been monitored for both the proteins at all temperatures (277 K, 298 K, 373 K, 423 K and 473 K). Up to 298 K, the structures of both the proteins are stable and remained close to the native structure during the whole simulation (Fig. 6A and B). At 373 K lpAFP maintains its native structure (Fig. 6A) whereas structural changes start to appear in hAFP at this temperature (Fig. 6B). β1 and β5 of hAFP denature at around 1.9 ns and 5.7 ns respectively. At 423 K, denaturation of β-sheets is also observed in lpAFP, however extent of loss of β-sheet is more pronounced in hAFP at this temperature. α helices of hAFP start to denature at this temperature, although they are maintained for the rest of the simulation in a reduced form. On the other hand, α helices are stably maintained in lpAFP at this temperature. At 473 K, at around 5 ns hAFP loses all the helices. The structure becomes highly coiled, only a few rearranged β-sheets remain in the structure. α helices of lpAFP are partially denatured at this temperature; however they are maintained in a reduced form for the rest of the simulation. Thus the study of unfolding pathways of lpAFP and hAFP reveal that, the rate of unfolding is faster in hAFP than lpAFP. Thus in our simulation the lowest temperature at which unfolding events first observed is 373 K, although it affects only hAFP. At 423 K and 473 K the unfolding process accelerates for both the proteins with again hAFP shows higher level of unfolding compared to lpAFP. Two dimensional plots of Rg and RMSD for both the proteins (Supplementary Fig. 3A and B) reveal distinct differences in the kinetic behavior of the two proteins at 473 K. The diagonal movement of Rg and RMSD in hAFP shows two-state kinetics representing simultaneous collapse of the chain and the formation of denatured structure. The plot exhibits approximately simultaneous increase in Rg and RMSD in hAFP. The rise of these two structural parameters is quite discrete in lpAFP which results in formation of two clusters. Although structural differences between the two clusters are not clearly evident. Both the clusters contain native one, intermingled with denatured ones.

3.8. Potential energy map

Potential energy of the two proteins at different points of the 473 K simulations is plotted as a function of PC1 and PC2 (Supplementary Fig. 4A and B). The energy landscape of the two proteins differs significantly. The potential energy landscape of lpAFP shows a sharp funnel shaped energy basin corresponding to the native state. hAFP on the other hand displays a more complex potential energy diagram indicating higher degree of heterogeneity compared to lpAFP.

4. Conclusion

We report here a comparative study of Ca^{2+} dependent and independent type II AFPs at five different temperatures. The overall flexibility calculated by the trace of the diagonalized covariance matrix displays similar flexibility of both the proteins at 277 K. However at higher temperatures (up to 373 K) lpAFP shows increased overall flexibility than hAFP. On the contrary, at much elevated temperature (473 K) the Ca^{2+} dependent hAFP shows greater flexibility that the corresponding independent one. Principal component analysis also indicates that the essential subspaces explored by the simulations of the two proteins at different temperatures are non-overlapping and they show significantly different directions of motion. In hAFP, Ca^{2+} coordination remains unaltered up to 373 K, beyond that it gets lost and the residues become highly mobile in nature. Ca^{2+} binding motif of hAFP shows less flexibility than the corresponding aligned motif of lpAFP up to 373 K. The study of unfolding pathways of lpAFP and hAFP reveal that the rate of unfolding is similar up to 373 K, beyond that hAFP shows faster unfolding than lpAFP. From our simulation it can be interpreted that Ca^{2+} plays vital role in maintaining the stability of hAFP. Ca^{2+} coordination is maintained up to 373 K, beyond 373 K, the Ca^{2+} coordination becomes destroyed that may affect the stability of the entire protein also, so that beyond 373 K, the rate of unfolding of hAFP is faster than lpAFP. In order to determine the exact role of Ca^{2+} ion much longer simulations are needed. Our study may help in elucidating the molecular basis of thermostability of two structurally similar proteins, which perform the same function in different manner, one in presence of Ca^{2+}, and the other in absence of Ca^{2+}.

Acknowledgement

This study was supported by grants from Department of Biotechnology Govt. of India.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.05.004.

References

[1] G. Chen, Z. Jia, Ice-binding surface of fish type III antifreeze, Biophysical Journal 77 (1999) 1602–1608.
[2] S. Kundu, D. Roy, Temperature-induced unfolding pathway of a type III antifreeze protein: insight from molecular dynamics simulation, Journal of Molecular Graphics and Modelling 27 (2008) 88–94.
[3] S. Kundu, D. Roy, Comparative structural studies of psychrophilic and mesophilic protein homologues by molecular dynamics simulation, Journal of Molecular Graphics and Modelling 27 (2009) 871–880.
[4] Y. Liu, Z. Li, Q. Lin, J. Kosinski, J. Seetharaman, J.M. Bujnicki, J. Sivaraman, C. Hew, Structure and evolutionary origin of Ca^{2+}-dependent herring type II antifreeze protein, PLoS One 6 (2007) e548.
[5] R. Wang, P. Zhang, Z. Gong, C.L. Hew, Expression of the antifreeze protein gene in transgenic goldfish (*Carassius auratus*) and its implication in cold adaptation, Molecular Marine Biology and Biotechnology 4 (1995) 20–26.
[6] C.L. Hew, P.L. Davies, G. Fletcher, Antifreeze protein gene transfer in Atlantic salmon, Molecular Marine Biology and Biotechnology 1 (1992) 309–317.
[7] M. Griffith, K.V. Ewart, Antifreeze proteins and their potential use in frozen foods, Biotechnology Advances 13 (1995) 375–402.
[8] J. Kaiser, New prospects for putting organs on ice, Science 295 (2002) 1015.
[9] K.V. Ewart, D.S. Yang, V.S. Ananthanarayanan, G.L. Fletcher, C.L. Hew, Ca^{2+}-dependent antifreeze proteins, modulation of conformation and activity by divalent metal ions, Journal of Biological Chemistry 271 (1996) 16627–16632.

[10] Y. Nishimiya, H. Kondo, M. Takamichi, H. Sugimoto, M. Suzuki, A. Miura, S. Tsuda, Crystal structure and mutational analysis of Ca^{2+}-independent type II antifreeze protein from long snout poacher, *Brachyopsis rostratus*, Journal of Molecular Biology 382 (2008) 734–746.
[11] E. Lindahl, B. Hess, D. van der Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, Journal of Molecular Modeling 7 (2001) 306–317.
[12] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, Journal of Computational Chemistry 26 (2005) 1701–1718.
[13] W.F. van Gunsteren, S.R. Billeter, A.A. Eising, P.H. Hunenberger, P. Kruger, A.E. Mark, W.R.P. Scott, I.G. Tironi, Biomolecular Simulation: The Gromos 96 Manual and User Guide, Hochschulverlag AG an der Zurich, Zurich, Switzerland, 1996.
[14] J.D. Thompson, D.G. Higgins, T.J. Gibson, ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice, Nucleic Acids Research 22 (1994) 4673–4680.
[15] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), Intermolecular Forces, D Reidel Publishing Company, Dordrecht, 1981, pp. 331–342.
[16] H.J.C. Berendsen, J.P.M. Postma, A. DiNola, J.R. Hakk, Molecular dynamics with coupling to an external bath, Journal of Chemical Physics 81 (1984) 3684–3690.
[17] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, Journal of Computational Chemistry 18 (1997) 1463–1472.
[18] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald: an N-log (N) method for Ewald sums in large systems, Journal of Chemical Physics 98 (1993) 10089–10093.
[19] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pederson, A smooth particle meshes Ewald potential, Journal of Chemical Physics 103 (1995) 8577–8592.
[20] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[21] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, H.J.C. Berendsen, Gromacs User Manual Version 3.3, 2005, http://www.gromacs.org
[22] R.A. Sayle, E.J. Millner-White, Rasmol-biomolecular graphics for all, Trends in Biochemical Sciences 20 (1995) 374–376.
[23] A. Amadei, A.B. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, Proteins 17 (1993) 412–425.
[24] A. Amadei, M.A. Ceruso, A. Di Nola, On the convergence of the conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations, Proteins 36 (1999) 419–424.
[25] B.L. de Groot, D.M.F. van Aalten, A. Amadei, H.J.C. Berendsen, The consistency of large concerted motions in proteins in molecular dynamics simulations, Biophysical Journal 71 (1996) 1707–1713.
[26] D.M.F. van Aalten, B.L. de Groot, J.B.C. Findlay, H.J.C. Berendsen, A. Amadei, A comparison of techniques for calculating protein essential dynamics, Journal of Computational Chemistry 18 (1997) 169–181.