BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2011,** *50,* 181–193 181
DOI: 10.1021/bi101411q

Calmodulin-like Protein from *Entamoeba histolytica*: Solution Structure and Calcium-Binding Properties of a Partially Folded Protein[†]

Ashok K. Rout,[‡,||] Narendra Padhan,[§,||] R. P. Barnwal,[‡] A. Bhattacharya,[§] and Kandala V. R. Chary*[,‡]

[‡]*Tata Institute of Fundamental Research, Mumbai, India, and* [§]*Jawaharlal Nehru University, New Delhi, India.*
[||]*A.K.R. and N.P. have equally contributed to this work*

Received April 12, 2010; Revised Manuscript Received November 28, 2010

ABSTRACT: The mechanism of Ca^{2+}-signaling in the protozoan parasite *Entamoeba histolytica* is yet to be understood as many of the key regulators are still to be identified. *E. histolytica* encodes a number of multi-EF-hand Ca^{2+}-binding proteins (EhCaBPs). Functionally only one of these molecules, EhCaBP1, has been characterized to date. The calmodulin-like protein from *E. histolytica* (abbreviated as EhCaM or EhCaBP3) is a 17.23 kDa monomeric protein that shows maximum sequence identity with heterologous calmodulins (CaMs). Though CaM activity has been biochemically shown in *E. histolytica*, there are no reports on the presence of a typical CaM. In an attempt to understand the structural and functional similarity of EhCaM with CaM, we have determined the three-dimensional (3D) solution structure of EhCaM using NMR. The EhCaM has a well-folded N-terminal domain and an unstructured C-terminal counterpart. Further, it sequentially binds only two calcium ions, an unusual mode of Ca^{2+}-binding among the known CaBPs, notably both in the N-terminal domain of EhCaM. Further, EhCaM is present in the nucleus in addition to the cytoplasm as detected by immunofluorescence staining, unlike other EhCaBPs that are detected only in the cytoplasm. Therefore, this protein is likely to have a different function. The presence of unusual and a diverse set of CaBPs in *E. histolytica* suggests a distinct Ca^{2+}-signaling process in *E. histolytica*. The results reported here help in understanding the structure−function relationship of CaBPs including their Ca^{2+}-binding properties.

The nature, function, and variety of calcium-binding proteins (CaBPs)[1] found in the genome of *Entamoeba histolytica* suggest an intricate Ca^{2+}-signaling system in this protozoan parasite. *E. histolytica* is the causative agent of amoebiasis, a pandemic disease affecting millions of people throughout the world, particularly in tropical countries (*1*). One of the mechanisms by which this parasite causes cellular damage is through phagocytosis of epithelial cells, red blood cells (RBCs), and immune cells (*2*−*5*). Ca^{2+} signaling plays a pivotal role in the pathogenesis of *E. histolytica* and is involved in exocytosis and phagocytosis (*6*, *7*). Cells respond to an extracellular stimulus by a transient change in intracellular Ca^{2+} concentration, that is sensed by a set of different CaBPs, through downstream effectors (*8*). Most of the CaBPs consist of two domains separated by a flexible linker of varying length. Recent studies suggest that the two domains are functionally and structurally distinct. However, both of the domains are required for structural stability and full range of functional diversity (*9*). Each of these domains has two Ca^{2+}-binding EF-hand motifs. The EF-hand motif almost always occurs in pairs with the two Ca^{2+}-binding motifs stacked against one another in a face-to-face manner and forms a four-helix bundle that is packed together to make a hydrophobic core (*10*). The structural integrity of such domain is further stabilized by a short antiparallel β-sheet formed between the individual pairs of EF-loops. Each EF-hand motif consists of a contiguous stretch of 12 amino acid residues forming the loop, which plays an important role in defining the structure of the Ca^{2+}-binding site. To date, more than 300 three-dimensional (3D) structures have been solved both by NMR spectroscopy and by X-ray crystallography (*11*−*13*). In addition, there have been several attempts to understand the affinity, cooperativity, selectivity, and displacement of Ca^{2+} among the various EF-hand units present in a given protein and their role in the pathogenesis of different organisms such as *E. histolytica* and *Listeria*.

The genome of *E. histolytica* encodes 27 CaBPs with multiple EF-hand Ca^{2+}-binding domains. The majority of EhCaBPs are novel as no homologues have been found in other organisms. Therefore, these proteins are likely to be involved in amebic biology including pathogenesis (*14*). It has also been reported that a number of CaBPs are observed in *Plasmodium*, the parasite that causes malaria. This suggests that the presence of a diversity of CaBPs probably helps pathogens to adjust to different environmental conditions and survive in the host (*15*). The structural topology of EhCaBP1 and EhCaBP2 (1JFK and 2JNX, respectively) is similar to CaM. Further, these proteins are functionally

[†]This work was supported by Tata Institute of Fundamental Research (TIFR), Mumbai, India. N.P. thanks the University Grant Commission (UGC), India, for a research fellowship.

*Corresponding author. E-mail: chary@tifr.res.in. Phone: +91-22-2278 2489. Fax: +91-22-2280 4610.

[1]Abbreviations: *E. histolytica*, *Entamoeba histolytica*; CaM, calmodulin; EhCaM, calmodulin-like partially unfolded protein isolated from the genome of *E. histolytica*; BMRB, BioMagResBank; CaBP, calcium-binding protein; EF I, first EF-hand of EhCaM; EF II, second EF-hand of EhCaM; NMR, nuclear magnetic resonance; CARA, computer-aided resonance assignment; HSQC, heteronuclear single-quantum coherence; CPMG, Carr−Purcell−Meiboom−Gill; NOE, nuclear Overhauser effect; TOCSY, total correlation spectroscopy; NOESY, nuclear Overhauser enhancement spectroscopy; ITC, isothermal titration calorimetry; CD, circular dichroism; DSC, differential scanning calorimetry; PBS, phosphate-buffered saline; BSA, bovine serum albumin.

Published on Web 11/29/2010

quite different from each other and from CaM (*13*, *14*, *16*). While CaMs are highly conserved within the family, irrespective of species, EhCaM has very low sequence homology with EhCaBP1 and EhCaBP2. However, it has better homology (~50%) with heterologous CaM. It is expressed in trophozoites and belongs to the EF-hand Ca^{2+}-sensor protein superfamily. Though CaM activity has been biochemically shown in *E. histolytica*, there is no molecular evidence for the presence of a typical CaM in this organism (*17*). In this backdrop, we have carried out characterization of the EhCaM and study its Ca^{2+}-binding properties, with an intention that the similarities and differences in structural properties between CaM and EhCaM may provide clues to the latter's function. The EhCaM is found to be partially folded with a well-folded N-terminal domain and an unstructured C-terminal counterpart, first of its kind among the known CaBPs, Further, it sequentially binds two calcium ions, both of them to the N-terminal domain of EhCaM, an unusual mode of Ca^{2+} binding among the known CaBPs. This study further highlights the relative Ca^{2+}-binding affinities and specificities among the metal-binding sites and cellular distribution of EhCaM.

MATERIALS AND METHODS

Protein Expression and Purification. The EhCaM coding sequence was amplified by PCR using genomic DNA as a template. Gene-specific primers (forward, 5′GGGAAACATATGAGTGAACAAAAGAAGGTA3′; and reverse, 5′CGCGGATCCTTATTTGCTGGTAATCAATTTAAC3′) were used for PCR amplification. The product was then cloned into a pET-30a expression vector. The plasmid construct was then transformed into competent *Escherichia coli* BL21 (DE3) cells for overexpression and purification of recombinant EhCaM. The presence and orientation of the insert in the vector were confirmed by sequencing. The protocol used for overexpression and purification of EhCaM in its Ca^{2+}-bound form is as described in our earlier paper (*18*). The Ca^{2+}-free form of EhCaM was prepared by EGTA and EDTA treatment followed by dialysis with Chelex-100 treated buffer (50 mM Tris-HCl, 100 mM NaCl, pH = 7.0) and avoiding glass apparatus to take care of the trace amount of Ca^{2+} contamination.

Sequence Analysis. The amino acid sequence of EhCaM (accession number EAL46322) was retrieved from the *E. histolytica* genome database (http://www.ncbi.nlm.nih.gov) and used for the BLAST sequence similarity search against the *E. histolytica* genome database (12.5X coverage, that is the final assembled genome sequence) at The Institute of Genome Research (http://www.tigr.org/tdb/e2k1/eha1), the Wellcome Trust's Sanger Institute (http://www.sanger.ac.uk/projects/E_histolytica), and the National Center for Biological Information (http://www.ncbi.nlm.nih.gov). All editing of sequences, simple operations, and multiple sequence alignment were carried out using a suite of programs called "Bio-Edit" (Tom Hall, version 7.0). The phylogenetic tree was constructed using Dayhoff's amino acid matrix of the Phylip package using the Gene Bee server (http://www.genebee.msu.su).

Calcium-Binding Affinity of EhCaM. From the sequence analysis, EhCaM is found to have EF-hand motifs. To demonstrate that EhCaM indeed binds Ca^{2+}, like other EF-hand-containing CaBPs, we carried out several assays and biophysical analysis. These include ruthenium red staining (Supporting Information Figure S2A), $^{45}Ca^{2+}$-binding overlay assay (Supporting Information Figure S2B), isothermal titration calorimetry (ITC), and NMR spectroscopy.

ITC was used to quantify the binding energetics (macroscopic binding constant and enthalpy changes) associated with metal binding to the protein. ITC measurements were performed with a Microcal Omega titration calorimeter at 25 °C. Apo-EhCaM was prepared as described in the Protein Expression and Purification section above, centrifuged, and degassed prior to the titration with Ca^{2+}. Each titration consisted of injecting 3 μL aliquots of 10 mM Ca^{2+} solution (diluted from 1 M standard $CaCl_2$ solution supplied by Sigma-Aldrich Chemicals) into 145 μM protein solution (1.7 mL) after every 3 min to ensure that the titration peak returned to the baseline prior to the next injection. The ITC data were baseline corrected by titrating the buffer with Ca^{2+} with identical parameters used during the experiment and analyzed using the software ORIGIN (supplied with Omega Microcalorimeter). The amount of heat released per addition of the titrant was fitted to different models to find out the number of binding sites and the binding affinities of the protein. The experimental data were reproduced by repeating the same experiment three times with different protein samples of different concentrations.

In order to find out the residues involved in Ca^{2+} binding, we carried out Ca^{2+} titration with uniformly ^{15}N-labeled apo-EhCaBP and recorded a series of [$^{15}N-^{1}H$] HSQC spectra. The NMR sample for this purpose was prepared as described in the NMR Spectroscopy section later. The Ca^{2+} titration was carried out with standard 100 mM $CaCl_2$ (Fluka). The protein (apo-EhCaM) concentration was 1.0 mM in 50 mM Tris-HCl (pH = 7.0) and 100 mM NaCl. For each titration, an aliquot of 1 μL of the stock solution was added to the NMR tube containing the protein solution, mixed, equilibrated, and followed by recording 2D [$^{15}N-^{1}H$] HSQC at 30 °C. HSQC spectra thus recorded were processed with identical processing parameters. Integral volumes of the individual $^{15}N-^{1}H$ cross-peaks in the resultant spectra were measured using Felix 2002 and monitored during the course of NMR titration. The reverse titration (titrating the Ca^{2+}-loaded protein with EGTA) was also carried out to confirm the residual Ca^{2+}-binding sites, by monitoring the disappearance of downfield-shifted peaks.

Immunofluorescence Localization of EhCaM. In order to find out the biological role of this protein, the localization of the protein was checked first by immunofluorescence staining. *E. histolytica* cells, resuspended in warm incomplete TYI-S-33 medium, were transferred onto prewarmed, acetone-cleaned coverslips placed in a Petri-dish and allowed to adhere for 10 min at 37 °C. The culture medium was removed, and cells were fixed with 3.7% prewarmed paraformaldehyde (PFA) for 30 min. After fixation, the cells were permeabilized with 0.1% Triton X-100/PBS for 3 min. The cells were then washed with PBS and quenched for 30 min in PBS containing 50 mM NH_4Cl. The coverslips were blocked with 1% bovine serum albumin (BSA)/PBS for 30 min, followed by incubation with primary antibody at 37 °C for 1 h. The coverslips were washed with PBS followed by 1% BSA/PBS before incubation with secondary antibody for 30 min at 37 °C. Antibody dilutions used were EhCaM (polyclonal antibody raised in rabbit against recombinant protein) at 1:200, control bleed (prebleed) at 1:200, and anti-rabbit Alexa-Flour 555 (Molecular Probes) at 1:200. The fixed cells were incubated with 200 μg/mL Hoechst 33342 (Sigma, USA) for 15 min at room temperature to stain the DNA and then washed three times with PBS/BSA. The preparations were further washed with PBS and mounted on a glass slide using DABCO (1,4-diazbicyclo(2,2,2)octane; Sigma), 10 mg/mL in 80% glycerol.

The edges of the coverslips were sealed with nail paint to avoid drying. The cells were visualized using a laser scanning confocal microscope equipped with argon−krypton lasers (Leica Microsystems Heidelberg GmbH, Mannheim, Germany). This method has been repeated several times to get consistent results. Prebleed from rabbit used for raising antibody was used as negative control.

Subcellular Fractionation. The protocol used to separate the nuclear fraction from the cytoplasmic and the membrane fraction was slightly modified as described by Dey et al. (*19*). In brief, 10^7 cells growing in log phase were harvested at 280*g* for 7 min at 4 °C, and the cell pellet was washed three times with PBS. The washed pellet was resuspended in 2 mL of lysis buffer (10 mM HEPES, pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5% Nonidet P-40 detergent, and 7.4 μg/mL aprotinin, 2 μg/mL leupeptin, 1 μg/mL E-64, 10 μg/mL 7-amino-1-chloro-3-tosylamido-2-heptanone (N^{α}-(*p*-tosyl)lysine chloromethyl ketone)) and incubated on ice for 15 min. Nuclei were pelleted by centrifugation at 3000*g* for 15 min at 4 °C and washed with lysis buffer minus NP-40. The supernatant was further ultracentrifuged at 100000*g* for 30 min at 4 °C to obtain cytoplasmic (soup) and membrane (pellet) fractions. Nuclei pellet was resuspended in 50 μL of lysis buffer, and the protein content of each fraction was estimated by BCA assay.

Secondary Structural Propensity of EhCaM. The secondary structure of the protein was predicted by CD spectroscopy and later compared with the same information obtained using chemical shift index (CSI) values (*18*). The CD spectra were recorded on a JASCO J-810 spectropolarimeter, using 1 cm and 1 mm path length cuvettes for near-UV and far-UV CD, respectively. All spectra were recorded in 50 mM Tris buffer (pH 7.0) and 100 mM NaCl, in the presence or absence of EDTA as per the requirements of the experiment, and the appropriate blank was subtracted. Protein solutions of 40 and 75 μM were used for far- and near-UV CD spectra, respectively. The CD data thus obtained were expressed in terms of mean residue ellipticity [Θ], in deg cm^2 $dmol^{-1}$. The same sets of experiments were repeated three times with different samples to check the reproducibility of the data.

Thermal Stability of EhCaM from Differential Scanning Calorimetry (*DSC*). It is well established that metal ion binding to CaBPs leads to a stabilizing effect on the secondary and tertiary structures of the protein. In order to test whether Ca^{2+} binding indeed increases the thermal stability of EhCaM, heat denaturation measurements were carried out on a Microcal VP-DSC instrument. Apo and holo forms of EhCaM (150 μM) prepared in 50 mM Tris-HCl (pH = 7.0) and 100 mM NaCl were used in this study, and corresponding T_m values were estimated. Curves were corrected for the instrumental baseline obtained by heating the solvent used in the preparation of the protein sample. The data were reproduced by repeating the experiments twice with different protein samples.

NMR Spectroscopy. For NMR experiments on EhCaM, unlabeled, uniformly ^{15}N-labeled (U-^{15}N-EhCaM), uniformly ^{13}C/^{15}N doubly labeled (U-^{13}C/^{15}N-EhCaM), and uniformly ^{13}C-labeled (U-^{13}C-EhCaM) samples (50 mM Tris-HCl, pH = 7.0, 100 mM NaCl, 30 mM $CaCl_2$) were prepared in both 100% ^2H_2O and a mixed solvent of 90% H_2O and 10% ^2H_2O as described in our earlier paper (*18*). All NMR experiments were carried out at 30 °C with ∼1 mM protein samples on a Bruker Avance 800 MHz spectrometer equipped with a 5 mm triple-resonance cryogenic probe. The two-dimensional (2D) experiments with unlabeled EhCaM in ^2H_2O included clean total correlation spectroscopy (TOCSY) with a mixing time (τ_m) of 80 ms and nuclear Overhauser enhancement spectroscopy (NOESY) with τ_m of 120 ms.

Experiments with U-^{15}N or U-^{13}C included sensitivity-enhanced 2D [^{15}N−^{1}H] HSQC using pulsed-field gradients for coherence selection, 2D [^{13}C−^{1}H] HSQC, 3D TOCSY-[^{15}N−^{1}H] HSQC (T_m = 80 ms), 3D NOESY-[^{15}N−^{1}H] HSQC at two different mixing times (τ_m = 100 and 130 ms), 3D NOESY-[^{13}C−^{1}H] HSQC (τ_m = 100 ms), and GFT (3,2) HNHA (*20*). The experiments recorded with the U-^{13}C/^{15}N doubly labeled sample for backbone and side-chain assignments were as described earlier (*18*). All of the spectra were processed using Felix 2002 (Accelrys Inc., San Diego, CA) and analyzed by TATAPRO (*21*) and Cara (*22*). The ^{1}H chemical shifts were referenced with respect to an external standard 2,2-dimethyl-2-silapentene-5-sulfonate (DSS). ^{13}C and ^{15}N chemical shifts were calibrated indirectly (*23*). The chemical shift data (BMRB accession number 15563) were reported earlier as an assignment note (*18*).

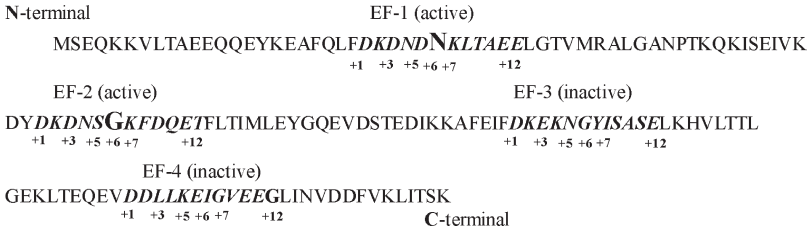
Constraints for Structure Calculation. (*a*) *Distance Constraints Derived from Nuclear Overhauser Enhancement and Dihedral Angle Constraints*. The 3D structure of EhCaM was calculated using the information derived from the NMR data, which included NOE distance restraints, backbone torsion angle restraints, restraints for hydrogen bonds, and Ca^{2+}−ligand coordination (described below in the Ca^{2+}−Ligand Distance Constraints section) using standard simulated annealing and torsion angle dynamics protocol with the program Cyana 3.0 beta (*24*, *25*). NOE cross-peaks from 3D [^{1}H−^{1}H]-NOESY-[(^{15}N/^{13}C−^{1}H)]-HSQC and 2D [^{1}H−^{1}H] NOESY spectra were identified, assigned, and linked in the NOESY spectra using the program CARA (*22*). The lower bound and upper bound distance constraints were set to 2.4 and 6.0 Å, respectively. A total of 942 distance constraints (on an average 12 constraints per residue), which include 281 intraresidue, 362 interresidue (sequential), 224 medium-range, and 75 long-range distance constraints, were used in the structure calculation.

Hydrogen bond constraints were added for residues that are involved in α-helical and β-sheet conformation, as characterized by the chemical shift indices, $^3J(H^N-H^{\alpha})$ values as derived from GFT (3,2)D HNHA (*24*), and deuterium exchange rates. A lower limit of 2.0 Å was used for H−O distance in all hydrogen bond constraints. A total 60 H-bond constraints were used in the structure calculation (Table 2).

The dihedral angle constraints were derived from TALOS using the $^{1}H^{\alpha}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shift values. A total of 116 (φ and ψ) dihedral angle constraints were used in the structure calculation of the EhCaM in its Ca^{2+}-bound state. One hundred structures were calculated, from which 20 structures with lowest target function values of ∼1 were selected for analysis. The quality of the lowest energy structures was validated by the program PROCHECK-NMR (*26*) and PSVS-1.3 web server. The structures were visualized, and the surface charge topology of the molecules was generated by using PyMOL (*27*) and NOC (*28*). Coordinates for the N-terminal domain of Ca^{2+}-bound (holo) EhCaM have been deposited in the PDB (2KTG), and the structural statistics for N-terminal EhCaM are given in Table 2. We could not determine the structure of the C-terminal domain of EhCaM due to insufficient NMR constraints.

(*b*) *Ca^{2+}−Ligand Distance Constraints*. It has been found from crystallographic data on several EF-hand proteins that in a canonical Ca^{2+}-binding loop there is a strong hydrogen bond

Scheme 1: Complete Amino Acid Sequence of EhCaM[a]



[a]Accession number EAL46322. The EF-hand motifs with the amino acid at the sixth position of the EF-hand loop used as a probe for cation binding are highlighted. Residues belonging to the active and inactive Ca^{2+}-binding loops are shown in italics.

between the side-chain carboxyl oxygen of Asp at position 1 and the backbone $^1H^N$ of the residue at position 6. This results in the downfield shift of the corresponding [$^{15}N-^1H$] cross-peak. It has also been found from crystallographic study on several EF-hand CaBPs that Ca^{2+} coordinates directly to the highly homologous residues at positions 1, 3, 5, 7, and 12 of individual Ca^{2+}-binding loops. In the structure calculation of the N-terminal domain of EhCaM, Ca^{2+} coordinates were incorporated in the last stages of structure refinement for the residues in the Ca^{2+}-binding loops as shown in Scheme 1. The upper bound distance constraints between Ca^{2+} and residues at positions 1, 3, 5, 7, and 12 were varied to fix the calcium atom around the individual regions of Ca^{2+}-binding loops.

(*c*) *^{15}N Relaxation Dynamics of EhCaM*. The spin−lattice relaxation time (T_1), spin−spin relaxation time (T_2), and $^{15}N-^1H$ nuclear Overhauser effect (NOE) were measured on Varian Unity$^+$ 600 MHz and Bruker Avance 800 MHz spectrometers for studying the dynamics of Ca^{2+}-bound EhCaM. The required ^{15}N relaxation data were obtained using a 2D [$^{15}N-^1H$] correlation spectroscopy based inversion recovery experiment (for the measurement of individual T_1 values) (*29*), Carr−Purcell−Meiboom−Gill sequence (for the measurement of individual T_2 values) (*30*), and steady-state [$^{15}N-^1H$] NOE experiment. In this endeavor, pulsed-field gradients were used for both coherence transfer and sensitivity enhancement (*30*, *31*). A total of ten T_1 experiments with different inversion recovery delays ranging from 16 to 1023 ms were recorded. Likewise, ten T_2 experiments were recorded with CPMG delays in the range from 10 to 190 ms. The [$^{15}N-^1H$] NOE spectra were recorded with and without proton saturation during relaxation delay, with 64 scans for each complex t_1 point. A 2.5 s period of proton saturation was used in the NOE experiment. The relaxation data thus obtained were analyzed using the Model-free approach (*32*) for the N-terminal domain alone as the C-terminal domain is unstructured.

For the analysis of R_1, R_2, and NOE, peak heights were measured from the NMR spectra using Felix software. Signal-to-noise ratios were calculated from the ratio of peak heights to base plane noise in the spectra acquired with a relaxation delay of 0.016 s for the R_1 experiments, 0.010 s for R_2 experiments, and without proton saturation for the NOE experiment. For R_1 and R_2 measurements, uncertainties in the peak heights were determined from duplicate spectra (recorded with relaxation delays of 0.016 and 0.195 s for R_1 and 0.010 and 0.110 s for R_2) (*33*, *34*). For NOE measurements, uncertainties in the peak heights were calculated from the standard deviation of the base plane noise in the spectra. Phenomenological R_1 and R_2 values and uncertainties were determined by nonlinear least-squares fitting for the experimental data to single exponential equations. The NOE values were determined as the ratios of the peak intensities measured from spectra acquired with and without proton saturation, and uncertainties in the NOE values were calculated by propagating the uncertainties in the peak heights (*35*). The model selection strategy used Monte Carlo numerical simulations for the best fit between the dynamical models and the experimental data. The relaxation data were fitted to Model-2 (S^2, τ) (*34*).

An exchange on a millisecond to microsecond time scale contributes to the observed transverse relaxation rates, $R_2 = R_{2,int} + R_{ex}$, where R_2 is the observed rate, $R_{2,int}$ is the intrinsic rate, and R_{ex} is the contribution from slow exchange. Thus, the measurement of R_{ex} provides a means to assess slow topological fluctuations in the ensemble. The contribution to R_2 arising from slow exchange on a millisecond to microsecond time scale (R_{ex}) was estimated from the relaxation data (R_1, R_2, and $^{15}N-^1H$ NOE) at two fields (600 and 800 MHz) (*36*).

(*d*) *HX Exchange Rates*. The hydrogen exchange experiment was carried out to get information about the structural stability of EhCaM. The HX exchange rates of individual backbone amide protons were measured indirectly by monitoring the decays of the nonoverlapping peaks. The decay of the individual peak volumes was fitted to a single exponential decay. The protection factors (PF) for each residue were calculated as $PF = K_{int}/K_{exp}$, where K_{int} and K_{exp} are the intrinsic and experimental exchange rates, respectively (*37*, *38*).

RESULTS

Expression of EhCaM in E. histolytica Trophozites. It is difficult to predict the function of a gene purely on the basis of computational analysis. To rule out the possibility that we were studying a pseudogene, because genomic DNA was used to amplify the EhCaM gene, the expression of EhCaM in *E. histolytica* cells was determined by RT-PCR (Supporting Information Figure S1A), Northern hybridization (Supporting Information Figure S1B), and Western blot (Supporting Information Figure S1C) using specific antibody against EhCaM. The details of RT-PCR and Northern hybridization are given in the Supporting Information. The analysis suggests that the gene is indeed expressed in *E. histolytica*, and the specific transcript of the expected size was observed. The immunofluorescence staining also indicated the expression of the EhCaM protein.

EhCaM Exists as a Monomer. The purity and molecular size of the protein were checked using gel-filtration chromatography. The peak corresponding to the protein showed relatively higher molecular weight compared to its expected value (data not shown). This may be because of partly unstructured polypeptide stretches in the protein. This ambiguity in the molecular size was resolved by MALDI-TOF mass spectrometry. The MALDI-TOF results showed three peaks in the spectrum: one at m/z 17254 corresponding to the monomeric state and the other two weak peaks corresponding to the doubly charged and dimeric species, respectively (Supporting Information Figure S1D).

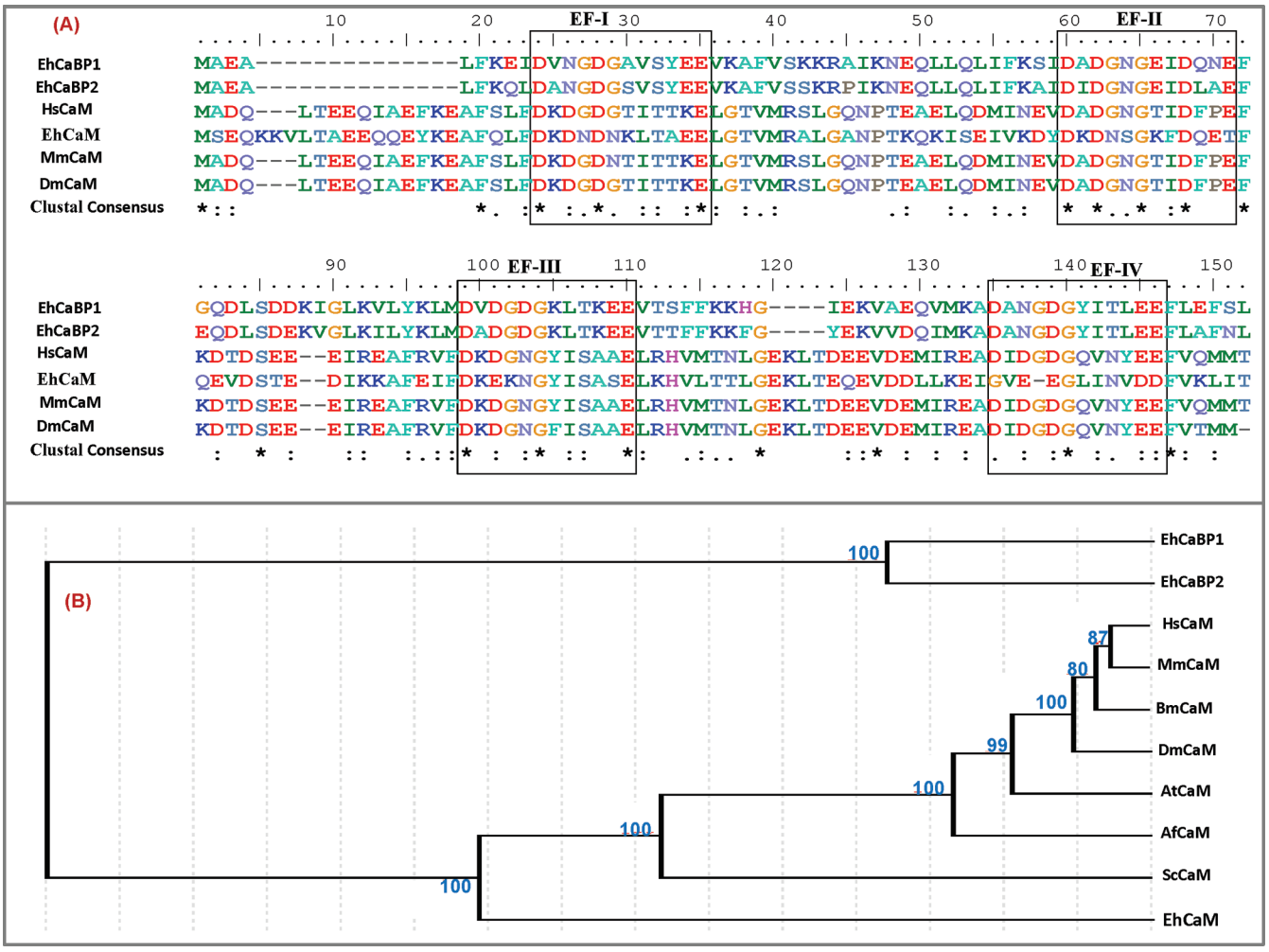


FIGURE 1: (A) Multiple sequence alignments (ClustalW) of the calcium-binding proteins. Asterisks represent conserved residues; dots represent similar residues. The EF-hand loops are boxed. Numbers indicate the positions of the amino acid residues in each sequence. Species names are abbreviated: Hs, *Homo sapiens*; Eh, *Entamoeba histolytica*; Mm, *Mus musculus*; Dm, *Drosophila melanogaster*. CaM, calmodulin; CaBP, calcium-binding protein. (B) Phylogenetic analysis of the calcium-binding proteins. Molecular phylogeny of EF-hand domains. The phylogenetic trees were constructed using the Dayhoff amino acid distance matrix. GenBank accession numbers for full-length proteins used in this analysis are as follows: TcCaM, *Trypanosoma cruzi* calmodulin (XP_808089); MmCaM, *M. musculus* calmodulin (AAA66182); HsCaM, *H. sapiens* calmodulin (AAD45181); DmCaM, *D. melanogaster* CaM (NP_523710); AfCaM, *Aspergillus fumigatus* calmodulin (EDP50378); BmCaM, *Bombyx mori* CAM5 (NP_001040234); AtCaM, *Arabidopsis thaliana* calmodulin 5 (NP_180271); ScCaM, *Saccharomyces cerevisiae* calmodulin (CAA55612); EhCaBP1, *E. histolytica* CaBP1 (EAL48959); EhCaBP2, *E. histolytica* (EAL51694); EhCaM, *E. histolytica* (EAL46322). The bootstrap values are indicated at the nodes and were based on 100 replicates. A value close to 100 represents a more significant grouping.

These observations suggest that the protein exists largely in the monomeric state in solution under the chosen experimental conditions. Further, the monomeric state of EhCaM is also supported from the overall tumbling time, which is as expected ($\tau_c = 9.6$ ns; discussed in the NMR Relaxation section).

Divergency in the Sequence of EhCaM. Based on the primary sequence and motif analysis, 12 of the 27 EhCaBPs (each around 150 amino acid residues long) are expected to be involved in Ca^{2+} binding. Sequence analysis of EhCaM carried out using the sequence alignment algorithm ClustalW (Figure 1A) did not show any significant sequence similarity among other CaBPs. The phylogenetic tree constructed using the aligned sequences (Figure 1B) and Dayhoff's amino acid distance matrix analysis showed that EhCaBP1 and EhCaBP2 are the only closely related CaBPs compared to the rest present in the genome and form a separate cluster. Interestingly, EhCaM is found to be closer to known calcium sensors such as CaM. These observations suggest that the EhCaM is a CaM-like protein that might have diverged considerably during evolution.

EhCaM Localization. Subcellular distribution of EhCaM was investigated by both biochemical and immunofluorescence methods. The EhCaM was found in all of the three subcelluar fractions (nucleus, membrane, and cytoplasm). This result was validated by cell fractionation studies. *E. histolytica* lysate was separated into cytoplasmic, membrane, and nuclear fractions as described under Materials and Methods. The protein in these fractions was identified with specific antibody in Western blot analysis. Bands of appropriate sizes were visible in all of the three subcellular fractions (Figure 2 inset).

EhCaM Is Predominantly α-Helical. The CD experiments were performed with EhCaM to characterize the percentage of secondary and tertiary structures that can be induced upon binding to Ca^{2+}. The changes in the molar ellipticity in far-UV CD spectra were used to characterize the changes in the secondary structure of the protein, and similar changes in the near-UV CD spectra were used to characterize the changes in its tertiary structure (*39*). Panels A and B of Figure 3 show the far- and near-UV regions of the CD spectra for apo and holo forms of

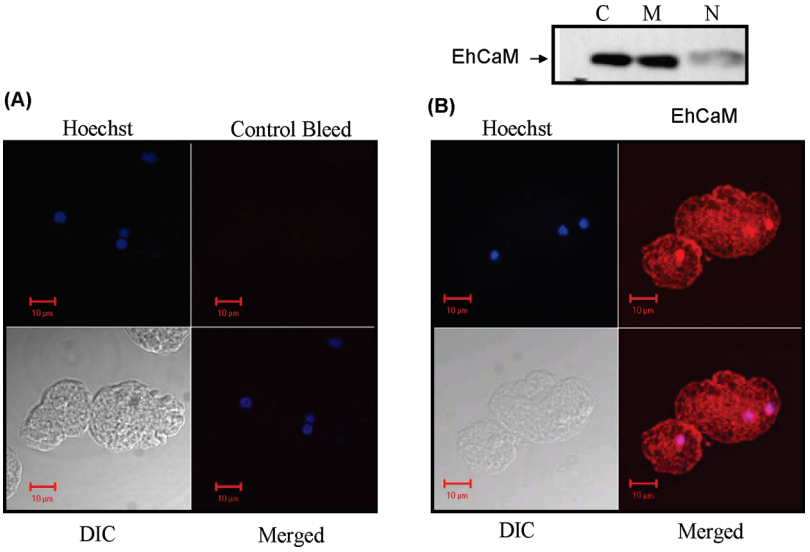


FIGURE 2: Localization of EhCaM in *E. histolytica* trophozites. Immunostaining: Cells were grown for 48 h and were then transferred to prewarmed, acetone-washed coverslips for 10 min at 37 °C. The attached cells were fixed with 3.7% paraformaldehyde/PBS, permeabilized with 0.1% Triton X-100/PBS, and then stained with various antibodies as indicated. Secondary antibody used was anti-mouse/rabbit Alexa-Flour 488. Scale bars represent 10 μm. Hoechst was used to stain the nuclei of the trophozites. Frame A: Control antibody. Frame B: Anti-EhCaM antibody. Inset: The *E. histolytica* whole cell lysate, prepared from mid-log-phase cells, was fractionated biochemically into cytoplasmic, membrane, and nuclear fractions as described in Materials and Methods. Fifty micrograms of proteins of indicated fractions was separated on a SDS−PAGE, electrophoretically transferred to PVDF membranes, and then immunostained with anti-EhCaM antibody. C, cytoplasmic fraction; M, membrane fraction; N, nuclear fraction.

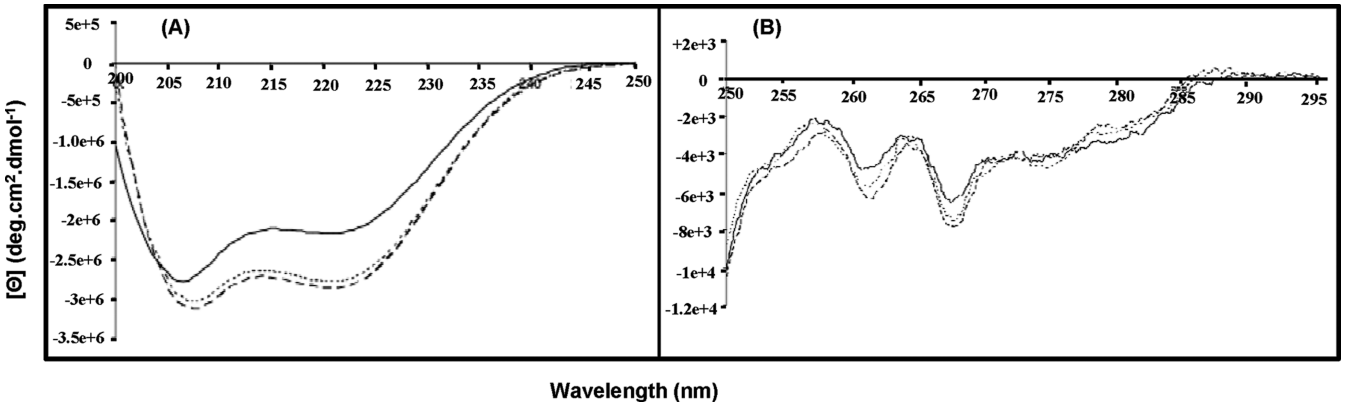


FIGURE 3: (A) Far-UV CD spectra of apo- (solid line) and holo- (dotted line) EhCaM, respectively. The concentration of EhCaM was 40 μM in 50 mM Tris-HCl (pH = 7.0) containing 100 mM NaCl. (B) Near-UV CD spectra recorded during the Ca^{2+} titration with apo-EhCaM. The protein concentration was 75 μM in 50 mM Tris-HCl (pH = 7.0) containing 100 mM NaCl. A significant change in going from apo to holo form suggests that the addition of Ca^{2+} brings the protein from a molten globule state to a well-folded state.

EhCaM, respectively. As is evident in Figure 3A, the far-UV CD spectrum of apo-EhCaM exhibits a minimum near 208 ($\pi-\pi^*$) and 222 ($n-\pi^*$) nm that is characteristic of α-helical conformation. Upon addition of Ca^{2+}, the helicity increased. These observations reveal that Ca^{2+} indeed induces helicity in EhCaM that can be calculated from the molar ellipticity at 208 and 222 nm (Θ_{208} nm and Θ_{222} nm). Though EhCaM is devoid of Trp, it is rich in Tyr and Phe (4 and 7 residues, respectively), with the bands for the corresponding residues well resolved in the near-UV CD spectra. The spectrum shows two negative broad bands at 278 and 285 nm that correspond to Tyr. The vibronic bands arising from Phe residues are well resolved with minima at 262 and 268 nm. Large changes seen in this region of the spectrum reveal that the protein undergoes significant conformational change in its tertiary structure upon Ca^{2+} binding.

Ca^{2+} Binding Increases the Thermal Stability. DSC experiments with EhCaM were carried out to characterize its thermal stability in both its apo and holo forms. The peak maximum of apo-EhCaM (Supporting Information Figure S3A) (transition temperature, T_m = 34.28 ± 0.05 °C) was found to be almost 32 °C lower than the peak maximum of the Ca^{2+}-bound form (Supporting Information Figure S3B; T_m = 66.53± 0.03 °C), indicating that Ca^{2+} binding enhances thermal stability. The transition peaks did not fully reappear upon rescanning each of the samples. After the fifth scan the peaks were not visible, suggesting that irreversible unfolding may have occurred due to aggregation and/or denaturation. The DSC thermogram of apo-EhCaM is found to be broad as compared to Ca^{2+}-bound EhCaM, suggesting that the unfolding of apo-EhCaM is a slow process. In contrast, Ca^{2+}-bound EhCaM showed a sharp transition peak indicating that it unfolds rapidly. A sharp one-step transition also suggests that the unfolding occurs in a single step, unlike in the case of CaM, in which the two domains were reported to unfold independently (*40*). The spectra, after baseline correction and concentration normalization, optimally fitted to a two-state model. The optimal fitting parameters are the

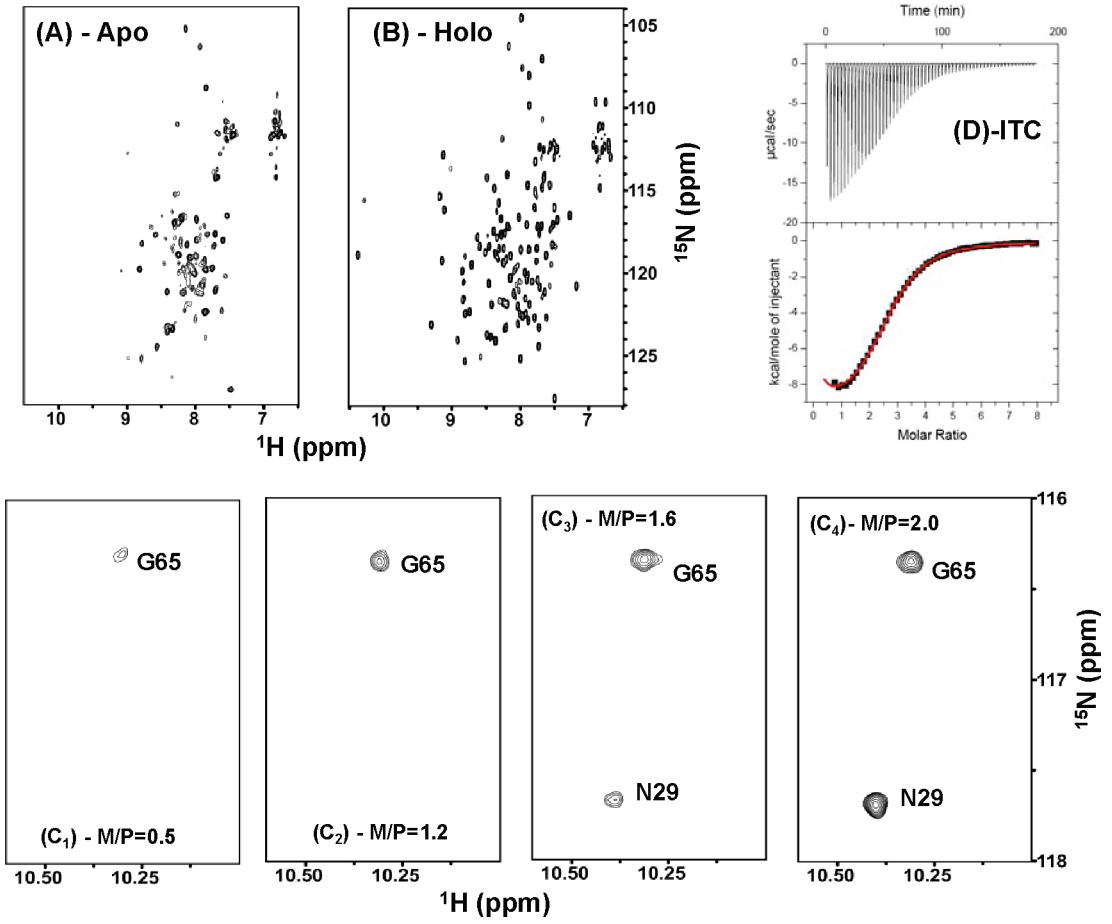


FIGURE 4: 2D [^{15}N−^{1}H] HSQC spectra of (A) apo-EhCaM and (B) holo-EhCaM. Spectra were recorded with ~1 mM protein, labeled with ^{15}N in a mixed solvent of 90% H_2O and 10% ^2H_2O. Increasing concentration of $CaCl_2$ was added to the protein solution and allowed to equilibrate before recording the spectra on a Bruker 800 MHz NMR spectrometer at 303 K. (C_1−C_4) Selected region of 2D [^{15}N−^{1}H] HSQC spectra recorded during the Ca^{2+} titration of EhCaM showing N29 and G65 peaks at different M/P ratio. The M/P ratio is indicated in each panel. (D) Thermogram of Ca^{2+} binding to EhCaM. The protein concentration was 145 μM in 50 mM Tris-HCl (pH = 7.0) containing 100 mM NaCl. The data fitted best to a two-site sequential binding model. The best-fit values of macroscopic dissociation constant and associated ΔH and ΔS values are listed in Table 1.

Table 1: Thermodynamic Parameters for EhCaM from ITC

K (M^{-1})	ΔH_1 (kcal·mol^{-1})	ΔS (kcal·mol^{-1}·K^{-1})
$K_1 = 4.00 \times 10^6 \pm 5.3 \times 10^3$	$\Delta H_1 = -1.605 \times 10^4 \pm 86.6$	$\Delta S_1 = -23.6$
$K_2 = 7.28 \times 10^4 \pm 2.1 \times 10^2$	$\Delta H_2 = -7573 \pm 104$	$\Delta S_2 = -3.16$

following: the transition temperature (T_m) = 34.28 ± 0.051 °C, the calorimetric enthalpy (ΔH_c) = 5393 ± 54.5 cal/mol, and the van't Hoff enthalpy (ΔH_v) = $6.619 \times 10^6 \pm 827$ cal/mol for the apo-EhCaM and T_m = 66.53 ± 0.038 °C, $\Delta H_c = 2.092 \times 10^4 \pm 132$ cal/mol, and $\Delta H_v = 6.785 \times 10^4 \pm 530$ cal/mol for the Ca^{2+}-bound EhCaM.

Energetics of Ca^{2+} Binding. Ca^{2+} binding to EhCaM was studied using ITC (Figure 4D). The ITC data show clearly two sequential Ca^{2+}-binding sites, an unusual mode of Ca^{2+} binding among the known CaBPs. The data could be fitted to only the two-site sequential binding model with the lowest χ^2. Either of these sites is found to be exothermic in nature. One binding site has affinity in the micromolar range ($K = 4.00 \times 10^6 \pm 5.3 \times 10^3$), while the other has affinity in the submicromolar range ($K = 7.28 \times 10^4 \pm 2.1 \times 10^2$). The thermodynamic parameters thus obtained are shown in Table 1. The affinity of Ca^{2+} for this protein is lower than that of other Ca^{2+}-sensor proteins reported in the literature (*41*, *42*).

Identification of EF-Hand Motifs by NMR. It has been well established that many CaBPs consists of two domains separated by a linker, having two Ca^{2+}-binding motifs in each domain, and the metal binding is cooperative in nature. NMR experiments were carried out with EhCaM to identify the two Ca^{2+}-binding sites that were established based on the ITC data. The two nonoverlapping peaks arising from the residues at the sixth position of the EF-hand motifs were used as markers to identify the Ca^{2+}-binding motifs (*43*).

Panels A and B of Figure 4 show the [^{15}N- ^{1}H] HSQC spectra of apo- and holo-EhCaM, respectively. The spectrum of apo-EhCaM shows a narrow dispersion in backbone $^{1}H^{N}$ chemical shifts (7−9.1 ppm) and accounts for only one-third of the expected peaks. Most of the peaks are broad due to constrained conformational fluctuations on the micro- and millisecond time scale (*44*−*46*). The peaks arising from the residues at the sixth position of the individual Ca^{2+}-binding motifs did not show any downfield-shifted spectral signature, indicating the absence of any bound metal ion (*47*).

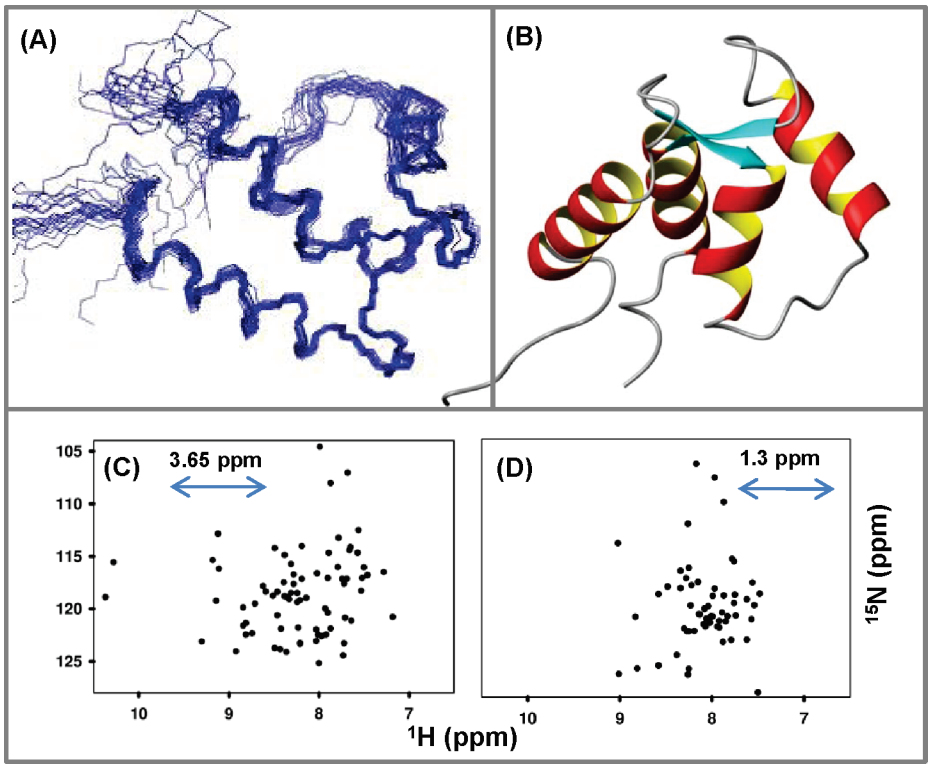


FIGURE 5: (A) An ensemble of 20 superimposed minimum energy NMR-derived structures of the N-terminal domain of holo-EhCaM (rmsd = 0.61). (B) Representative lowest energy structure of holo-EhCaM. (C) A mimic plot of ^{15}N and ^{1}H assignment for the N-terminal domain and (D) that of ^{15}N and ^{1}H assignment for the C-terminal domain. The poor dispersion along the $^{1}H^{N}$ chemical shift (1.3 ppm) seen in panel D suggests that the C-terminal domain may have some collapsed structure like other unstructured proteins while the dispersion of 3.65 ppm in the case of N-terminal domain suggests a well-folded structure for this domain, which is also verified from the 3D solution structure of EhCaM derived from NMR.

The [^{15}N−^{1}H] HSQC spectrum (Figure 4B) of Ca^{2+}-bound EhCaM displays relatively sharper peaks compared to those in the apo state, revealing that the Ca^{2+}-bound form is more structured. Also, the poor chemical shift dispersion in the proton dimension, which was still present in the middle of the HSQC, suggests the presence of a largely unstructured polypeptide stretch within the protein. Backbone resonance assignments established that the unstructured part corresponds to its C-terminal domain. The appearance of two downfield-shifted peaks in the HSQC arising from the residues at position 6 of EF-hands I and II (N29 and G65) indicates the presence of two Ca^{2+}-binding sites. This is the result of the involvement of the $^{1}H^{N}$ protons of N29 and G65 in hydrogen bonding with the side-chain carboxyl oxygen atom (C′O) of an invariant Asp at position 1 of the respective Ca^{2+}-binding loops (*43*). Since NMR can provide residue level information, we monitored the Ca^{2+}-filling pathway and noticed that the Ca^{2+} binds first to EF-II and then to EF-I (Figure 4C_1−C_4). During the course of Ca^{2+} titration, the peak corresponding to G65 of EF-II appears first, followed by the peak corresponding to N29 of EF-I. This reveals that the Ca^{2+} binding of EF-I is enhanced by the filling of EF-II due to cooperativity. It is found that binding affinity of EhCaM is in the millimolar range in contrast to most of the CaBPs that are in the micromolar range. This low binding affinity of EhCaM for Ca^{2+} may be due to the changes of the residues at positions 3, 5, and 7 of the Ca^{2+}-binding loop that are involved in Ca^{2+} coordination. Though the amino acid sequence analysis shows three putative Ca^{2+}-binding sites in EhCaM, NMR studies revealed only two Ca^{2+}-binding sites, both in the N-terminal domain of EhCaM, which is in conformity with the observation made using ITC. Sequence homology shows that EF-III (Scheme-1) has conserved residues for binding to Ca^{2+}, as in most of the CaBPs. However, most of the residues in EF-IV (Scheme 1) are changed in comparison to the canonical EF-hand motifs of CaBPs. Hence, it is unable to bind Ca^{2+} in the C-terminal domain due to lack of cooperativity, although EF-III has conserved residues. Thus EF-III and EF-IV of the C-terminal domain of EhCaM do not show any affinity for Ca^{2+}, and both of the metal ions bind sequentially to the N-terminal domain of EhCaM.

EhCaM, Partially Unfolded Protein. The 3D structure of EhCaM was deduced using the NMR data as described in Materials and Methods. Due to insufficient distance constraints seen in the C-terminal domain, we could not concurrently obtain the structural information. The C-terminal domain of EhCaM is thus found to be almost a random-coil-like structure with short polypeptide stretches, widely dispersed, showing α-helical signatures. Hence, we calculated the 3D structure of the N-terminal domain (M1−S85) alone. Figure 5A shows an ensemble of 20 superimposed structures of the N-terminal domain of EhCaM obtained with the lowest rmsd ~0.61 (residues 10−79), and Figure 5B shows its representative lowest energy structure. Structural statistics of the ensemble of 20 refined conformers of the N-terminal domain are shown in Table 2. The mimic plots of HSQC for N- and C-terminal domains show that the dispersion in the $^{1}H^{N}$ chemical shifts is 3.65 and 1.3 ppm for the N- and C-terminal domains, respectively (Figure 5C,D). This further supports that the EhCaM has a well-defined N-terminal domain and an unstructured C-terminal counterpart.

Structural Comparison between EhCaM and CaM. In order to find out structural similarity between EhCaM and CaM, if any, we have carried out detailed analysis of the N-lobes of

Table 2: Structural Statistics for the Ensemble of 20 Refined Conformers of the N-Terminal Domain of Holo-EhCaM (PDB ID 2KTG)

NMR derived restraints			
total interproton restraints			942
intraresidue ($[i-j]=0$)			281
sequential ($[i-j]=1$)			362
medium range ($1<[i-j]<5$)			224
long range ($[i-j]\geq 5$)			75
hydrogen bonds			60 (30 × 2)
dihedral angles (φ, ψ)			116
no. of structures used for analysis			20
average CYANA target function			0.64 ± 0.07
constraint violations			
NOE distance violations >0.1 Å			none
van der Waals violations >0.1 Å			2
dihedral angle violations >3 Å			none
rmsd values	all	ordered[b]	selected[a]
all backbone atoms (Å)	2.1	0.61	0.61
all heavy atoms (Å)	2.4	1.2	1.2
Ramachandran plot summary from PROCHECK[a]			
residues in most favored regions (%)			89.9
residues in additionally allowed regions (%)			7.2
residues in generously allowed regions (%)			2.9
residues in disallowed regions (%)			0.0
structure quality factors: overall statistics	mean score	SD	*Z*-score[c]
Procheck *G*-factor[b] (φ, ψ only)	0.03	N/A	0.43
Procheck *G*-factor[b] (all dihedral angles)	−0.62	N/A	−3.67
Verify3D	0.30	0.0289	−2.57
ProsaII (−ve)	0.71	0.0516	0.25
MolProbity clash score	15.19	3.8346	−1.08

[a]Residues selected based on dihedral angle order parameter with $S(\varphi)+S(\psi)\geq 1.8$. Selected residue ranges: 4A−7A, 9A−43A, and 47A−79A. [b]Residues with sum of φ and ψ order parameter >1.8. Ordered residue ranges: 4A−7A, 9A−43A, and 47A−79A. [c]With respect to mean and standard deviation for a set of 252 X-ray structures <500 residues, of resolution ≥1.8 Å, *R*-factor ≥0.25, and *R*-free ≥0.28; a positive value indicates a "better" score. Generated using PSVS 1.4.

EhCaM and CaM. For this, we separately overlaid individual Ca^{2+}-binding loops of EhCaM and CaM by using PyMol and found that the conformation of individual Ca^{2+}-binding loops of EF-I and EF-II belonging to EhCaM matches well with those of CaM with an rmsd of 1.09 and 1.13 Å, respectively (Supporting Information Figue S4). Besides, we have also overlaid the EF-I and EF-II of EhCaM and CaM and found that the overall conformation of the aforementioned EF-hands of the two proteins is also similar (Supporting Information Figure S4). Thus, we strongly believe that the structured N-terminal domain of the EhCaM may be involved in binding target molecules as in the case of CaM (*48*) and thus be part of a novel Ca^{2+}-signaling pathway. Therefore, it is unlikely that a functional homologue of CaM has been missed in *E. histolytica*.

EhCaM Dynamics. The R_1 rates are quite constant across the polypeptide chain with an average value of 1.12 ± 0.06 s^{-1} (Figure 6A). The R_2 values (Figure 6B) show larger variation along the sequence, ranging from 11.7 to 21.4 s^{-1}. Higher values of R_2 (>12 s^{-1}) for most of the residues in the N-terminal domain support the earlier observation of its well-folded structure, while higher R_2 values seen for a few residues belonging to the C-terminal domain are due to the exchange contribution, R_{ex} (Figure 6D) calculated from the relaxation data at two fields (600 MHz, data not shown, and 800 MHz). The structured N-terminal domain is further established from the observation of higher heteronuclear ($^{15}N-^{1}H$) NOE values (Figure 6C) for this domain compared to its floppy C-terminal counterpart. The average NOE value for EhCaM is 0.77 ± 0.04 s^{-1}. Negative $^{15}N-^{1}H$ NOE values at N- and C-terminals show regions of high flexibility. The blank regions in the relaxation data, seen mostly in the C-terminal domain, were primarily due to severe spectral overlap.

During model-free analysis, simultaneous fitting of the R_2/R_1 ratios gave a $D_{\parallel}/D_{\perp}$ of 0.96 for the N-terminal domain of EhCaM. The axially symmetric diffusion tensors derived from experimental R_2/R_1 ratios were used for the model-free analysis of the relaxation data. Based on these relaxation data, the final overall tumbling time (τ_c) was found to be 9.6 ns, which supports the monomeric state of EhCaM. Further insights into the dynamics of the ordered regions of EhCaM could be obtained by analyzing the correlation between the transverse relaxation parameter (R_2) and the average area buried upon folding (Figure 6F_1-F_3) (as described in the Discussion section).

During the hydrogen exchange experiment, most of the solvent-accessible residues exchanged very fast with deuterium, upon addition of $^{2}H_2O$ to the protein, resulting in the disappearance of many of their respective peaks in the $[^{15}N-^{1}H]$ HSQC in the dead time before recording the first spectrum. The sudden disappearance of most of the peaks belonging to the C-terminal domain reveals it is a partially folded nature. The rate of disappearance of the remaining peaks was studied as a function of time by recording a series of HSQC, and associated PFs were calculated following the procedure described in the Materials and Methods section. As evident from Figure 6E, higher values of PFs observed for the residues in the N-terminal domain compared to the C-terminal counterpart suggest a greater structural stability for the N-terminal domain.

All of the spectroscopic evidence taken together, namely, slightly higher molecular size of the protein as observed in the gel-filtration chromatographic profile and absence of Ca^{2+}-binding sites, poor dispersion in the $^{1}H^{N}$ chemical shifts, nonobservance of NOEs, sudden disappearance of $[^{15}N-^{1}H]$ peaks in the HX experiment, and negative heteronuclear $^{15}N-^{1}H$

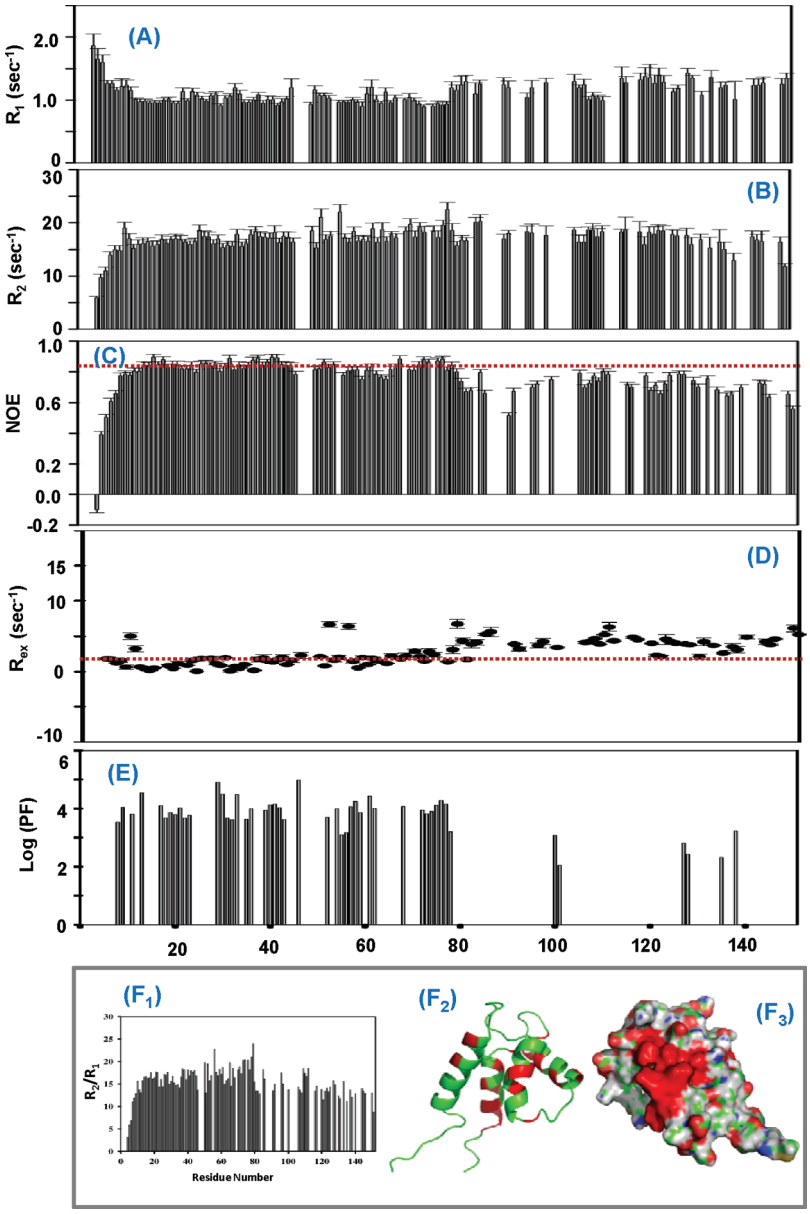


FIGURE 6: ^{15}N relaxation data recorded on a Bruker 800 MHz spectrometer. (A) Longitudinal relaxation rates, R_1 ($1/T_1$). (B) Transverse relaxation rates, R_2 ($1/T_2$). (C) [^{15}N−^{1}H] NOE enhancements are defined as I_{sat}/I_{eq}, where I_{sat} and I_{eq} are the intensities of peaks in the 2D experiments with and without proton saturation, respectively. Error bars of the T_1 and T_2 data denote curve-fitting uncertainties, and errors in the [^{15}N−^{1}H] NOEs are estimated from the signal/noise ratio of the spectra. The blank regions found in the plots correspond to residues showing extensive spectral overlap, unassigned residues, broadening of a few spectral signatures, and one Pro residue. (D) Chemical exchange contributions, R_{ex} (estimated) to the transverse relaxation rates, and R_2, calculated based on the relaxation data at two fields, 600 and 800 MHz. (E) Histograms showing the residue-wise distribution of protection factors (PFs) from amide hydrogen exchange for holo-EhCaM. The sudden disappearance of most of the peaks corresponding to the C-terminal domain supports the unstructured part of it. (F_1) Residue-wise plot of R_2/R_1. (F_2) Secondary structure of EhCaM showing residues having higher R_2/R_1 values (in red). (F_3) Exposed hydrophobic surface (shown in red) of the N-terminal domain of EhCaM. These data correlate transverse relaxation parameter (R_2) and average area buried upon folding (F_1−F_3).

NOEs in the C-terminal domain, support that EhCaM has a folded N-terminal domain and unstructured C-terminal domain.

DISCUSSION

Ca^{2+} plays an important role in amebic pathogenesis, and it is likely that one or many of the CaBPs, encoded in amebic genome, may be involved in some of the signaling pathways that determine host−pathogen interaction. In order to identify the key CaBPs involved in pathogenesis, we set out to characterize some of the CaBPs present in *E. histolytica*. Two of the EhCaBPs (EhCaBP-1 and -2) have been structurally characterized earlier in our laboratory (*13*, *16*). In this paper, we characterize yet another CaBP from *E. histolytica* (named EhCaM) using various biophysical and biochemical methods. The absence of any homologue of EhCaM in the database, when the sequence alignments are viewed outside the EF-hand loops, suggests that it is a novel protein. Therefore, it may have a unique role in *E. histolytica* biology. A number of observations suggest that EhCaM is distinct from EhCaBP-1 and -2 as well as CaM: (1) While EhCaBP-1 and -2 display a sequence identity of 79% at both nucleotide and amino acid level among themselves, EhCaM is quite divergent, showing a low level of identity with either EhCaBP-1 or -2. (2) Both EhCaBP-1 and -2 are cytoplasmic proteins while EhCaM is present in all of the three subcellular

fractions (membrane, cytoplasm, and nucleus). (3) EhCaM cannot stimulate the activity of cAMP phosphodiesterase, a signature activity of all CaMs (*14*).

Due to the fact that EhCaM is also expressed inside the nucleus, it is likely that the unusual structure of this molecule may reflect this localization (Figure 2). Interestingly, no nuclear localization signal was detected in the primary sequence of EhCaM. Recently, a few other CaBPs were found in the nucleus at very low concentrations (*49*). Ca^{2+} signaling in the nucleus is recognized to be important in many organisms (*50*) perhaps including *Entamoeba*.

In most cases, the ability of a CaBP to regulate or be regulated by the Ca^{2+} concentration is central to its function. Some CaBPs like CaM and TnC can transmit the signal of elevated free Ca^{2+} to regulate other proteins, whereas the role of others is to buffer the free Ca^{2+} within a range and may be significant in real time. In either of these classes, viz., Ca^{2+} regulatory or Ca^{2+} buffers, the positive cooperativity increases their sensitivity and response to Ca^{2+} concentrations (*51*, *52*). It was observed that many CaBPs, belonging to EF-hand superfamily, bind Ca^{2+} with affinity constants varying from 10^4 to 10^9 M^{-1} (*13*, *14*, *53*). In contrast, EhCaM bind Ca^{2+} with affinity constants ranging from 10^4 to 10^6 M^{-1}. It is worth mentioning here that, in general, amebic proteins display lower affinity for Ca^{2+} as compared to CaBPs from other organisms (*14*).

Despite having 79% sequence similarity with one another, EhCaBP1 and EhCaBP2 are reported to bind different sets of proteins and peptides, suggesting that they may be involved in distinct Ca^{2+}-signaling pathways (*7*, *14*, *54*). As mentioned earlier, thermodynamic data reveal that EhCaM has two Ca^{2+}-binding sites. The NMR titration data revealed the sequential occupancy of two sites, with the first Ca^{2+} ion binding to site II, followed by the binding to site I. These sites have been identified using the amino acid residues at the sixth position of the individual Ca^{2+}-binding loops, namely, N29 (belonging to EF-hand I) and G65 (belonging to EF-hand II) as markers. These markers appear highly downfield shifted and do not overlap with the rest of the spectral signatures arising from other residues. Interestingly, both the above-mentioned Ca^{2+}-binding sites are located in the N-terminal domain (M1−Y79). Though the presence of Asn at the sixth position in the Ca^{2+}-binding loop is unusual, the downfield shift of N29 is attributed to the hydrogen bonding it is involved with, which is confirmed by HX data (Figure 6E).

Coming to the structure of EhCaM, the CD spectral data showed significant collapse of secondary and tertiary structure in its apo form. Upon addition of Ca^{2+}, the molten globular structure gets transformed into its native 3D fold, with the protein predominantly in an α-helical conformation. The observed increase in intensity of the aromatic bands in the far- and near-UV CD spectra is indicative of large changes in the protein conformation upon Ca^{2+} binding. This is also reflected in the NMR data. The secondary structural elements of the protein published in our earlier paper (*18*) were identified from the well-established empirical relationship of ^{13}C chemical shifts ($\Delta C^{\alpha}-\Delta C^{\beta}$) (*55*, *56*), suggesting that EhCaM is predominantly α-helical in conformation with two short stretches of β-strands in the holo form. The narrow dispersion in backbone $^1H^N$ chemical shifts (7−9.1 ppm) and observation of only one-third of the expected peaks reveal that the protein in its apo form has a collapsed tertiary structure or molten globule-like structure. On the other hand, holo-EhCaM displays relatively sharper peaks compared to those in the apo state, revealing that the Ca^{2+}-bound form is more structured.

The NMR-derived 3D structure of EhCaM converges well within its N-terminal domain with a lowest rmsd (0.61 Å). On the other hand, the C-terminal domain of the protein could not converge due to insufficient NOEs. Thus, the structure of EhCaM is partially folded. It has a well-folded N-terminal domain and an unstructured C-terminal domain. Although the CSI plot (*18*) shows that the C-terminal domain has some secondary structure propensity, it does not show any tertiary fold. The observation that the C-terminal domain is unstructured is supported by low chemical shift dispersion of amide protons, insufficient NOEs, sudden disappearance of [$^{15}N-^{1}H$] peaks in the HX experiment, negative heteronuclear $^{15}N-^{1}H$ NOEs, and absence of Ca^{2+}-binding sites. EhCaM thus having an unstructured C-terminal domain may be involved in a novel signal transduction process.

The distribution of R_2/R_1 ratio (Figure 6F_1) in a sequence-dependent manner can be used to correlate the presence of conformational exchange and fast internal motions. The presence of internal motions in the time scale of τ_c results in the reduction of R_2/R_1 ratio. On the other hand, any conformational exchange in the time scale of millisecond to nanosecond range or/and fast exchange of the $^1H^N$ with solvent will increase the value of R_2/R_1. The information content of R_2/R_1 ratio is of great importance to determine the presence of conformational exchange and/or fast internal motion, which is not evident from the trend of sequence-wise distribution of R_2 values alone. Many residues (E18, Q21, M40, G44, and L77) are involved either in conformational exchange or in fast exchange with the solvent. This is primarily due to the enhanced R_2/R_1 ratio although they show lower R_2 values. The residues that show higher R_2 values as well as higher R_2/R_1 ratio are probably due to conformational exchange or due to higher solvent accessibility. G37, T38, Q50, I52, V56, Y59, G65, F67, T71, L73, T74, M76, and Y79 which show both higher R_2 and R_2/R_1 are exposed (Figure 6F_2,F_3) to solvent when mapped on the surface of the protein. Further, five residues (L73−T74 and M76−Y79) belonging to the F-helix of EF-II also show higher R_2 and R_2/R_1 ratio, as this helix is highly exposed to the solvent (Figure 6F_2,F_3). Apart from these, the residues that reside on the termini of the secondary structural elements (Q21, G37, Q50, Y59, E78, and Y79) also show higher R_2/R_1 ratio and hence may be involved in conformational exchange.

CaMs are highly conserved ubiquitous proteins, and EhCaM shows the highest sequence similarity among all CaMs. Therefore, it is unlikely that a functional homologue of CaM has been missed in *E. histolytica*. Overall, it appears that EhCaM may have diverged considerably from CaM in order to serve a specific function in *E. histolytica*. It is worth mentioning here that, in both CaM and TnC, the N-terminal domain is regulatory and interacts with the target molecules while the C-terminal domain acts as an anchoring domain (*48*, *57*). In the case of EhCaBP1, the N- and C-terminal domains have been reported to behave differently in their functional assays (*9*, *58*). It has also been reported that the N-terminal domain of EhCaBP1 behaves more like the whole molecule (*9*), and thus this domain was considered as a regulatory one. In the present study, with the 3D structure of EhCaM showing a structured N-terminal domain and an unstructured C-terminal domain, we strongly believe that the N-terminal domain is also regulatory in nature. Though such functional characterization is yet to be carried out for EhCaM, the domain structures mentioned earlier suggest its possible involvement in

binding target molecules and thus be part of the novel Ca^{2+}-signaling pathway.

CONCLUSION

In conclusion, we have described the NMR-derived high-resolution 3D solution structure of a CaM-like protein EhCaM, from the protozoan parasite *E. histolytica*, which consists of a structured N-terminal domain and an unstructured C-terminal domain. This characteristic is the first of its kind among the known CaBPs. This is supported by the observation of slightly slower mobility in the gel-filtration column compared to that expected of its molecular mass. Although the CSI plot showed that the C-terminal domain has some secondary structural propensity, no tertiary fold could be seen. The unstructured C-terminal domain is supported by the absence of Ca^{2+}-binding sites, poor dispersion in the $^1H^N$ chemical shifts, nonobservance of NOEs, sudden disappearance of [$^{15}N-^1H$] peaks in the HX experiment, and negative heteronuclear $^{15}N-^1H$ NOEs in the C-terminal domain. Thus, EhCaM is the first partially structured CaBP found in the EF-hand CaBPs, suggesting that it may be involved in a unique signal transduction pathway of this parasite. Since this is the first EhCaBP that has been found in the nucleus, it is likely that the structure may reflect its localization. Further, the fact that EhCaM shows highest sequence similarity among all CaMs, it is likely to be a functional homologue of CaM in *E. histolytica*. Overall, it appears that EhCaM has diverged considerably from CaM in order to serve a specific function in *E. histolytica*. Thus, taken together, the above studies on EhCaM provide some important and preliminary results for further study to answer some more questions pertaining to this parasite, its nature of containing such a large number of calcium-binding proteins, mechanisms of Ca^{2+}-signaling in this parasite, the role of the unstructured part of EhCaM on its function, and the information about its evolution.

ACKNOWLEDGMENT

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

SUPPORTING INFORMATION AVAILABLE

Some figures and information about RT-PCR, purification, Ca^{2+}-binding assays, and DSC. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Walsh, J. A. (1986) Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. *Rev. Infect. Dis. 8*, 228–238.
2. Bracha, R., Kobiler, D., and Mirelman, D. (1982) Attachment and ingestion of bacteria by trophozoites of *Entamoeba histolytica*. *Infect. Immun. 36*, 396–406.
3. Ravdin, J. I., Moreau, F., Sullivan, J. A., Petri, W. A., Jr., and Mandell, G. L. (1988) Relationship of free intracellular calcium to the cytolytic activity of *Entamoeba histolytica*. *Infect. Immun. 56*, 1505–1512.
4. Tsutsumi, V., Ramirez-Rosales, A., Lanz-Mendoza, H., Shibayama, M., Chavez, B., Rangel-Lopez, E., and Martinez-Palomo, A. (1992) *Entamoeba histolytica*: erythrophagocytosis, collagenolysis, and liver abscess production as virulence markers. *Trans. R. Soc. Trop. Med. Hyg. 86*, 170–172.
5. Meza, I. (2000) Extracellular matrix-induced signaling in *Entamoeba histolytica*: its role in invasiveness. *Parasitol. Today 16*, 23–28.
6. Ikura, M. (1996) Calcium binding and conformational response in EF-hand proteins. *Trends Biochem. Sci. 21*, 14–17.
7. Jain, R., Santi-Rocca, J., Padhan, N., Bhattacharya, S., Guillen, N., and Bhattacharya, A. (2008) Calcium-binding protein 1 of *Entamoeba histolytica* transiently associates with phagocytic cups in a calcium-independent manner. *Cell Microbiol. 10*, 1373–1389.
8. Berridge, M. J., Lipp, P., and Bootman, M. D. (2000) The versatility and universality of calcium signalling. *Nat. Rev. Mol. Cell Biol. 1*, 11–21.
9. Jain, R., Kumar, S., Gourinath, S., Bhattacharya, S., and Bhattacharya, A. (2009) N- and C-terminal domains of the calcium binding protein EhCaBP1 of the parasite *Entamoeba histolytica* display distinct functions. *PLoS ONE 4*, e5269.
10. Biekofsky, R. R., and Feeney, J. (1998) Cooperative cyclic interactions involved in metal binding to pairs of sites in EF-hand proteins. *FEBS Lett. 439*, 101–106.
11. Strynadka, N. C., and James, M. N. (1989) Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem. 58*, 951–998.
12. Nelson, M. R., and Chazin, W. J. (1998) Structures of EF-hand Ca(2+)-binding proteins: diversity in the organization, packing and response to Ca^{2+} binding. *Biometals 11*, 297–318.
13. Atreya, H. S., Sahu, S. C., Bhattacharya, A., Chary, K. V., and Govil, G. (2001) NMR derived solution structure of an EF-hand calcium-binding protein from *Entamoeba histolytica*. *Biochemistry 40*, 14392–14403.
14. Bhattacharya, A., Padhan, N., Jain, R., and Bhattacharya, S. (2006) Calcium-binding proteins of *Entamoeba histolytica*. *Arch. Med. Res. 37*, 221–225.
15. Nagamune, K., and Sibley, L. D. (2006) Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the apicomplexa. *Mol. Biol. Evol. 23*, 1613–1627.
16. Mustafi, S. M., Mutalik, R. B., Jain, R., Chandra, K., Bhattacharya, A., and Chary, K. V. (2009) Structural characterization of a novel Ca^{2+}-binding protein from *Entamoeba histolytica*: structural basis for the observed functional differences with its isoform. *J. Biol. Inorg. Chem. 14*, 471–483.
17. Munoz, M. L., Moreno, M. A., Perez-Garcia, J. N., Tovar, G. R., and Hernandez, V. I. (1991) Possible role of calmodulin in the secretion of *Entamoeba histolytica* electron-dense granules containing collagenase. *Mol. Microbiol. 5*, 1707–1714.
18. Rout, A. K., Barnwal, R. P., Padhan, N., Bhattacharya, A., and Chary, K. V. (2008) Sequence specific ^1H, ^{13}C and ^{15}N resonance assignments of a calmodulin-like calcium-binding protein from the protozoan parasite *Entamoeba histolytica* (EhCaM). *Biomol. NMR Assign. 2*, 77–79.
19. Dey, I., Keller, K., Belley, A., and Chadee, K. (2003) Identification and characterization of a cyclooxygenase-like enzyme from *Entamoeba histolytica*. *Proc. Natl. Acad. Sci. U.S.A. 100*, 13561–13566.
20. Barnwal, R. P., Rout, A. K., Chary, K. V., and Atreya, H. S. (2007) Rapid measurement of ^3J(H N-H alpha) and ^3J(N-H beta) coupling constants in polypeptides. *J. Biomol. NMR 39*, 259–263.
21. Atreya, H. S., Sahu, S. C., Chary, K. V., and Govil, G. (2000) A tracked approach for automated NMR assignments in proteins (TATAPRO). *J. Biomol. NMR 17*, 125–136.
22. Keller, R. (2004) The computer aided NMR resonance assignment tutorial, *CANTINA Verlag Goldau*.
23. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ^1H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR 6*, 135–140.
24. Gunter, P., Qian, Y. Q., Otting, G., Muller, M., Gehring, W., and Wuthrich, K. (1991) Structure determination of the Antp (C39---S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for structure the calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. *J. Mol. Biol. 217*, 531–540.
25. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for the NMR structure calculation with the new program DYANA. *J. Mol. Biol. 273*, 283–298.
26. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR 8*, 477–486.
27. DeLano, W. L. (2002) The PyMOL molecular graphics system, San Carlos, CA.

28. Cang, H. X., Chen, M. E., and Nymeyer, H. (2006) NOC: a free molecular explorer for protein structure visualization, validation and analysis, Florida:Nymeyer's group, Institute of Molecular Biology, Florida State University.
29. Vold, R. L., J., W., Klein, M. P., and Phelps, D. E. (1968) Measurement of spin relaxation in complex systems. *J. Chem. Phys. 48*, 483831–483832.
30. Cavanagh, J., Palmer, A. G., Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear relay spectroscopy. *J. Magn. Reson. 91*, 429–436.
31. Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J. Magn. Reson. 93*, 151–170.
32. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry 28*, 8972–8979.
33. Skelton, N. J., Palmer, A. G., Akke, M., Kordel, J., Rance, M., and Chazin, W. J. (1993) Practical aspects of two-dimensional proton-detected ^{15}N spin relaxation measurments. *J. Magn. Reson., Ser. B 102*, 253–264.
34. Palmer, A. G., Rance, M., and Wright, P. E. (1991) Intramolecular motions of a zinc finger DNA-binding domain from xfin characterized by proton-detected natural abundance ^{13}C heteronuclear NMR spectroscopy. *J. Am. Chem. Soc. 113*, 4371–4380.
35. Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., and Torchia, D. A. (1992) Dynamics of methyl groups in proteins as studided by proton detected ^{13}C NMR spectroscopy. Application to the leucine residues of staphylococal nuclease. *Biochemistry 31*, 5253–5263.
36. Wong, K. B., Fersht, A. R., and Freund, S. M. (1997) NMR ^{15}N relaxation and structural studies reveal slow conformational exchange in barstar C40/82A. *J. Mol. Biol. 268*, 494–511.
37. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins 17*, 75–86.
38. Mukherjee, S., Kuchroo, K., and Chary, K. V. (2005) Structural characterization of the apo form of a calcium binding protein from *Entamoeba histolytica* by hydrogen exchange and its folding to the holo state. *Biochemistry 44*, 11636–11645.
39. Christova, P., Cox, J. A., and Craescu, C. T. (2000) Ion-induced conformational and stability changes in Nereis sarcoplasmic calcium binding protein: evidence that the APO state is a molten globule. *Proteins 40*, 177–184.
40. Daniel, L., Philippe, O. T., Francois, D., Rene, T., Frederic, B., Claudette, B., Alexander, A. M., and Jacques, H. (2002) Cation binding mode of fully oxidised calmodulin explained by the unfolding of the apostate, *Biochim. Biophys. Acta*.
41. Mukherjee, S., Mohan, P. M., and Chary, K. V. (2007) Magnesium promotes structural integrity and conformational switching action of a calcium sensor protein. *Biochemistry 46*, 3835–3845.
42. Linse, S., Helmersson, A., and Forsen, S. (1991) Calcium binding to calmodulin and its globular domains. *J. Biol. Chem. 266*, 8050–8054.
43. Biekofsky, R. R., Martin, S. R., Browne, J. P., Bayley, P. M., and Feeney, J. (1998) Ca^{2+} coordination to backbone carbonyl oxygen atoms in calmodulin and other EF-hand proteins: ^{15}N chemical shifts as probes for monitoring individual-site Ca^{2+} coordination. *Biochemistry 37*, 7617–7629.
44. Baum, J., Dobson, C. M., Evans, P. A., and Hanley, C. (1989) Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry 28*, 7–13.
45. Schulman, B. A., Kim, P. S., Dobson, C. M., and Redfield, C. (1997) A residue-specific NMR view of the non-cooperative unfolding of a molten globule. *Nat. Struct. Biol. 4*, 630–634.
46. Rao, B. D. (1989) Nuclear magnetic resonance line-shape analysis and determination of exchange rates. *Methods Enzymol. 176*, 279–311.
47. Mukherjee, S., Mohan, P. M., and Chary, K. V. (2007) Magnesium promotes structural integrity and conformational switching action of a calcium sensor protein. *Biochemistry 46*, 3835–3845.
48. Crivici, A., and Ikura, M. (1995) Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct. 24*, 85–116.
49. Bachs, O., Agell, N., and Carafoli, E. (1994) Calmodulin and calmodulin-binding proteins in the nucleus. *Cell Calcium 16*, 289–296.
50. Malviya, A. N., and Klein, C. (2006) Mechanism regulating nuclear calcium signaling. *Can. J. Physiol. Pharmacol. 84*, 403–422.
51. Williams, R. J. (1992) Calcium and calmodulin. *Cell Calcium 13*, 355–362.
52. Williams, R. J. (1992) Calcium fluxes in cells: new views on their significance. *Cell Calcium 13*, 273–275.
53. Mustafi, S. M., Mutalik, R. B., Jain, R., Chandra, K., Bhattacharya, A., and Chary, K. V. (2009) Structural characterization of a novel Ca^{2+}-binding protein from *Entamoeba histolytica*: structural basis for the observed functional differences with its isoform. *J. Biol. Inorg. Chem. 14*, 471–483.
54. Mustafi, S. M., Mutalik, R. B., Jain, R., Chandra, K., Bhattacharya, A., and Chary, K. V. (2009) Structural characterization of a novel Ca^{2+}-binding protein from *Entamoeba histolytica*: structural basis for the observed functional differences with its isoform. *J. Biol. Inorg. Chem. 14*, 471–483.
55. Spera, S., and Bax, A. (1991) Empirical correlation between protein backbone conformation and C^a and C^b ^{13}C nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc. 113*, 5490–5492.
56. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR 5*, 67–81.
57. Filatov, V. L., Katrukha, A. G., Bulargina, T. V., and Gusev, N. B. (1999) Troponin: structure, properties, and mechanism of functioning. *Biochemistry* (*Moscow*) *64*, 969–985.
58. Mukherjee, S., Mohan, P. M., Kuchroo, K., and Chary, K. V. (2007) Energetics of the native energy landscape of a two-domain calcium sensor protein: distinct folding features of the two domains. *Biochemistry 46*, 9911–9919.