J Mol Model (2005) 11: 481–488
DOI 10.1007/s00894-005-0256-7

ORIGINAL PAPER

Subramanian Jyothi · Sourajit M. Mustafi
Kandala V. R. Chary · Rajani R. Joshi

Structure prediction of a multi-domain EF-hand Ca^{2+} binding protein by PROPAINOR

Received: 16 May 2004 / Accepted: 21 February 2005 / Published online: 11 August 2005

Abstract PROPAINOR is a new algorithm developed for ab initio prediction of the 3D structures of proteins using knowledge-based nonparametric multivariate statistical methods. This algorithm is found to be most efficient in terms of computational simplicity and prediction accuracy for single-domain proteins as compared to other ab initio methods. In this paper, we have used the algorithm for the atomic structure prediction of a multi-domain (two-domain) calcium-binding protein, whose solution structure has been deposited in the PDB recently (PDB ID: 1JFK). We have studied the sensitivity of the predicted structure to NMR distance restraints with their incorporation as an additional input. Further, we have compared the predicted structures in both these cases with the NMR derived solution structure reported earlier. We have also validated the refined structure for proper stereochemistry and favorable packing environment with good results and elucidated the role of the central linker.

Keywords Computational protein structure prediction · Distance geometry · NMR · Nonparametric statistics

Introduction

In this paper, we present the three-dimensional (3D) structure prediction of a calcium-binding protein from *Entamoeba Histolytica* (hereafter referred to as EhCaBP) using the ab initio algorithm PROPAINOR (PROtein structure Prediction by AI and Nonparametric Regression) [1–3]. This protein of 134 amino-acid residues has long been believed to play a major role in the pathogenesis of amoebiasis. This monomeric protein shows homology with many calcium-binding proteins only in the calcium-binding loops and not in the inter-loop regions, which are suspected to be the sites that interact with other proteins. Chary and coworkers have determined the three-dimensional (3D) solution structure of this protein in its holo form (PDB ID: 1JFK) by NMR [4].

The PROPAINOR algorithm is derived from a knowledge-based nonparametric regression method [3], which solves the computational problem of protein 3D structure prediction using the primary sequence as a probabilistic programming problem. Apart from the nonrequirement of sequence-homology, the modularity and computational efficiency of this algorithm are its important features. The algorithm has been proved to be better in terms of prediction accuracy as compared to existing ab initio computational methods [1, 2].

Earlier, we have used the algorithm for the prediction of a large sample of proteins of length 70–150 amino acids with different protein folds. They included both α-helical and β-sheet dominant proteins. Most of them were earlier studied by either NMR or X-ray crystallography and hence could be good models to test the robustness of our algorithm [2, 3]. Further, we predicted the 3D structure of human seminal plasma inhibin (HSPI), a biologically important protein, using our refined PROPAINOR algorithm. This structure gave good validation results with respect to the available experimental observations. Moreover, functional studies on this structure provided significant explanation and insight into its binding activities and related biological and immunogenic functions and also offered new directions for its potential applications [1].

In this paper, we have undertaken the atomic structure prediction of a multi-domain (two-domain) protein,

S. Jyothi · R. R. Joshi (✉)
Department Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India
E-mail: rrj@math.iitb.ac.in

S. Jyothi
Statistics Division, Nicholas-Piramal (India) Ltd., Mumbai, India

S. M. Mustafi · K. V. R. Chary
Department of Chemical Sciences, Tata Institute of Fundamental Research, Mumbai, 400005, India

EhCaBP, based on the knowledge of only the protein primary sequence. We have studied the sensitivity of the predicted structure to NMR distance restraints with their incorporation as an additional input. Further, we have compared the predicted structures in both these cases with the NMR-derived solution structure reported earlier [4]. We have validated the refined structure of the protein for proper stereochemistry and favorable packing environment with good results (Section 4). We have obtained the calcium-binding sites in the predicted structure and analyzed the role of the central linker region (Section 5).

Methodology

We had developed a nonparametric statistical approach to regress the 3D-distances between residues (centroids or C^{α} atoms) as a function of the primary distances and some important features of the primary sequence [3, 5].

In our model, the unfolded polypeptide chain is represented as a linear sequence of amino acid residues. Each residue is modeled as a sphere having as its center the centroid of the residue and radius equal to the van der Waal radius as used in Ref. [6]. The 3D-distance between the C^{α} atoms of residue i and j denoted by d_{ij} is then estimated as a function of the corresponding *primary distance* p_{ij}, where

$$p_{ij} = r_i + 2r_i + 1 + 2r_i + 2 + \cdots + 2r_{j-1} + r_j$$

Here, r_i denotes the van der Waals radius of the ith residue; $i = 1, 2, \ldots, N$ (= the total number of amino acids in the primary chain). Computational experiments revealed that the primary distances alone are not sufficient to explain the random variation observed in the 3D-distances in native proteins. Hence, in our model, we use a sliding window approach incorporating some parameters (such as the length of the primary chain, the proportion of *hydophilic* residues, etc.) that were found to be statistically significant. The list of these parameters and detailed methodology are given in our earlier paper [2].

The training sample used for estimation consisted of proteins from 70 to 150 residues selected randomly from the PDB. Only high-resolution X-ray structures were used. As the idea was to develop an ab initio prediction method that did not rely on homology, homologous sequences having homology above 65% among these were discarded. The resulting training sample had 93 proteins. The proteins in this training set are such that much less than 1% of the pairwise sequence alignments show more than 40% identity.

We had fitted a nonparametric regression model [7] to this training sample using the SAS software. This model was used to estimate the short and medium-range 3D-distances d_{ij} between the ith and jth residues such that $|j - i| \leq 4$ for any given primary sequence in the validation sample. These distance estimates were used in the distance constraints in the Distance Geometry program dgsol [8] to predict the C^{α} coordinates. Some long-range distance restraints were also imposed using the fact that hydrophobic core residues would form a compact fold. These were obtained in terms of certain geometric constraints using the theoretical results on the *radius of gyration* and *hydrophobic* residue probability distribution [9].

It may be noted that our approach does not require any energy-function based optimization and is therefore computationally simpler than other computational methods of structure prediction. Comparison of the local aspects of the protein structures predicted by PROPAINOR with those by the threading method [10, 11] showed that the accuracy of the former is higher and more consistent [3]. The global RMSD for the protein structures predicted by PROPAINOR varied between 5 and 9 Å for proteins with sizes between 70 and 120 residues. The performance of our method was also validated against DRAGON and XPLOR [12] and was found to be better in accuracy and computational time than other extensively used distance based computational approaches [3].

The modified PROPAINOR algorithm

The sensitivity analysis of the protein structures predicted using PROPAINOR to *long-range* distances revealed that the addition of a very small number of correct *contacts* (derived from NMR data) vastly improves the accuracy of the predicted structures. The results of this analysis motivated us to look into the possibility of enhancing the accuracy of our algorithm by improving upon the distance estimation and structure computation procedures.

We have thus refined our basic PROPAINOR algorithm by optimal consideration of *short*, *medium* and *long-range* effects in the distance-constraints. As in the original PROPAINOR algorithm [3], the primary sequence is divided by a sliding window; namely the ith window includes residues at primary chain position numbers i to $i + 4$.; for $i = 1, 2, \ldots$ In case of short and medium range constraints, only the inter-residue distances within a window are considered whereas for long range those between ith and $(i+k)$th windows are taken into account for $k \geq 20$. The details are given in Refs. [2, 5], respectively.

The refined PROPAINOR algorithm [2] incorporates modification of long range constraints in terms of possibilities of β-turns; likely *contacts* between pairs of residues that are separated by more then 20 residues. The globular fold heuristics are also refined using β-turn propensity [13] and *hydrophilicity* profile plots. In particular, the residues in the sequence with a β-turn propensity greater than a specified threshold or falling in the regions of local maxima of the *hydrophilicity* profile plots are constrained to lie on the surface of the protein.

As the long-range constraints were obtained by heuristics, it was necessary to assign some measure of their accuracy. For this, the posterior probabilities were

measured by nonparametric discriminant analysis [14] of different subsets of the training sample. Nonparametric discriminant analysis was also used to refine the estimated of correlation between the primary and 3D-distances within each window and this refined estimate was used for the computation of the short and medium range distances by nonparametric regression [5]. The corresponding posterior probabilities were associated with the short and medium distance constraints.

The probabilities with each distance constraint were incorporated in the probabilistic programming formulation of the distance geometry problem. The validation of modified PROPAINOR had shown a significant reduction in RMSD, ranging between 4 and 7 Å [2].

Structure prediction of EhCaBP

We have used the refined version of PROPAINOR [2] to determine the 3D structure of a 134 amino-acid residue long Ca^{2+} binding protein from a protozoan (herein after abbreviated as *Eh*CaBP; $M^r = 14.7$ kDa). We have carried out two independent 3D-structure prediction experiments—one without any NMR restraints and the second with the inclusion of all the backbone NOE restraints given in Ref. [4]. These were 104 restraints, which included both the long and medium range distances. It is worth mentioning here that no NOE (NMR) based restraints involving side-chain spins were used in the prediction as the present algorithm calculates only the C^{α} trace of the protein to start with.

For predictions without NMR constraints, only the 526 short and medium-range distance restraints obtained from the nonparametric regression model and 123 distance restraints from the long-range interval estimation procedures of PROPAINOR were used. Modifications of these together with some additional distance constraints were used as per the NMR restrains were used in the second experiment.

Incorporation of NMR restrains

For example, if the regression estimate for a distance d_{mj} was (l, u) but the NOE restrain between the mth and jth residue was found to give a smaller interval then the latter was used in place of (l, u). If, because of the sliding window approach, the distance between residue numbers 20 and 28 is not estimated by regression, but the NOE restraint is available for the C^{α} atom for these residues then the latter would be added in the set of distance restrains.

Structure prediction without NMR restraints

The initial structure obtained without any NMR restraints showed a two-domain architecture for EhCaBP. As all the proteins in the training sample of PROPAINOR were single domain, we incorporated some more heuristics to refine this structure. These heuristics were defined for distance constraints between residues in the *hydrophobic* patches in the segments that fall in the neighborhood of the domain-separating residues, 67 and 68 in the predicted 3D fold.

The distance between the *hydrophobic* patches in the segments H_1 and H_4 (consisting of residues 37–42 and 97–101, respectively) was therefore restrained to lie in the range 28.74–58.96 Å . Also, the distance constraints between the *hydrophobic* patches of the segments H_2 and H_3 (consisting of the residues 58–61 and 77–84, respectively) was set in the range 14.74–58.96 Å, as these segments lie on the nearest boundaries of the two domains. An intermediate range (20.74–58.96 Å) was imposed between the *hydrophobic* patches in H_1 and H_3 and between H_2 and H_4. These numbers were obtained in terms of the theoretical estimate of the *radius of gyration* of the predicted structure (which was 14.74 Å) and certain heuristics. The heuristics were based on the same logic of *compactness* and as those devised for the single domain case except noting that the *hydrophobic* cores well within the two domains should be widely separated from each other.

Further, it is interesting to note that the prediction by the modified PROPAINOR showed the presence of two anti-parallel β-strands in each of the two domains (residues 16–18 and 52–54 in the N-terminal domain and residues 91–93 and 123–125 in the C-terminal domain). The distances between the residues were therefore restrained to pair up the corresponding β-strands in each domain into anti-parallel β-sheets using the β-it strand packing heuristics [2].

Twenty-five long-range distance constraints obtained by the above heuristics were added to the earlier set of 526 short and medium-range and 123 long-range restraints. Thus, twenty-two distinct solutions were obtained from 300 runs of **dgsol**. The optimal structure was selected as the one that was the best in terms of the bump distance and globular diameter restraint satisfaction, the theoretical *radius of gyration* and the *hydrophobic* residue distribution. We identify this structure as CaBP-1 (Fig. 1b). As seen in Fig. 1b, the 3D structure thus derived has two domains, one from 1 to 67 and the other from 68 to 134.

Structure prediction with the incorporation of NMR restraints

In our structure prediction with NMR restraints, we have used 104 NOE-derived backbone distance constraints [4]. Thirty distinct structures were obtained from 300 runs of **dgsol**. The optimal structure from these solutions was obtained using the same criteria as that of CaBP-1. We identify this structure as CaBP-0 (Fig. 1a). As seen in Fig.1a, the 3D structure thus derived also has two domains, one from 1 to 67 and the other from 68 to 134, as in the case of CaBP-1.

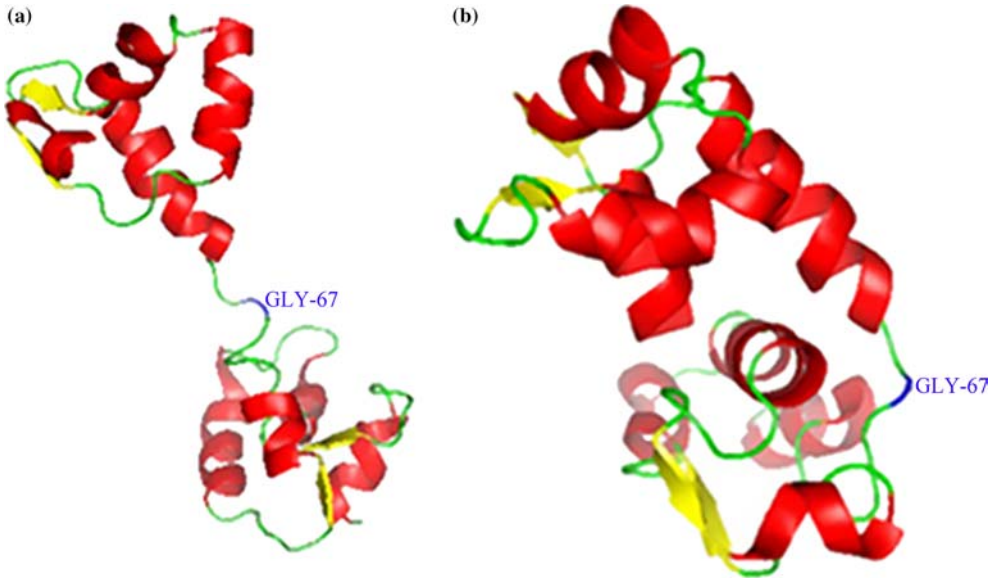

Fig. 1 Ribbon Model of the C^{α} structures of EhCaBP predicted by PROPAINOR. **a** With some NMR restraints (CaBP-0); **b** without any NMR restraints (CaBP-1). The α-helices are shown in pink while β-strands are shown in yellow. The figures were drawn using PyMol molecular graphics system (http://www.pymol.sourceforge.net)

Finally, the full atomic coordinates for CaBP-0 and CaBP-1 were then obtained using the program Max-Sprout [15]. These were further refined through restrained energy minimization using GROMOS96 [16] implemented in Swiss-PDBViewer [17].

Results

We compared the C^{α} traces of CaBP-0 and CaBP-1 with the NMR-derived solution structure of EhCaBP [4; PDB ID: 1jfk.pdb; identified as CaBP*]. The RMSD of CaBP-0 and CaBP-1 with CaBP* for the entire protein and also for both the individual domains are listed in Table 1. The superposition of the C^{α} traces of CaBP-0 and CaBP-1 on the structure CaBP* are shown in Figs. 2 and 3, respectively.

The structures for EhCaBP thus obtained by our method have double domain architecture with eight α-helices and two anti-parallel β-sheets (one in each domain), identical to CaBP* (Fig. 1). The *radius of gyration* and *hydrophobic core* residue distribution for CaBP-0 and CaBP-1 are given in Table 1. These show best agreement of the latter with the theoretical estimates [9] noting that the theoretical *radius of gyration* for EhCaBP would be 14.15 Å and 74 of the residues may be expected in the *hydrophobic core* region. We also computed the reliability measures [2] for the predicted structures. The precision score and the probability for CaBP-0 are 1750.94 and 0.8928, respectively. The same measures for CaBP-1 are 1766.69 and 0.8914, respectively. The reliability coefficient is 1.14 for both CaBP-0 and CaBP-1. This denotes a good prediction. All these indicate good prediction.

We have also tested and validated the refined atomic structures CaBP-0 and CaBP-1 of EhCaBP for good stereochemistry and packing using the WHAT_IF and PROCHECK suite of programs [18]. The overall G-factor for the structure CaBP-0, which is a carefully weighted average of all the tests performed by PROCHECK is−**0.55**, which indicates a well-refined structure. The corresponding score for CaBP-1 is−**0.58**, which is also good. (As per PROCHECK criteria a G-factor above −1.0 is considered good.) The G-factor for the structure CaBP* (1jfk) was **0.02**.

The predicted atomic structures CaBP-0 and CaBP-1 were also evaluated for the pattern of nonbonded interactions using the ERRAT program [19]. This gives a plot of the value of the error function versus position of a nine-residue sliding window. This plot also provides confidence limits based on comparison with statistics from highly refined crystal structures. Around 90% of the residues are found to have favorable nonbonded contacts (below 95% limit) in CaBP-0 (Fig. 4a). Around 80% of the residues in CaBP-1 satisfy this criterion (Fig. 4b).

Table 1 comparison of the predicted 3D structures derived with some (CaBP-0) and without any (CaBP-1) NMR restraints with that of the structure in reported in the pdb file1jfk [4]

Predicted structure	Radius of gyration (Å)	No. of residues in the core	RMSD (Å)		
			Entire protein	Fragment 1–67	Fragment 68–134
CaBP-0	15.76	47	4.94	2.54	2.27
CaBP-1	14.87	62	7.86	5.36	5.28

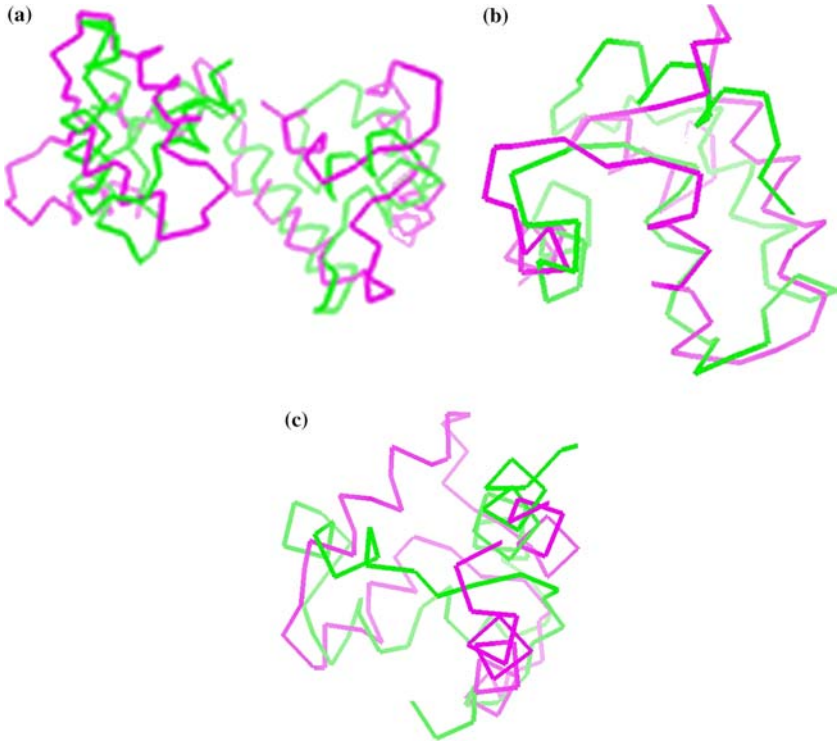

Fig. 2 Superposition of the predicted C^{α} chain structure CaBP-0 with C^{α} chain structure of the pbd file 1jfk (CaBP*): **a** entire protein; **b** fragment 1-67 (N-terminal domain); **c** fragment 68-134 (C-terminal domain). Here CaBP* (1jfk) is shown in *mazenda pink* color while CaBP-0 is in *green* color. The 3D-structure superimposing was done using the Molecular Operating Environment (*MOE*)—a software product of CCG (http://www.chemcomp.com)

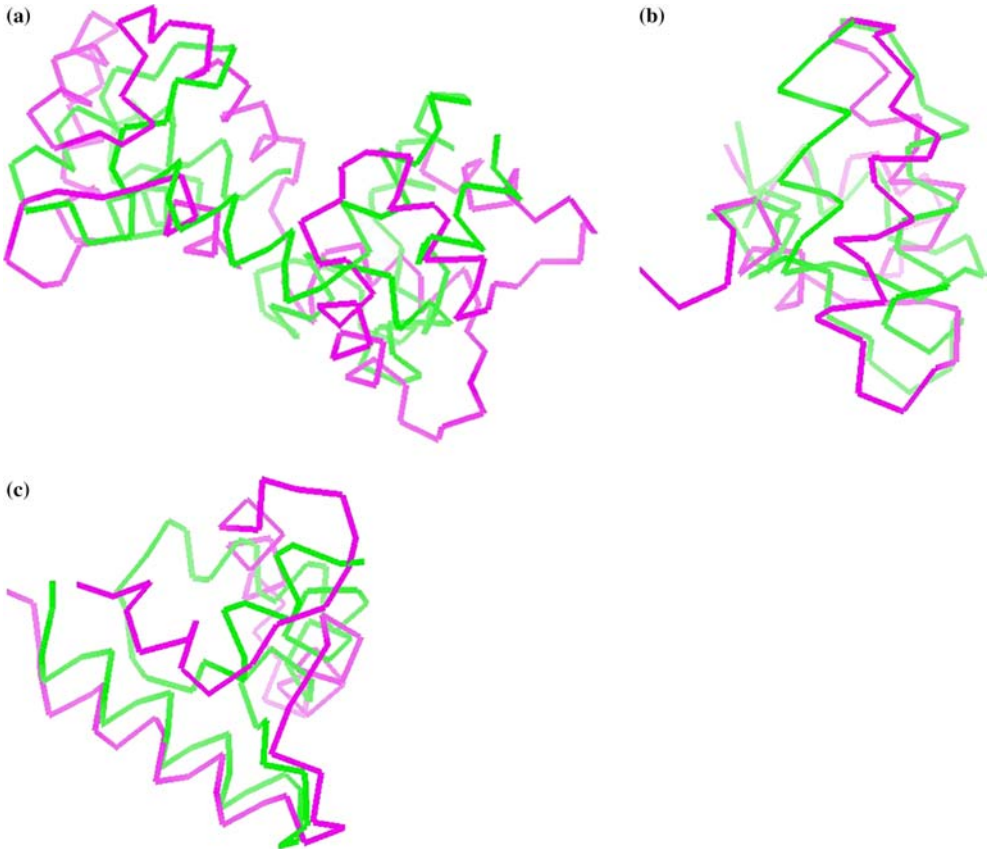

Fig. 3 Superposition of the predicted C^{α} chain structure CaBP-1 with C^{α} chain structure of the pbd file 1jfk (CaBP*) (1jfk): **a** entire protein; **b** fragment 1-67 (N-terminal domain); **c** fragment 68-134 (C-terminal domain). Here CaBP* (1jfk) is shown in *mazenda pink* color while CaBP-1 is in *green* color. The 3D-structure superimposing was done using the Molecular Operating Environment (*MOE*)—a software product of CCG (http://www.chemcomp.com)

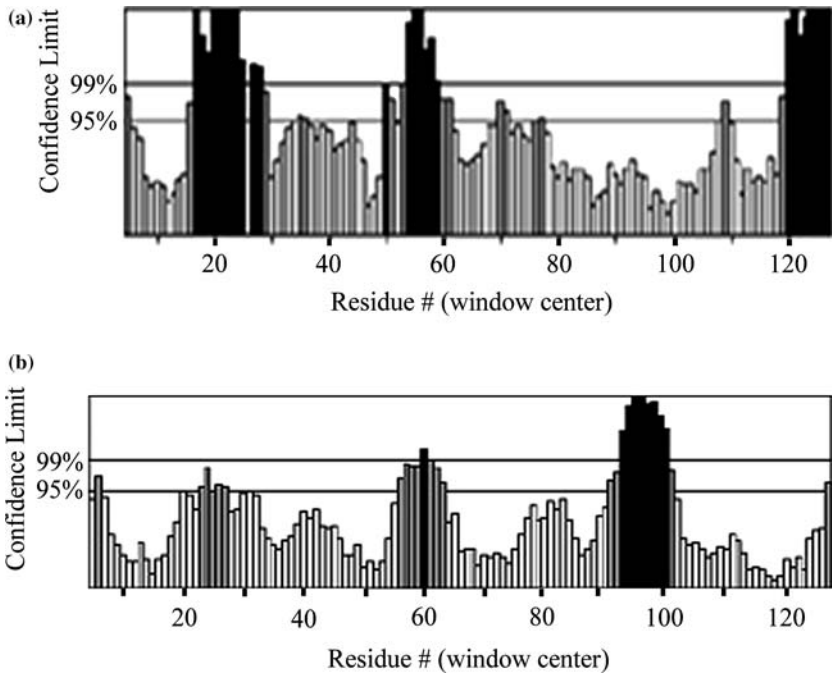

Fig. 4 ERRAT plots for predicted structures: **a** with some NMR constraints (CaBP-0) and **b** without any NMR constraints (CaBP-1). The *white* and *gray bars* show range of good refinement in terms of favorable nonbonded contacts

Calcium-binding loops and central linker

As illustrated in Fig. 1, the predicted structures for EhCaBP show that it is a dumbbell-shaped protein with two globular domains connected by a flexible linker region. Each domain contains a pair of helix -*loop-helix* motifs similar to the EF-hand motifs common to EF-hand calcium-binding proteins. This topology is found to be highly similar to calmodulin (CaM) despite the low sequence homology (approximately 29% for the entire sequence; the homology however is more than 70% in the four calcium-binding *loops*).

The advantages of our approach over the experimental studies are that we can incorporate different kinds of hypotheses about functional sites as well as search for them ab initio. As an illustrative example, this application is used here for the prediction of the four calcium-binding *loops* present in the protein.

The primary structure of the EhCaBP reveals four Ca^{2+} -binding loops. In order to predict the Ca^{2+} -binding sites on the computed structures, we placed four Ca^{2+} ions at randomly selected sites on the predicted structure CaBP-0 and carried out energy minimization for complex formation. The binding energies were computed using the program GRAMM [20, 21]. A binding energy of −3922.4 units was obtained when the Ca^{2+} ions were placed on the sites that were also the experimentally obtained calcium binding *loops* in EhCaBP, whereas the binding energies obtained when the Ca^{2+} ions were placed at randomly selected sites on the surface of the protein varied between −3861.2 and −3884.2 units. The 3D structures of the holo form (Ca^{2+} bound) of EhCaBP thus obtained are shown in Fig. 5. These show the optimal binding of Ca^{2+} ions in the predicted calcium-binding *loops* in both CaBP-0 and CaBP-1; these are in agreement with that of the structure CaBP* [4].

It may be noted that we had chosen the approach of trials at random sites, because our algorithm is ab initio and we want to show its potential without using any kind of homology or sequence alignments. However, as a cross check we have compared the results with those obtained by some standard programs on the Internet.

Comparison with PROSITE, etc

The calcium-binding sites identified by running PROSITE (http://www.expasy.org/prosite), BLOCKS (http://blocks.fhcrc.org/blocks) and BLASTP (http://www.ncbi.nlm.nih. gov/BLAST/Blast.cgi) on CaBP* (pdb file 1jfk) and CaBP-0 and CaBP-1 (our predicted structures with and without NMR) were also the same as those identified above, namely the following (output of PROSITE): 10–22: DvNGDGAVSyeEV; 46–58: DaDGNGEIDqnEF; 85–97: DvDGDGKLTkeEV and 117–129: DaNGDGYITleEF.

In each case, the *best*-it hit protein/protein family was 1vrk (in pdb) or its family. It gave the least '*E value*' ($= 2 \times 10^{-12}$) while using BLASTP, whereas these values for other eight well-aligned proteins were 6–10 times higher. About 22532 blocks were searched by BLOCKS and 3511793 (for CaBP-1) to 3579394 (for CaBP-0 and CaBP*) alignments were made. The '*combined e-values*' for each of the six *best-hit* protein families ranged from 0.00038 (0.00037 for CaBP-1) to 0.32; the least being for the Calcium binding proteins family of 1vrk.

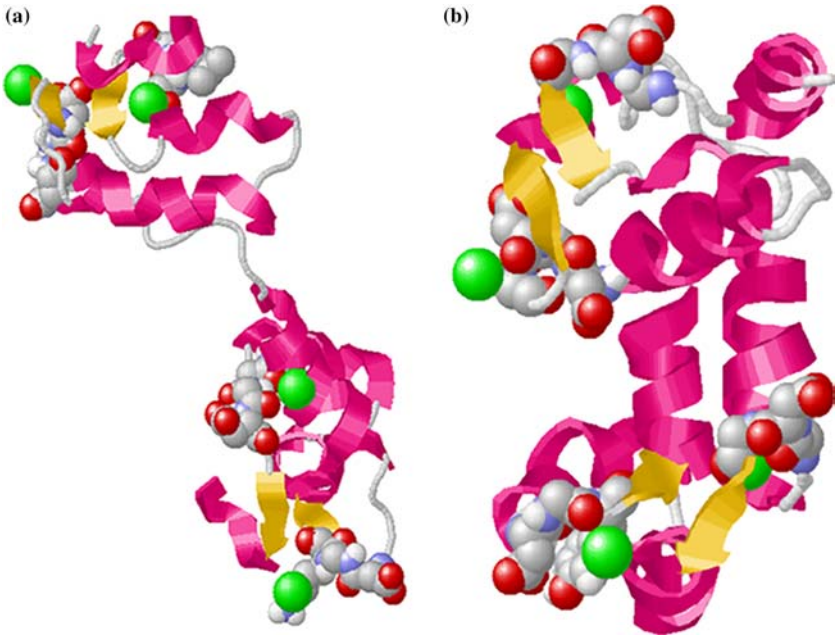

Fig. 5 The predicted 3D Structure of EhCaBP with bound Ca^{2+} ions. **a** CaBP-0, **b** CaBP-1. In both the structures, α-helices are shown in pink and the β-strands in yellow. Ca^{2+} ions are depicted as fluorescent green balls. Some of the residues in the calcium-binding *loops* are depicted in space-fill representation. The figures were drawn using PyMol molecular graphics system (http://www.pymol.sourceforge.net)

Role of the central linker

The two domains in EhCaBP are connected by a long central linker. This linker contains two Gly residues (G63 and G67), which impart the molecule with a high degree of flexibility. This central linker region is also thought to play a crucial role in the biological function of the molecule by binding to target peptides [22]. We have identified the possible active binding sites for EhCaBP in CaBP-0 and CaBP-1 using the algorithm proposed by Kolasker et al. [23], which is based on solvent accessibility. Interestingly, we find that the residues from 65 to 75 form part of the binding sites for EhCaBP in both the predicted structures CaBP-0 and CaBP-1. Incidentally, these residues have been proposed form part of the target peptide-binding site in the NMR derived 3D structure CaBP* [4].

The structural dissimilarity of our predicted structures with CaM are in good agreement with that of Chary and coworkers [4]. This structural dissimilarity and the greater flexibility of the central linker could be a possible reason for the novel signal transduction mechanism in EhCaBP, which is distinct from CaM.

Structural studies also indicate a more open structure for EhCaBP with a larger water-exposed surface area as compared to CaM. In this respect, we now note that the ratio of accessible atoms to buried atoms for CaM is 1.67, whereas for the predicted structures CaBP-0 and CaBP-1 it is 1.99 and 2.25, respectively. The same ratio is 2.19 for CaBP* (the minimized average NMR structure for EhCaBP obtained by Chary and coworkers [4]).

Discussion

In this paper, we have discussed the application of our refined PROPAINOR algorithm for the ab initio structure prediction of a multi-domain protein, namely a calcium-binding protein from *Entamoeba Histolytica*. We show here that even without any experimental restraints the double domain topology and the helix-packing features of EhCaBP are captured with clarity. Comparison of the structures predicted with and without NMR restraints with that of the NMR-derived solution structure [4] shows very significant topological match. The testing of both the predicted structures through structure-validation programs shows good stereochemistry and packing interactions.

Complexation of the predicted structures of EhCaBP with Ca^{2+} ions also confirms optimal binding in the calcium-binding *loops* as in the case of the NMR derived structural study [4]. Our predicted structure also supports the role of the flexible central linker connecting the two domains as the key factor in distinguishing the functions of EhCaBP as compared to those of CaM.

We have identified the active binding sites in EhCaBP and have shown that residues 65–75 in the central linker region are indeed part of the binding sites. This has been consistently the case, both for the structure predicted without any NMR restraints (CaBP-1) as well as for the structure predicted with a few long-range NMR distance restraints (CaBP-0).

The EhCaBP is found to be present only in pathogenic *Entamoeba Histolytica* but not in other non-pathogenic *Entamoeba* species. Hence, this protein is thought to play a major role in the pathogenesis of amoebiasis. Understanding the molecular basis of pathogenesis to find new targets for drug development would require further structural and functional studies on the proteins encoded by the pathogenic and non-pathogenic species of *Entamoeba*. Our immediate focus in this regard is to analyze its binding with melittin computationally [24] and also look into the possibilities of its immunosuppression and consequent inhibition of

its signal-transduction mechanism. Some of these computational results will be reported shortly.

Acknowledgments The facilities provided by the National Facility for High Field NMR, supported by Department of Science and Technology (DST) and Tata Institute of Fundamental Research, Mumbai, India, and computational research grants from the Department of Biotechnology (DBT), Council of Scientific and Industrial Research (CSIR), are gratefully acknowledged.

References

1. Joshi RR, Jyothi S (2002) J Mol Model 8:50–57
2. Joshi RR, Jyothi S (2003) Comput Biol Chem 27:241–252
3. Jyothi S, Joshi RR (2001) Comput Chem 25:283–299
4. Atreya HS, Sahu SC, Bhattacharya A, Chary KVR, Govil G (2001) Biochemistry 40:14392–14403
5. Jyothi S, Joshi RR (2003) Sankhya 65:593–611
6. Stanfel LE (1996) J Theor Biol 183:195–205
7. Buja A, Hastie T, Tibshirani R (1989) Ann Stat 17:453–555
8. More JJ, Wu Z (1999) J Global Optim 15:219–234
9. Kolinsi A, Skolnick J (1997) J Chem Phys 107:953–964
10. Sippl MJ (1990) J Mol Biol 213:859–883
11. Sippl MJ (1995) Curr Opin Struct Biol 5:229–235
12. Aszodi A, Gradwell MJ, Taylor WR (1995) J Mol Biol 251:308–326; DOI: 10.1006/jmbi.1995.0436
13. Chou PY, Fasman GD (1974) Biochemistry 13:211–222
14. Silverman BW (1986) Density estimation for statistics and data analysis. Chapman& Hall, New York
15. Holm L, Sander C (1991) J Mol Biol 218:183–194
16. van Gunsteren WF (1996) Biomolecular simulation: GROMOS96. BIOMOS bv, ETH Hönggerberg, Zürich
17. Guex N, Peitsch MC (1997) Electrophoresis 18:2714–2723
18. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G (1998) CABIOS 14:523–528
19. Colovos C, Yeates TO (1993) Protein Sci 2:1511–1519
20. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992) PNAS 89:2195–2199
21. Vakser IA (1995) Protein Eng 8:371–377
22. Anu KM, Gopal B, Satish PR, Bhattacharya S, Bhattacharya A, Murthy MRN, Surolia A (1999) FEBS Lett 461:19–24
23. Kolaskar AS, Kulkarni-Kale U (1999) Virology 261:31–42
24. Maulet Y, Cox JA (1983) Biochemistry 22:5680–5686