Secondary structure of a calcium binding protein (CaBP) from *Entamoeba histolytica*

Sarata C. Sahu\(^a\), A. Bhattacharya\(^b\), Kandala V.R. Chary\(^*\),*, Girjesh Govil\(^a\)

\(^a\) Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400 005, India
\(^b\) School of Life Sciences, Jawaharlal Nehru University, New Mehrauli Road, New Delhi 110 067, India

Received 3 August 1999

Abstract A calcium binding protein from *Entamoeba histolytica*, (EhCaBP, \(M_r \sim 15\) kDa) is the causative agent for amoebiosis and has a very low sequence homology (\(\sim 30\%\)) with other known CaBPs. Almost complete sequence specific resonance assignments for \(^1\)H, \(^{13}\)C and \(^{15}\)N spins in EhCaBP were obtained using double and triple resonance NMR experiments. Qualitative interpretation of the nuclear Overhauser enhancements, chemical shift indices and of hydrogen exchange rates threw valuable light upon the secondary structure of this protein. CaBP is found to have two globular domains each of which consists of two pairs of helix-loop-helix motifs. Though this protein has a very small sequence homology with calmodulins, the topological arrangement of the \(\alpha\)-helices and \(\beta\)-strands in EhCaBP resemble them.

© 1999 Federation of European Biochemical Societies.

Key words: Calcium binding protein; NMR assignment; Secondary structure; EF family; *Entamoeba histolytica*

1. Introduction

Calcium ions have diverse functions in biological systems. These range from biomineralization in bones, teeth and shells to a more complex role as an intra-cellular messenger. The physiological responses of Ca\(^{2+}\) are found to be mediated by a large number of Ca\(^{2+}\) binding proteins (CaBPs). These proteins can be classified into two general groups. One group includes extra-cellular enzymes and proteins that have an enhanced thermal stability or resistance to proteolytic degradation as a result of binding with Ca\(^{2+}\) ions. The other consists of a family of intra-cellular proteins that reversibly bind Ca\(^{2+}\) ions and thereby modulate the action of other proteins or enzymes. The second group is distinguished from the first in that its members have a common Ca\(^{2+}\) binding motif consisting of two helices which flank a loop of 12 contiguous residues. Such a structure is popularly known as helix-loop-helix (HLH) motif or ‘EF-hand’. Each of the 12 residues forming the loop plays an important role in defining the structure of the calcium binding site. Invariably, five of them are involved directly in providing oxygen ligands to the Ca\(^{2+}\) ion. The other residues provide hydrogen bonding via main chain NH groups to stabilize the geometry of the loop.

*Entamoeba histolytica*, a protozoan parasite, is the causative agent of amoebiosis and amoebic dysentery world-wide. It infects nearly 50 million people world-wide, resulting in about 50000 deaths every year [1]. Though ubiquitous in distribution, this parasite is more prevalent in tropical and subtropical regions. It can also invade extra-intestinal tissues such as liver and brain, leading to formation of abscesses. The mechanisms governing pathogenesis of this parasite are not clear yet. Calcium is thought to be involved in the pathogenetic mechanisms of amoebiasis [2,3]. Ravdin et al. [2] have reported that there is a transient increase in Ca\(^{2+}\) levels in target cells during the process of cell killing by *E. histolytica*. Ca\(^{2+}\) enhances the secretion of hydrolytic enzymes from *E. histolytica*. E. histolytica is known to have a Ca\(^{2+}\)-ATPase and inositol phosphates (IPs) (IP\(_3\) and IP\(_4\)) mobilizable internal Ca\(^{2+}\) stores. However, there is no information regarding various Ca\(^{2+}\) sensitive pathways and their roles in host parasite interaction in amoebiasis. In order to understand the mechanisms by which calcium affects virulence, a gene encoding a novel CaBP from *E. histolytica* (EhCaBP) has been isolated and characterized [4]. Though EhCaBP has four EF-hand domains similar to calcium signal transducing calmodulin (CaM), detailed analysis shows that it is functionally and structurally distinct from CaM and possibly involves a novel signal transduction pathway [5]. Analysis of different species of *Entamoeba* indicates that this protein may not be present in the non-pathogenic species *Entamoeba invadens* and *Entamoeba moshkovskii* and may be involved in pathogenesis [6].

EhCaBP is a 15 kDa monomeric protein showing homology only in the loops with other CaBPs. The inter-loop regions which are suspected to be the sites that interact with other proteins do not show homology. The three-dimensional (3D) structure of this protein is not yet known. As a step to characterize structure-function relationships of this protein, we have initiated 3D solution structure determination of the Ca\(^{2+}\) bound EhCaBP, using NMR. This paper presents details of \(^1\)H, \(^{13}\)C and \(^{15}\)N sequence specific resonance assignments and the secondary structure derived from the chemical shift indices, qualitative analysis of nuclear Overhauser effects and amide proton exchange rates.

2. Materials and methods

EhCaBP was expressed in the *Escherichia coli* BL21(DE3) strain containing a PET-3c expression system. The protocol described earlier [6] was modified so that expression could be carried out in the minimal (M9) media. Mid-log phase cells (OD = 0.60) were treated with isopropyl\(-\beta\)-d-galactoside for 4 h. EhCaBP gets expressed to an extent of \(\sim 30\%\) of the total cell proteins. EhCaBP was purified from the cell lysate by utilizing two of its major properties, namely, the heat stability and stable negative charge over a wide range of pH values. Thus, the simple purification protocol used involves heat treatment of induced recombinant *E. coli* cell lysate followed by ion-exchange chromatography. During heat treatment, the majority of *E. coli* proteins gets coagulated and may be removed by centrifugation. The EhCaBP remains in the supernatant, which was passed through a DE-52 anion-exchange column matrix. The binding of the protein to the matrix was carried out in the presence of EGTA. Subsequently, EhCaBP was

---

*Corresponding author. Fax: (91) (22) 2152110. E-mail: chary@tifr.res.in

0014-5793/99/$20.00 © 1999 Federation of European Biochemical Societies. All rights reserved.

PII: S0014-5793(99)01204-1
eluted specifically with the same buffer containing Ca\textsuperscript{2+} instead of EGTA. The purity was checked by SDS-PAGE. Isotopically \textsuperscript{15}N-labelled and doubly labelled (\textsuperscript{13}C and \textsuperscript{15}N) \textit{Eh}CaBP were also produced using M9 medium containing 1 g \textsuperscript{15}NH\textsubscript{4}Cl and 2.5 g \textsuperscript{13}C glucose per liter of culture as sole source of \textsuperscript{15}N and \textsuperscript{13}C, respectively. The yield of \textit{Eh}CaBP was \sim 50 mg of purified protein per liter of culture.

NMR measurements were carried out on a Varian Unity\textsuperscript{600} MHz NMR spectrometer equipped with a pulsed field gradient unit and triple resonance probe with actively shielded Z-gradients, operating at a $^1$H frequency of 600.051 MHz. NMR measurements were performed with a sample of 0.6 ml of 3 mM \textit{Eh}CaBP in 30 mM CaCl\textsubscript{2} and 50 mM deuterated Tris buffer, at an optimized pH (6.0) and temperature (35°C), either in 99.9\% H\textsubscript{2}O or in a mixed solvent of 90\% H\textsubscript{2}O and 10\% D\textsubscript{2}O. The two-dimensional (2D) experiments with the unlabeled \textit{Eh}CaBP in H\textsubscript{2}O include two quantum-filtered correlation spectroscopy \cite{7}, three quantum-filtered correlation spectroscopy \cite{8}, clean total correlation spectroscopy (clean-TOCSY) \cite{9} with a mixing time ($t\textsubscript{m}$) of 80 ms and nuclear Overhauser enhancement spectroscopy (NOESY) \cite{10} with a $t\textsubscript{m}$ of 150 ms. Spectra in 90\% H\textsubscript{2}O +10\% D\textsubscript{2}O included WGRAF-NOESY \cite{11} with a $t\textsubscript{m}$ of 150 ms and clean-TOCSY with a $t\textsubscript{m}$ of 100 ms. The experiments with uniformly \textsuperscript{15}N-labelled \textit{Eh}CaBP included 2D (\textsuperscript{15}N-\textsuperscript{1}H)-HSQC, 3D TOCSY-(\textsuperscript{15}N-\textsuperscript{1}H)-HSQC \cite{12} ($t\textsubscript{m}$ = 80 ms, with a DIPSI-2 mixing sequence \cite{13}), 3D NOESY-(\textsuperscript{15}N-\textsuperscript{1}H)-HSQC. Further, using the doubly labelled (both \textsuperscript{13}C and \textsuperscript{15}N) \textit{Eh}CaBP, the following triple resonance experiments were recorded: HNCA, HN(CO)CA, HNCO.

Fig. 1. The nOe connectivity diagram for \textit{Eh}CaBP.
CBCANH and CBCA(CO)NH spectra. Data transformation and processing were done on a Silicon Graphics workstation (R10000 based Indigo II Solid Impact Graphics) using the FELIX95.0/FE-
LIX97.0 software (Biosym Technologies).

Deuterium exchange studies on EhCaBP were carried out by re-
cording a series of one-dimensional (1D) 1H-NMR spectra and 2D
(15N, 1H)-HSQC spectra recorded immediately after the lyophilized
fully protonated U-15N-labelled EhCaBP had been dissolved in 2H2O.

For the estimation of J(HN-HK) values, a 3D HNHA [14] spectrum
was recorded with the U-15N-labelled EhCaBP.

Proton chemical shifts (N) were calibrated relative to 2,2-dimethyl-
2-silapentane-5-sulfonate (DSS) at 308 K (0.0 ppm). Carbon chemical
shifts were calibrated indirectly relative to DSS. Nitrogen shifts were
calibrated with respect to an external standard of 15N-labelled ammo-
nium chloride (2.9 M in 1 M HCl) set at 24.93 ppm [15]. All the
reported 1H shifts are derived from 3D NOESY-15N-1H-HSQC, while
15N Ns were measured from a well-resolved 2D 15N-1H-HSQC spec-
trum. 13C Ns were measured from 3D CBCA(CO)NH,
while 13C OS were measured from 3D HNCO.

3. Results and discussion

3.1. Resonance assignments

Sequence specific resonance assignments for nearly all 1HN,
15N, 13CN, 13CL, 13CO, 1Hk and 1Hl spins were obtained by the
concerted use of several triple resonance experiments, HNCA,
HN(CO)CA, HNCO, HNCACO, CBCANH and CBCA(CO)-
NH, as reported earlier [19]. From the knowledge of the
chemical shifts of the aforementioned spins, it was straightforward
to assign the side-chain 1H resonances. For this purpose, the
15N-edited 3D TOCSY-HSQC and 3D NOESY-HSQC
were used. Thus, we could almost complete the sequence spe-
cific resonance assignments for all the detectable protons. Fig.
1 shows the nuclear Overhauser enhancement (nOE) connec-
tivity diagram obtained from the 3D NOESY-HSQC spec-
trum. The resonances for M1, A2 and 15N spin of E3 at the
N-terminal end of the polypeptide chain could not be as-
signed. The chemical shifts thus obtained have been deposited in
the BioMagResBank (http://www.bmrb.wisc.edu) under BMRB accession number 4271. Further, we have used the 3D HNHA [14] spectrum with the 15N-labelled protein and estimated the J(HN-Hk) values in the case of 100 amino acid
residues.

3.2. Deuterium exchange study

Deuterium exchange studies from a series of 1D 1H-NMR
and 2D 15N, 1H-HSQC spectra at regular intervals have re-
vealed that the protein has a strong hydrophobic core. These
2D 15N, 1H-HSQC spectra were used to estimate the deute-
rium exchange rates. There are several slowly exchanging la-
bile proton resonances that show up even after 100 days of
addition of 99.99% 2H2O (data not shown here).

3.3. Secondary structure

Chemical shift differences from their random coil values of
13CN, 13CL, 13CO and 1Hk spins [16–18], qualitative interpreta-
tion of the short and medium range nOes, amide proton ex-
change rates and the knowledge of the J(HN-Hk) values [14]
throw valuable light on the secondary structure of a protein.

3.4. Chemical shift index (CSI)

The knowledge of the individual chemical shifts of 13CN,
13CL, 13CO and 1Hk spins in EhCaBP has been used to arrive
at the consensus CSI [18] (Fig. 2) and thus to identify the
individual secondary structure elements. A residue was as-
signed to a helical structure when two of the three 13CN,
13CO and 1Hk CSIs agreed with the one expected for an
α-helical structure (Fig. 2). On the other hand, a residue
was assigned to a β-strand when two of the three 13CN,
13CO and 1Hk CSIs satisfy with the one expected for a β-strand
(Fig. 2). The consensus plot reveals that EhCaBP possesses

Residue number

Fig. 2. Plot of the chemical shift indices for each amino acid residue in EhCaBP. +1 and −1 correspond to a low-field and high-field shift rela-
tive to a random coil chemical shift. The consensus secondary structural elements are indicated at the bottom.
3.5. α-Helical structures

The eight α-helical segments identified by the CSI analysis exhibit strong $d_{NN}$ and medium $d_{KN}$ connectivities. Most of the amino acid residues belonging to these α-helical segments expectedly show $3J(H^N-H^K)$ values which are less than 6 Hz. These stretches were further characterized by the appearance of medium $d_{ON}(i, i+3)$ and $d_{ON}(i, i+4)$ NOEs and absence of $d_{ON}(i, i+2)$ and $d_{ON}(i, i+4)$ connectivities. Further, deuterium exchange studies confirm these secondary structural segments. Most of the amide protons belonging to the α-helices exchange slowly.

3.6. Anti-parallel β-sheet structures

The four short β-strands consisting of three amino acids each identified by the CSI analysis exhibit strong $d_{KN}$ connectivities and large $3J(H^N-H^K)$ values in the range 7–10.5 Hz. All the amide protons belonging to β-strands expectedly exchange very slowly. At the end of the sequence specific resonance assignments, three $d_{ON}$ connectivities seen in the 2D NOESY spectrum in $^2$H$_2$O (see Fig. 3) have been characterized as arising from the amino acid residues belonging to the above mentioned β-strands. These connectivities, A16(H$^N$)-D54(H$^N$) (i.e. between H$^N$ of Ala-16 and Asp-54), K91(H$^N$)-T125(H$^N$) and T93(H$^N$)-Y123(H$^N$), unambiguously establish that β-strands 1 and 2, and 3 and 4 form two short anti-parallel β-sheet structures. Another expected $d_{ON}$ connectivity between S18(H$^N$) and E52(H$^N$), however, could not be seen because of the near degeneracy of the S18(H$^N$) and E52(H$^N$) resonances. The two anti-parallel β-sheet structures are further characterized from the observation of several long range NOEs in the 3D NOESY-(15N-1H)-HSQC spectrum. These include $d_{NN}(V17-I53)$, $d_{NN}(L92-I124)$, $d_{ON}(D54-V17)$, $d_{ON}(S18-I53)$, $d_{ON}(T125-L92)$ and $d_{ON}(I124-T93)$. Further, the deuterium exchange studies substantiate characterization of these anti-parallel β-sheet structures.

Fig. 3. A selected region of the 2D NOESY spectrum of EhCaBP recorded in 99.99% D$_2$O. This shows H$^N$-H$^N$ connectivities, seen between the individual β-strands forming two short anti-parallel β-sheets in EhCaBP.

Fig. 4. Amino acid sequence alignment for CaM (top row, 148 residues) and EhCaBP (bottom row, 134 residues). The Ca$^{2+}$ binding loop regions in both the sequences are underlined.
3.7. Ca\textsuperscript{2+} binding loops

There are four Ca\textsuperscript{2+} binding loops in \textit{EhCaBP}, which have been identified by a homology study [18]. Each of these loops consists of 12 contiguous residues, with invariant aspartate, glycine and glutamate residues at the first, sixth and 12th position. These loops are D10-E21, D46-E57, D86-E96 and D117-E128. Each of these loops is flanked by two of the eight K\textsuperscript{H}-helices discovered earlier by CSI analysis in the form of HLH motif or EF-hand. For example, helices 1 and 2 flank loop 1 (D10-E21), helices 3 and 4 flank loop 2 and so on. The four short \( \beta \)-strands consisting of three amino acids each form the amino acid residues at the seventh, eighth and ninth position of each of these Ca\textsuperscript{2+} binding loops. For example, the A16-V17-S18 amino acid stretch of \( \beta \)-strand 1 forms the amino acids at the seventh, eighth and ninth position of loop 1 and so on.

On the whole, \textit{EhCaBP} is thus found to have two globular domains, each of which consists of two pairs of HLH motifs, like in other homologous proteins. The two domains are connected by an extended conformation.

3.8. Comparison of the secondary structure with that of vertebrate Ca\textit{M}

The secondary structural features described above can be used to compare to those of the other known CaBPs. Two more CaBPs from \textit{E. histolytica} that have been reported are Ca\textsuperscript{2+}-ATPase and \textit{E. histolytica} CaM (\textit{EhCaM}). While Ca\textsuperscript{2+}-ATPase is thought to be involved in Ca\textsuperscript{2+} transport, \textit{EhCaM} has not been characterized yet. Hence, it is difficult to draw comparisons among these three CaBPs from \textit{E. histolytica}. However, biochemically \textit{EhCaBP} resembles vertebrate CaM [20]. The primary sequence of \textit{EhCaBP} has been compared with sequences in the protein database for identification of possible homologues. Although the overall similarity was only 30–35\%, the Ca\textsuperscript{2+} binding domains are highly homologous (> 75\% similarity) to a number of heterologous CaMs in contrast to the high degree of sequence homology among different CaMs. Despite of the lack of significant similarity in their primary sequences (Fig. 4), CaM and \textit{EhCaBP} have many features in common, e.g. four EF-hand domains, mobility shift in presence of Ca\textsuperscript{2+}, heat stability and cytoplasmic localization. These features suggest that \textit{EhCaBP} and CaM may have similar structural features and hence exhibit similar biochemical and physiological functions. In the light of this, it is interesting to compare their secondary structure elements. Fig. 5 shows the secondary structure topology of the \textit{EhCaBP} derived from preliminary DYANA calculations [21], wherein only sequential and hydrogen bond constraints have been used. The only long range connectivities that have been used are those between the \( \beta \)-strands 1 and 2 and \( \beta \)-strands 3 and 4. The most striking feature of this \textit{EhCaBP} structure is its similarity to the 3D fold of CaM [20]. The presence of eight \( \alpha \)-helices (1–8) and two short anti-parallel \( \beta \)-sheets is typical of the CaM fold. Each of the four Ca\textsuperscript{2+} binding loops is flanked by two each of the eight \( \alpha \)-helices mentioned above, in the form of a HLH motif or EF-hand. Each of these four HLH motifs together form independent domains (A and B) in pairs which are linked by a polypeptide segment. The linker between the two domains has 10 residues in \textit{EhCaBP}, while it consists of eight residues in CaM. On the other hand, the polypeptide chain between the loops 3 and 4 is four residues shorter in \textit{EhCaBP} as compared to that of CaM. Other differ-
ences between *EhCaBP* and CaM are with respect to the lengths of the α-helices. The α-helix 1, shortest in *EhCaBP*, is found to be only six units long, compared to 14 units corresponding to α-helix 1 in CaM. The other discrepancy is seen in the length of helix 8, which is seven units long (the second shortest helix) in *EhCaBP*, while it has nine amino acids in CaM.

In conclusion, the secondary structure of *EhCaBP* derived from NMR data reveals that even though this protein has a very low sequence homology with CaMs, its 3D fold is similar to that of other known CaMs. However, substantial differences are seen between individual secondary structure elements.

Acknowledgements: The NMR spectra were recorded at the National Facility for High Field NMR, TIFR, located at Mumbai. Support by the Department of Science and Technology, Government of India is gratefully acknowledged.

References