Protein Folding and Conformational Stability Studies on Cyclophilin and its Mutants

By

SOURAV ROY

Enrolment no: LIFE05200804004

Saha Institute of Nuclear Physics, Kolkata

A thesis submitted to the

Board of Studies in Life Sciences

In partial fulfillment of requirements

For the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

JANUARY 2015

HOMI BHABHA NATIONAL INSTITUTE[1]

RECOMMENDATIONS OF THE VIVA VOCE COMMITTEE

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Sourav Roy entitled " Protein folding and conformational stability studies on cyclophilin and its mutants" and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman – Prof. Dipak Dasgupta D Dasgupta	Date: 20/4/15
Guide / Convener – Prof. Rahul Banerjee R Banerjee	Date: 20/4/15
Co-guide -	Date:
Examiner – Prof. Debasisa Mohanty Debasisa Mohanty	Date: 20/4/15
Member 1- Prof. Dhananjay Bhattacharyya Dhananjay Bhattacharyya	Date: 20/4/2015
Member 2- Prof. Udayaditya Sen Udayaditya Sen	Date: 20/4/15

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 20.4.15

Place: KOLKATA

Co-guide (if applicable)

R Banerjee

Guide

Dr. Rahul Banerjee
Professor
Saha Institute of Nuclear Physics
1/AF, Bidhannagar, Kolkata - 700 064

[1] This page is to be included only for final submission after successful completion of viva voce.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Sourav Roy

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.

(Sourav Roy)

List of Publication(s) from Thesis

Journal

1. Equilibrium unfolding of cyclophilin from *Leishmania donovani*: Characterization of intermediate state(s). **Sourav Roy**, Sankar Basu ,Alok K. Datta , Dhananjay Bhattacharyya, Rahul Banerjee Dipak Dasgupta. *Intenational Journal of Biological macromolecules.* (2014) 69,353-360, doi: 10.1016/j.ijbiomac.2014.05.063

2. Application of Network Based Metrics to Probe Patterns in The Unlocking of Close Packing During The Unfolding of Cyclophilin. **Sourav Roy**, Sankar Basu, Dipak Dasgupta, Dhananjay Bhattacharyya, Rahul Banerjee (Under Review).

Conferences Attended

1) 'National workshop on Biocalorimetry', organized by Saha Institute of Nuclear Physics, 2009

2) Poster presentation in 'International Conference on Physics Biology Interface' organized by Saha Institute of Nuclear Physics, Kolkata and Indian Biophysical Society, Kolkata, 2009.

3) Poster Presentation in 'Recent Advances in Chemical and Physical Biology' organized by Saha Institute of Nuclear Physics and the Mechanobiology Institute, Singapore.

4) Poster Presentation in 'International Conference on Protein Folding and Dynamics' organized by the National Centre for Biological Sciences, Bangalore, 2012

5) Poster Presentation in 'National conference and workshop on Fluorescence Correlation Spectroscopy' organized by Saha Institute of Nuclear Physics and Tata Institute of Fundamental Research, 2013

6) Poster Presentation in 'Annual conference of the Indian Biophysical Society MADALS' organized by Saha Institute of Nuclear Physics, 2014

(Sourav Roy)

Dedicated to my parents, my wife

and

all my well wishers

ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere and heartfelt gratitude to the help and cooperation extended by my friends and well-wishers during my PhD tenure. First and foremost, I would like to convey my sincere regards and gratitude to my guide Prof. Rahul Banerjee for his untiring cooperation, suggestions and guidance that has led to the successful completion of this thesis.

I am indebted to Prof. Dipak Das Gupta (Biophysics and Structural Genomics Division, SINP) and Prof. Dhananjay Bhattacharyya (Computational Science Division, SINP) for their valuable suggestions, constant support and guidance in spectroscopy and molecular dynamics simulations. I would also like to extend my acknowledgements to my seniors in the lab Venuda, Sankarda, and Ambarnilda who had made my stay in the lab a memorable one. I am also grateful to Shibojyotida, Parijatdi, Sukanyadi, Manas, Sanchitadi, Sukanyadi, Angana, Amrita, Kasturidi, Shreyasi and Saptaparni for their kind cooperation. My post-MSc batch mates Suchismita, Kasturi, Avik, Shounak, Srijit, Barnali, Seema, Ramanuj who have stood by me in times of need. Financial assistance from the Department of Atomic Energy (DAE), Government of India and other laboratory facilities from Saha Institute of Nuclear Physics are also acknowledged. My sincere thanks are due to all the staff members of the Crystallography and Molecular Biology Division for their generous cooperation. I would specially like to mention my friends and lab mates for encouraging me through the highs and lows of my stay in this lab from the very beginning. Their friendly attitude shall always be a treasure to me. Lastly, I want to heartily thank my parents (Ma and Baba) and my wife Saswati whose constant support led my path towards this degree.

Sourav Roy

Kolkata

Contents

Chapter 3

Chapter 4

List of Figures with Figure Legends

Chapter 1

Figure 10: Crystal structure of BPTI (pdb id: 5PTI.pdb) with two 2 ß-sheets, a α-helix and a 3_{10} helical turn. The figure is drawn using Pymol (LLC Schrodinger)

Chapter 2

Figure 1: Native crystal structure of cyclophilin from *Leishmania donovani* (LdCyp) in green (cartoon representation) with the partially solvent exposed lone tryptophan residue Trp 143 (in yellow, ball and stick mode). Helix H1 is colored red and Helix H2 colored cyan. The Figure is generated using Pymol (Schrodinger LLC).

Figure2. Fluorescence unfolding spectra of native cyclophilin at various GdmCl concentrations ranging from (0-3) M a) (—) 4µM LdCyp b) (—) 4µM LdCyp+0.8 M GdmCl c) (—) 4µM LdCyp + 1.0 M GdmCl d) (—) 4µM LdCyp +1.4M GdmCl e) (—) 4µM LdCyp+3.0M GdmCl.

Figure3. Unfolding curve of LdCyp induced by GdmCl as monitored by intrinsic tryptophan fluorescence with emission maxima points at 360nm, fitted to a 3-state equation. The black circles (●) indicate the fluorescence intensity in arbitrary units at the corresponding denaturant concentrations (in mM) and (—) is the fitted line to the datapoints.

Figure4. Far-UV CD spectra of LdCyp (native) along with different denaturant concentrations a) (—) 4µM b) (—)4µ M+ 1M GdmCl c) (—)4µM+ 1.2 M GdmCl d)(—) 4µ M+1.4 M GdmCl e) (—) 4µM+ 3.0 M GdmCl.

Figure5. MRE_{222} [in black circles (●)] values obtained from far-UV CD spectra plotted as a function of the denaturant and fitted to a 3-state equation shown as a black (—) line.

Figure6. Far-UV CD unfolding curves of LdCyp monitored at MRE_{222} (■) black squares, MRE_{225} (●) red circle and at MRE_{215} (▲).

Figure7. The fraction unfolded versus GdmCl curve with independent probes of intrinsic tryptophan fluorescence monitored at 360nm, shown in filled red circles(●) , far UV CD monitored at 222nm in black squares (■) and near UV CD monitored at a maxima of 260nm shown in blue triangles(▲).

Figure8. Fluorescence unfolding curve of LdCyp monitored at 360 nm and fitted to a) two state model in Panel1 and b) three state model in Panel 2.

Figure9. Far UV CD unfolding curve of LdCyp monitored at 222nm being fitted to a) 2-state process in panel 1 b) and 3-state process in panel 2.

Figure 10: Near UV CD spectra of LdCyp for a) native protein LdCyp (—) 40 µM b) LdCyp + 0.2M GdmCl (—) maintaining a characteristic broad shoulder which considerably flattens out for c) LdCyp+ 1.0M GdmCl (—) d) LdCyp + 1.2M GdmCl (—) e) LdCyp + 3.0 M GdmCl (—), thereby depicting the loss of tertiary interactions with increasing denaturant concentrations.

Figure11. Panel1: Bis–ANS fluorescence (10 µM) spectra for LdCyp (4µM) incubated with various GdmCl concentrations, Panel 1 : a) (—) 4 µM LdCyp+0.2M GdmCl b) c) (—) 4 µM LdCyp + 0.7M GdmCl d) (—) 4 µM LdCyp + 1.0 M GdmCl e) (—) 4 µM LdCyp + 1.2 M GdmCl f) (—) 4 µM LdCyp + 1.4 M GdmCl g) (—) 4 µM LdCyp + 1.5 M GdmCl h) (—) 4 µM LdCyp + 1.9 M GdmCl i) (—) 4 µM LdCyp + 3.0M GdmCl. The native spectra (—) 4 µM of LdCyp is shown in Panel 2.

Figure12. Plot of Bis –ANS fluorescence (10 µM) intensity maximum in black circles (●) at maximum blue shifted wavelength of 488nm plotted as a function of GdmCl concentration, with maximum peak intensity obtained for protein sample (4 µM) incubated with 1.2 M GdmCl.

Figure13. F_0/F values plotted for both LdCyp at 25°C shown in inverted triangles in black (▼) and at 45°C, shown in filled black circles (●) as a function of acrylamide concentration.

Figure14. Thermal characterization of the intermediate states of LdCyp through DSC a) 10µM LdCyp b) 10µM LdCyp+0.2MGdmCl c) 10µM LdCyp+0.6M GdmCl d) 10µM LdCyp+1.0M GdmCl, two step (two T_m values) to a single step (one $T_{m\ value}$) transition. The black solid line (—) corresponds to the experimental data points while the solid red line (—) are the fitted lines.

Chapter 3

Figure 1: Crystal structure of cyclophilin from *Leishmania donovani* (LdCyp). The helices H1 and H2 are indicated in red and orange respectively (figure generated using Pymol [40])

Figure 2: Cα RMSD between snapshots and the native crystal structure (2HAQ) plotted as a function of simulation time at different simulation temperatures indicated by the colors (—) 310K_SIM1 (—) 400K_SIM1 (—) 450K_SIM1 and (—) 500K_SIM1.

Figure 3: C^{α} RMSD between native simulation snapshots (—) 310K_SIM1, (—) 310K_SIM2 and the native crystal structure of LdCyP(2HAQ).

Figure 4: C^{α} RMSD between native simulation snapshots (—) 400K_SIM1, (—) 400K_SIM2 and the native crystal structure of LdCyP (2HAQ).

Figure 5: C^{α} RMSDs of snapshots for simulations at 450K a) 450K_SIM1 (—) b) 450K_SIM2 (—) c) 450K_SIM3 (—) and d) 450K_SIM4 (—).

Figure 6: C^{α} RMSD of snapshots for simulations at 500K a) (—) 500K_SIM1 b) (—) 500K_SIM2 c) (—) 500K_SIM3 d) (—) 500K_SIM4 e) (—) 500K_SIM5.

Figure 7: Disnet values for simulation set of 400K_SIM2 shown in solid black (—) lines and 400K_SIM1 in (—) lines

Figure 8: Disnet values between the Surface Contact Networks (SCNs) of the snapshots and the SCN derived from the crystal structure (2HAQ) plotted as a function of simulation time for the simulations 310K_SIM1 (—), 400K_SIM1 (—), 450K_SIM1 (—), 500K_SIM1 (—). Residues constituting the hydrophobic core of LdCyp alone were considered in the construction of the SCNs.

Fig 9: Disnet values for the hydrophobic core of LdCyp at a) 450K_SIM1 b) 450K_SIM2 and c) 450K_SIM3 with the native crystal structure of cyclophilin (2HAQ) as a baseline

Figure10: a) Persf values for 2ns epochs indicated by filled circles considering the hydrophobic core of LdCyp as a function of simulation time with simulation sets of (310K_SIM1,400K_SIM1,450K_SIM1and500K_SIM1). The simulations at different temperatures are indicated by the colors : (—) 310K (—)400K (—)450K (—) 500K.

b) Dlf values for 2ns epoch indicated by filled circles considering the hydrophobic core of LdCyp ; plotted as a function of simulation time with simulation sets of (310K_SIM1,400K_SIM1,450K_SIM1 and 500K_SIM1). The simulations at different temperatures are indicated by the colors : (—) 310K (—)400K (—)450K (—) 500K.

Figure 11 Persf values for 450K simulation temperatures plotted as a function of simulation time a) (—) 450K_SIM2 b) (—) 450K_SIM3 and c) (—) 450K_SIM4.

Figure 12 Dlf values for simulations at 450K plotted versus simulation epochs, a) (—) 450K_SIM2 b) (—) 450K_SIM3 and c) (—) 450K_SIM4.

Figure 13: a) Bur_{epoch} values for every 2ns epoch indicated by filled circles plotted versus simulation time at different simulation temperatures (—) 310K_SIM1, (—) 400K_SIM1, (—)450K_SIM1, (—)500K_SIM1.

b) Fractional secondary content or SSC_{epoch} values for every 2 ns epochs indicated by filled circles plotted as a function of simulation time at different simulation temperatures , the color coding following the same convention given above.

Figure 14: Coordinates derived by MDS, based on the metric Disnet at simulation temperatures of a) 310K_SIM1 and b) 400K_SIM1.The color bar for the plotted points denoting snapshots at indicated time intervals during the course of the simulation.

Figure 15: Coordinates derived by MDS, based on the metric Disnet at simulation temperatures of a) 450K_SIM1 and b) 500K_SIM1.The color bar for the plotted points denoting snapshots at indicated time intervals during the course of the simulation

.

Chapter 4

Figure 1: Spatial positions of the residues which have been mutated are colored in red, located within the hydrophobic core of the molecule. The figure has been generated from the crystal structure of LdCyp (Pdb: 2HAQ) using Pymol (Schrodinger LLC).

Figure 2: Crystals of cyclophilin mutants L120A and V33A grown with PEG 3350 as a precipitant utilizing hanging drop vapor diffusion method. The crystals grew in temperature of 293K in about two weeks time.

Figure 3: DSC thermograms of the cyclophilin mutants a) V33A b) T55A c) L62A d) I85A e) L120A and f) F151V

Figure 4: Fluorescence unfolding curves for LdCyp (—) and mutants a) (—) V33A b) (—) T55A c) (—) L62A d) (—) I85A e) (—) L120A and f) (—) F151V in terms of fraction unfolded at various GdmCl concentrations

Figure 5: A *2Fo-Fc* electron density map for mutants V33A and L120A with cutoff of 1σ with side chains for alanine in place of valine in a) V33A and leucine in b) L120A marked with arrows. Figures generated using Coot v0.7.2.1.

List of Tables

Chapter 1

Chapter 2

Chapter 3

Table 2: <Disnet> values for different simulation temperatures for (310K_SIM1, 400K_SIM1, 450K_SIM1 and 500K_SIM1) calculated with native crystal structure (2HAQ) as baseline averaged over the entire simulation block of 50ns with standard deviations in parentheses.

Table 3: <Disnet> values for different simulation temperatures for (310_KSIM1, 400K_SIM1, 450K_SIM1 and 500K_SIM1) calculated with native crystal structure (2HAQ) as baseline averaged over the entire simulation block of 50ns with standard deviation in parentheses

Table 4: Dlf and persf values averaged over the entire simulation block (between 2 – 50ns) for the hydrophobic core of LdCyp at different simulation temperatures for (310K_SIM1,400K_SIM1,450K_SIM1 and 500K_SIM1). The standard deviations are given in parentheses.

Table 5: Dlf and persf values averaged over the entire simulation block (between 2 – 50ns) for the hydrophobic core of LdCyp at different simulation temperatures for (450K _SIM2, 450K_SIM3 and 450K_SIM4). The standard deviations are given in parentheses.

Table 6: Average persistence of contacts between residues within the core, divided into subsets of S1 (strand –strand or strand-loop contact), S2 (Strand-Helix1) and S3 (Strand-Helix2) at simulation temperatures of 310,400,450 and 500K with simulations sets of (310K_SIM1,400K_SIM1, 450K_SIM1 and 500K_SIM1) and the standard deviation given in parentheses. The abbreviations in bracket represent: L-Loop, S-strand, H1-Helix1 and H2-Helix2.

Table 7: Persistence of contacts between residues within the core, divided into subsets of S1 (strand–strand or strand-loop contact), S2 (Strand-Helix1) and S3 (Strand-Helix2) at simulation temperature of 400K_SIM2, the standard deviation given in parentheses. The abbreviations in bracket represent: L-Loop, S-strand, H1-Helix1 and H2-Helix2.

Table 8: Persistence of contacts between residues within the core, divided into subsets of S1 (strand –strand or strand-loop contact), S2 (Strand-Helix1) and S3 (Strand-Helix2) at simulation temperatures of 310,400,450 and 500K with simulations sets of (310KSIM1,400KSIM1, 450KSIM1 and 500KSIM1) and the standard deviation given in parentheses. The abbreviations in bracket represent: L-Loop, S-strand, H1-Helix1 and H2-Helix2. Perst (Persistence).

Table 9a: Epoch wise persistence values of strategic high persistent contacts in Set S1 for 450K_SIM1

Table 9b: Epoch wise persistence values of strategic high persistent contacts in Set S2 for 450K_SIM1.

Chapter 4

Table 1: Residues chosen for mutation in cyclophilin, their immediate neighbours in surface contact and the number of such neighbours as determined from the native crystal structure (2HAQ)

Table 2: Primer and template sequences for point mutants V33A and L120A with the region of nucleotide sequences (sense and anti-sense template) aimed for mutations highlighted in yellow and the desired change in nucleotide sequence for mutation highlighted in green in the forward and reverse primers.
Table 3: A typical PCR reaction mixture

Table 4: Primer and template sequences for point mutants procured from Genscript Inc. with the region of nucleotide sequences (sense and anti-sense template) aimed for mutations highlighted in yellow and the desired change in nucleotide sequence for mutation highlighted in green in the forward and reverse primers.

Table 5: Data collection summary of cyclophilin mutants of V33A and L120A, the values in the parentheses indicate values in the last resolution shell. The values in the parentheses indicate values in the last resolution shell.

Table 6: Relative thermodynamic parameters of cyclophilin mutants V33A, T55A, L62A, I85A, L120A and F151V

Table 7: Equilibrium thermodynamic parameters for native LdCyp and mutants V33A, L120A, T55A, L62A, I85A and F151V

Table 8: Summary of refinement statistics for V33A and L120A

Table 9: Side chain torsion angles of residues in surface contact with Val 33 and Leu 120 calculated from native crystal structure of LdCyp (2HAQ). In case of V33A and L120Avalues of χ_1, χ_2 in the top row corresponds to values in chain A and bottom row represents that for chain B.

Table 10: Side chain ΔS_m values of cyclophilin mutants along with the RMSD between the coordinates of the native crystal structure (2HAQ) and mutant crystal or modeled structures. Only side chain atoms of the residues in the immediate neighbourhood that is in surface contact of the mutated residue in the native crystal structure were used in the RMSD calculation.

Chapter 1: Equilibrium Unfolding of Cyclophilin from *Leishmania Donovani*: Characterization of the Intermediate State(s)

Introduction

Leishmaniasis, a broad spectrum of diseases caused by leishmania spp is widely prevalent in third world countries, among the poorer sections of the populace. The appearance of strains resistant to drugs (pentavalent antimonials) [**1**], traditionally used as the first line of defense against the pathogen, stress the need to continue the search for alternative drug targets, as second line drugs are generally expensive and have reportedly severe side effects [**2**]. Cyclophilin from *Leishmania donovani* (LdCyp) belongs to the ubiquitous class of peptidyl-prolyl cis-trans isomerases (PPIases), also known to be the intracellular receptor of the immunosuppressive drug cyclosporine A (CsA). In the present study we probed the unfolding of LdCyp by guanidium chloride (GdmCl) by means of fluorescence, circular dichroism (CD), spectroscopy and differential scanning calorimetry (DSC). The results from the studies indicate that most probably the unfolding of LdCyp proceeds via at least one equilibrium intermediate. We have made an attempt to structurally characterize this possible intermediate(s) using multiple spectroscopic tools and molecular dynamics simulations.

Materials and Methods

The protein cyclophilin from *Leishmania donovani* (LdCyp) was purified following a protocol describes elsewhere **[3]**. Intrinsic tryptophan fluorescence, near and far UV-CD, bis-ANS fluorescence, acrylamide quenching, differential scanning calorimetry (with and without denaturant) GdmCl (guanidinium chloride) and molecular dynamics simulation (with the initial coordinates obtained from crystal structure of LdCyp: 2HAQ) was employed to characterize the

unfolding of LdCyp and to characterize the possible intermediate(s) in the unfolding pathway(s) of LdCyp. All spectroscopic and calorimetry experiments were performed in 25 mM potassium phosphate buffer pH 7.5.

Results and Discussion

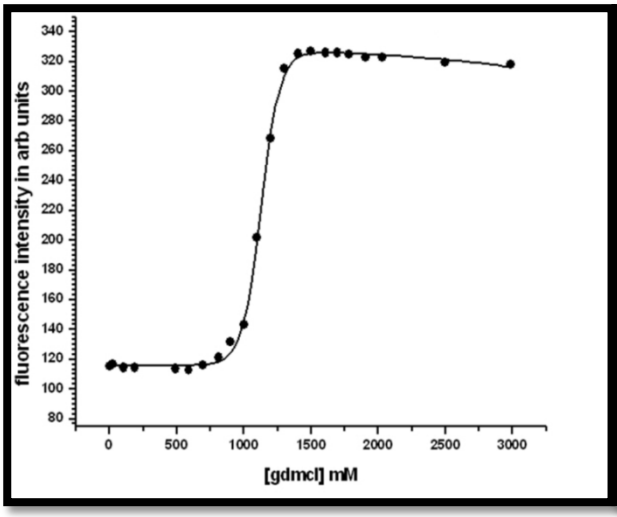
The unfolding curves of LdCyp induced by GdmCl obtained from two independent spectroscopic methods of intrinsic tryptophan fluorescence (Fig 1) and far-UV CD (Fig 2) were non-super imposable with different C_m values (see Table 1) and consequently suggested of an intermediate in the unfolding pathway of LdCyp. A 3-state model (see equation 1) was employed to fit the unfolding curves monitored by CD and fluorescence.

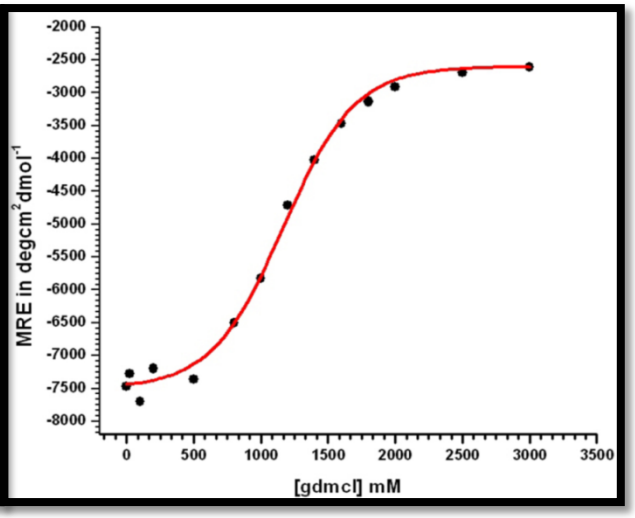
$$S_{obs} = \frac{S_N + S_I e^{\frac{-\Delta G_{NI}}{RT}} + S_U e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (1)$$

Bis-ANS fluorescence experiment also strongly suggested of an intermediate with heightened fluorescence intensities till 1.2 M of GdmCl and declining thereafter. Acrylamide quenching indicated the maximum solvent exposure of the protein to be around 1.4M GdmCl. DSC studies and molecular dynamics simulation studies show that the possible intermediates involved in the unfolding pathways may be due to the detachment of the helices from the core of the protein.

Table 1: Equilibrium thermodynamic parameters of unfolding for LdCyp mediated by GdmCl.

Spectroscopic Tools employed	Thermodynamic parameters					
	$\Delta G_{NI(H2O)}$ kcal mol^{-1}	m_{NI} Kcal mol^{-1} M^{-1}	$\Delta G_{NU(H2O)}$ kcal mol^{-1}	m_{NU} Kcal mol^{-1} M^{-1}	$C_m(N\leftrightarrow I)$ M	$C_m(N\leftrightarrow U)$ M
Intrinsic tryptophan fluorescence	12.13 (± 0.38)	-8.17 (± 0.92)	6.72 (± 0.54)	-5.89(±0.65)	1.47±0.06	1.16 ±0.04
Far UV Circular Dichroism	2.12 (±0.13)	-2.18 (± 0.29)	5.24 (±0.45)	-3.47 (± 0.43)	0.99 ± 0.02	1.51 ± 0.04





. **Fig 1** **Fig 2**

Fig 1: Unfolding curve of LdCyp induced by GdmCl as monitored by intrinsic tryptophan fluorescence with emission maxima points at 360nm, fitted to a 3-state equation. The black circles (●) indicate the fluorescence intensity in arbitrary units at the corresponding denaturant concentrations (in mM) and (—) is the fitted line to the data points.

Fig 2: MRE_{222} [in black circles (●)] values obtained from far-UV CD spectra plotted as a function of the denaturant and fitted to a 3-state equation shown as a red (—) line.

Conclusion

The above results strongly indicate that the unfolding of cyclophilin mediated by GdmCl passes through at least one intermediate state(s) and that the maximum solvent exposure of the protein happens at 1.2 M GdmCl.

Chapter 2: Unfolding molecular dynamics simulation of cyclophilin: characterizing the set of molecular events and identification of the transition state.

Introduction

All atom molecular dynamics (MD) simulations of protein unfolding has provided several key mechanistic insights into the process **[4]**. A central problem in the analysis of such simulations is the specification of an appropriate reaction coordinate to adequately describe strategic conformational transitions during the unfolding of the molecule **[5]**. Cα RMSD, radius of gyration, exposed hydrophobic solvent accessible area and secondary structural content are some of the most commonly used parameters to monitor the progress of unfolding. In such a scenario, definition of parameters based on the representation of protein structures as networks holds great promise **[6]**. In this work, we apply the methodology of protein surface contact networks based on a cutoff on S_m and O_v **[7]** to analyze protein unfolding trajectories. The evolution of these networks, have been monitored in the course of the reaction by the application of suitable metrics. Thus, the reaction coordinate for protein unfolding reduces to a set of specific geometrical relationships primarily between side chains based on a stringently defined mathematical criteria (S_m and O_v).

Materials and Methods

Unfolding molecular dynamics simulations were performed on cyclophilin (coordinates obtained from: 2HAQ) at temperatures 310,400,450 and 500K. Each simulation run was for a minimum duration of 50ns and was repeated five times with independently assigned initial velocities for 450 and 500K in order to confirm the statistical significance or the non-random character in the sequence of molecular events. AMBER 2002 force field was used for energy minimization and

salvation and charge neutralization of the molecule **[8]**. All simulations were performed using NAMD 2.0 **[9].**

Surface generation **[10]** and a cutoff of 0.40 on S_m and 0.08 on O_v **[15]** to construct the networks of the whole protein and that for the core residues were described in detail elsewhere. Cα RMSD was calculated using CHARMM **[11].** Secondary structural content was calculated using STRIDE **[12]**.

SCN's (Surface Contact Networks) were defined on the protein LdCyp at various stages of the unfolding simulation. The SCN for any snapshot was represented as an adjacency matrix and the distance between any two such adjacency matrices was determined by counting the number of links present in one and absent in other and then normalizing by the total number of links present in either of the two matrices.

Thus,

$$Dnet(P,P`) = \frac{\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left|P_{ij} - P`_{ij}\right|}{nL}$$

Where P_{ij} and P'_{ij} are the matrix elements of two adjacency matrices (P, P') corresponding to two SCN's (from their corresponding snapshots) and nL is the number of elements in the set E U E', where E and E' are the set of links prevalent in matrices P and P' respectively. In other words, the distance measure is the ratio of the uncommon links to the union of links between the two matrices.

Results and Discussion

Cα RMSD suggests that the simulation at 310K can be considered native with an average of 1.35 Å ± 0.12 (310KSIM1: from the 3rd to the 50th ns), exhibiting minimal fluctuations. At 400K, an abrupt increase was observed around 4ns (See Fig 1), subsequent to which the average RMSD increased to 2.76 ± 0.47 (400KSIM1 from 4-50ns). Careful visual examination of the snapshots around this region revealed the relative detachment of the helix H2 from its native position w.r.t the barrel, which was responsible for the abrupt rise in RMSD. A similar surge was observed around 15 ns at 450 K (450KSIM1-450KSIM3) rising to 4.0-4.5 Å due to the increased disjuncture of both helices from the body of the barrel. Both these events were reproducible in the other trajectories. At 500K, Cα RMSD rose to about 6-6.5 Å due to increasing non-native like characteristics.

For every MD simulation run from (2-50ns) average Dnet (<Dnet>) recorded a characteristic increase with rise in temperature. At 310K <Dnet> was 0.44 (±0.07) which is to say, that on an average 56% of the links in the core were identified to those found in the crystal structure. Relative to the native simulation (at 310K), little difference was observed at 400K (0.49±0.08), also evident from the superposition of their temporal fluctuations. However, at 450K <Dnet> increased rather abruptly to 0.68(0.12) and further rose to 0.80(0.11) at 500K. At 450K, Dnet values were stable up to ~23ns even though somewhat elevated relative to 310 or 400K. However, subsequent to 23ns a sharp rise in Dnet (see Fig 2) was observed till about 29ns after which it again resumed a more stable progression though at a much higher value. Thus it was tentatively assumed that sudden increase in Dnet from 24 to 29ns could correspond to the transition state, preceded by a relatively stable phase populated by intermediates.

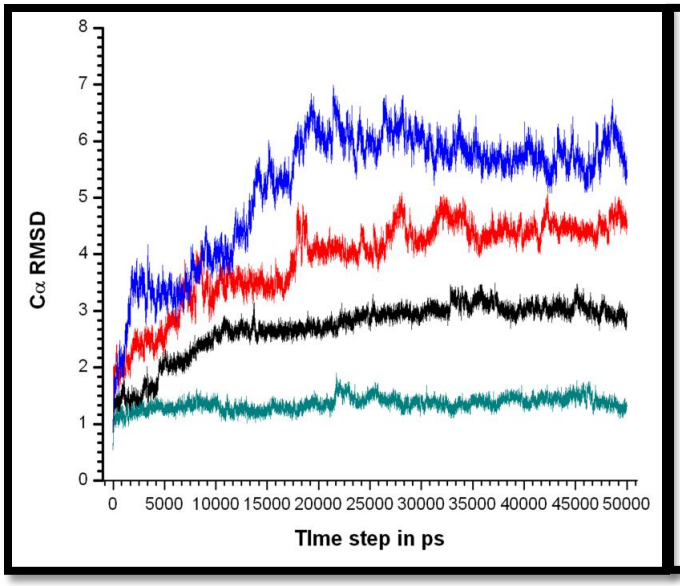


Fig 1

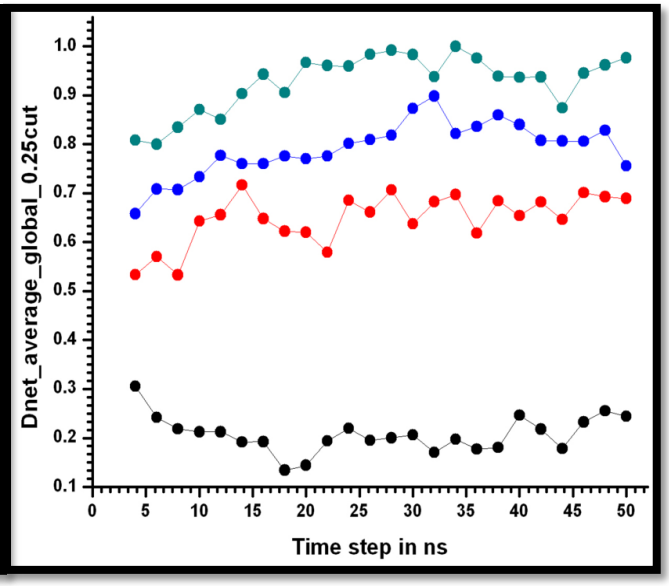


Fig 2

Fig 1: Cα RMSD at different temperatures of a) (—) 310K b) (—) 400K c) (—) 450K and d) (—) 500K calculated with crystal sctructure of cyclophilin (2HAQ) as reference.

Fig 2: Dnet values of the surface contact network of LdCyp at different simulation temperatures of a) (—)310K b) (—) 400K c) (—)450K d) (—)500K calculated with the average structure of 50ns simulation at 310K for the entire protein as reference.

Conclusion

Initial studies of the unfolding trajectories show that the protein is completely unfolded at 500 K and that it increasingly assumes a non-native like character in the simulation temperature of 450 K from 22 ns onwards till 30 ns, with a possible transition state region around 15-20 ns.

Chapter 3: Characterization of thermal stability of cyclophilin mutants correlation with structure.

Introduction

In this present study we have attempted to investigate the thermal stability of cyclophilin mutants (designed away from the active site and within the core) which are at key positions in terms of the protein contact network of cyclophilin. Attempts are also made to structurally characterize these mutants and to draw a correlation between the destabilization of these mutants with their structure and also in terms of network properties.

Materials and Methods

The mutant proteins (V33A, L120A, T55A, F151V, I85A and L62A) were scanned for their T_m's in 25mM potassium phosphate buffer (pH 7.5) with a scan speed of 60^0C/h with a protein concentration of 20µM in a Microcal VP-DSC calorimeter at an approximate pressure of 28 psi.

Cyclophilin mutants of L120A and V33A were crystallized by hanging drop vapor diffusion method and data was collected for them at 2.21 and 2.25 Å respectively.

Results and Discussion

Out of the six cyclophilin mutants T55A is the most stable thermodynamically as it bears minimal side chain alterations due to mutation with the least stable being I85A (See Table 1).

Table 1: Comparison of the T_m values and unfolding processes of cyclophilin mutants.

Mutants	Unfolding Process	T_m's (°C)
V33A	Non-2state(2 T_m's)	48.25,45.12
L120A	Non-2-state(2 Tm)	50.60,47.42
T55A	Non-2-state(2 Tm)	53.82,51.09
F151V	Non-2-state(1 Tm)	47.63
I85A	Non-2-state(1 Tm)	43.11
L62A	Non-2-state(1 Tm)	50.49

There is a wide variation in the Cp (T) values of the cyclophilin mutants (Fig 1) with the least being observed for I85A.

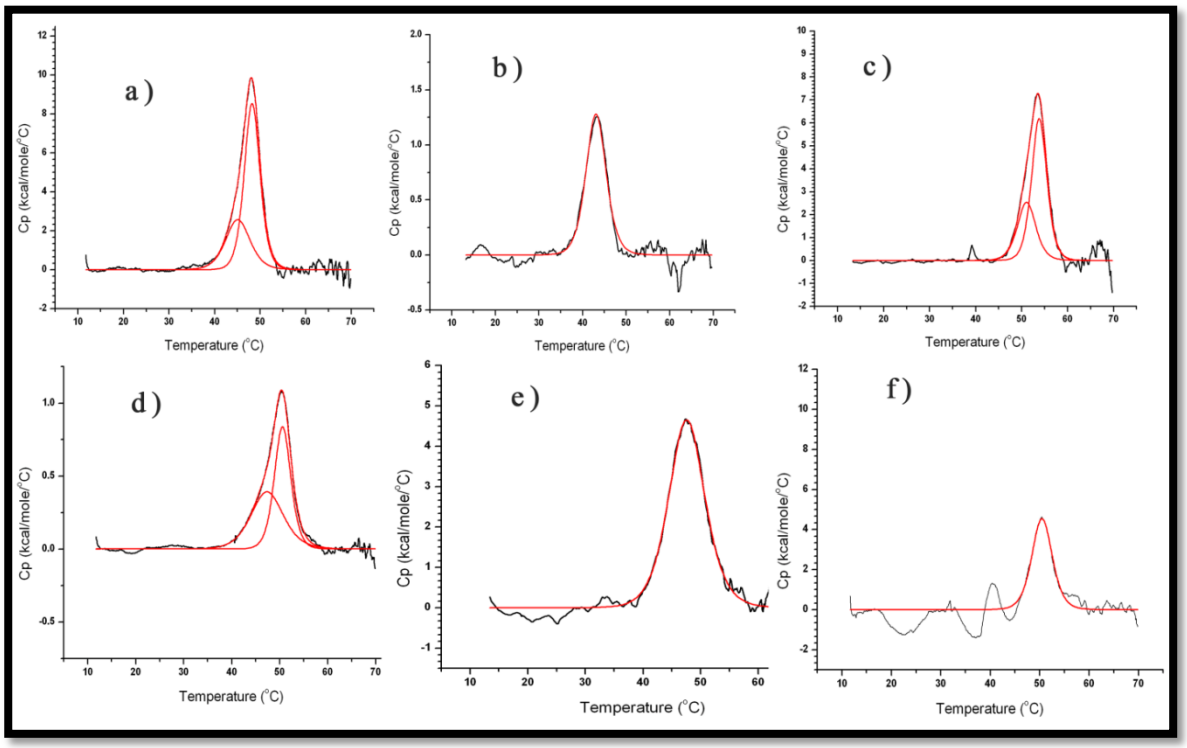


Fig 1: DSC thermogram of cyclophilin mutants of a) V33A b) I85A c) T55A d) L120A e) F151V f) L62A.

Crystallographic data collection summary for the mutants L120A and V33A are also given in (Table 2)

Table 2: Data Collection Statistics for cyclophilin mutant L120A and V33A (the numbers in the parentheses indicate the value in the last resolution shell)

	L120A	**V33A**
Space group	$P4_3$	$P4_3$
Unit Cell parameters	a=b=48.619, c=141.040Å $\alpha = \beta = \gamma = 90^0$	a=b=48.534, c=141.7 Å $\alpha = \beta = \gamma = 90^0$
Resolution range	30.0 - 2.21 Å	50.0 - 2.25 Å
Mosaicity	0.201	0.285
Rmerge %	4.45 (12.63)	4.8(16.9)
Completeness %	99.3 (93.7)	98.3 (86.8)
No of unique reflections	16104 (1497)	13585 (1369)
I/σ(I) ratio	15.3 (6.5)	14.3 (5.5)

Conclusion

Crystallographic studies and thermal characterization studies through DSC indicate that the mutant T55A is the most stable with minimum side chain alteration while I85A is the least stable mutant.

References

[1] Ashutosh et.al. J Infect Dis. 56, (2007), 143–153. http://dx.doi.org/10.1099/jmm.0.46841-0.

[2] D. Sereno et.al. Acta Trop. 74, (2000), 25-31.

[3] A.Chakraborty et.al. J.Biol.Chem. 277(2002), 47451-47460.

[4] A.Li et.al. J.Mol.Biol. 257(1996), 412-429.

[5] R.B.Best et.al. PNAS 102(2005), 6732-6737.

[6] A.Ghosh et.al. Biophys. J. 92(2007), 2523-2535.

[7] R.Banerjee et.al. J.Mol.Biol. 333(2003), 211-226.

[8] D.A.Pearlman et.al. Comput. Phys. Commun. 91(1995), 1-41.

[9] J.C.Phillips et.al. J.Comput.Chem. 26(2005), 1781-1802.

[10] S.Basu et.al. BMC Bioinformatics. 12 (2011), 95.

[11] B.R.Brooks et.al. J.Comput.Chem. 4(1983), 187-217.

[12] D.Frishman et.al. Proteins:Struct.Funct.Genet. 23(1995), 566-579.

Chapter 1

Introduction

1.1 General Overview: Structure and Function of Cyclophilins

Cyclophilins (CyP) are an ubiquitous class of proteins possessing peptidyl prolyl cis-trans isomerase activity, that is catalyzing the *cis-trans* isomerization of peptide bonds involving proline residues **[1,2]** (PPIase/isomerases/rotamases).Other proteins orthologous to cylophilins include FKBP's with identical catalytic activity and a highly conserved fold. Both these proteins and many of their isoforms are known to facilitate protein folding both *in vitro* and *in vivo* **[3-5]**. There are 7 major cyclophilins in humans – hCypA-18a (where 18 denotes molecular mass of 18 kDa), hCypB (Cyp-22/p), hCypC, hCypD, hCypE, hCyp40, and hCypNK (first identified from human natural killer cells) [**6, 7**]. Similarly, 9, 29 and 8 cyclophilins have been identified in the genomes of Drosophila, *Arabidopsis thaliana* and *Saccharomyces cerevisiae* respectively [**6, 8, 9**]. The immunosuppressive drugs Cyclosporine A (CsA) and FK506 when bound to cyclophilin and FKBP respectively, inhibits their isomerase activity [**10**]. However, the isomerase activity of these proteins was found to be independent of immunosuppression [**11**]. Owing to their involvement in immunosuppression, CyPs from humans have been extensively studied and characterized [**12**].

Cyclophilins are found in cellular compartments of most tissues and encode unique functions. The 18-kDa archetypal cyclophilin (CypA) is cytosolic (**Figure 1**) and found in all mammalian tissues, whereas other cyclophilins, whether they have a isolated CLD (Cyclophiln like Domain) or occurs in combination with other domains, are located in the endoplasmic reticulum (ER), the mitochondria, or the nucleus.

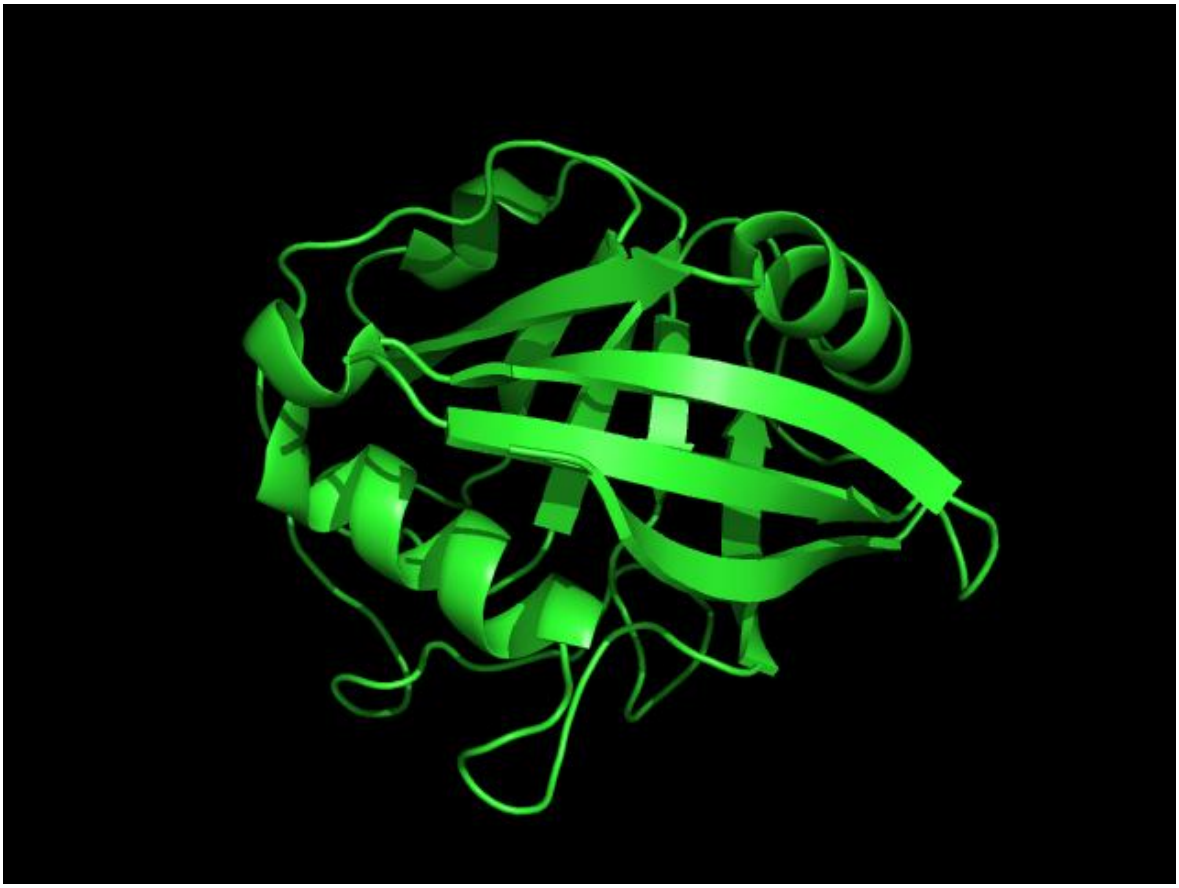


Figure 1: Crystal structure of Human cyclophilin A (1CPL.pdb) with 8 anti-parallel β-strands and two capping α-helices on either side of the barrel; drawn in the cartoon representation using Pymol (Schrodinger LLC)

In mammals, CypA and Cyp40 (molecular mass: 40kDa) are cytosolic whereas CypB and CypC have amino-terminal signal sequences that target them to the ER protein secretory pathway [**6, 13**]. CypD has a signal sequence that directs it to the mitochondria [**14, 15**]; CypE has an amino-terminal RNA-binding domain and is localized in the nucleus [**16**]. On the other hand, Cyp40 has TPRs (tetra-tri-copeptide repeats) and is located in the cytosol [**17**]. Amongst cyclophilins, human CypNK is the largest with pronounced hydrophilicity and positive charge at the carboxyl terminal end, and is found in the cytosol [**18, 19**]. Yeast cyclophilin Cpr1 is a homolog of hCypA with which it has a sequence identity of 65% and is present both in the cytoplasm and

nucleus [**20, 21**]. Cpr2, Cpr3, and Cpr5 have amino-terminal signal peptides directing them to the ER (Cpr2 and Cpr5 [**22, 23**] or the mitochondria (Cpr3 [**24, 25**]; Cpr4 and Cpr8 contain a single CLD (Cyclophilin like domain) and a long amino-terminal signal peptide and are located in vacuoles [**26**]. Lastly, Cpr6 and Cpr7 are homologs of the human Cyp40 protein and have long carboxy-terminal TPR repeats; they associate functionally with homologs of heat-shock proteins and other protein chaperones [**27**].

1.1.2 Biological Functions of Cyclophilins

Cyclophilins are the natural receptor of the immunosupressive drug CyclosporineA (CsA) originally derived from the fungi *tolypocladum inflatum*. Cyclophilins are also involved in a host of other important biological functions that include protein protein interactions, facilitation of protein folding, receptor maturation [**28-35**],functional association with HIV-1 virions [**36**], maintenance of mitochondrial pores [**37**], resolubilization of protein aggregates to restore functionality (Adenosine Kinase from *L.donovani*) [**38**]. CsA is an immunosuppressive drug most commonly used in organ transplants. CsA binds strongly to most of the cyclophilins and CsA-CypA binding to calcineurin inhibits the phosphatase activity and biological function of calcineurin [**39-42**]. The formation of this ternary complex is the prime step leading towards immunosuppression. The crystal structure of the complex has recently been determined to a resolution of 2.8 Å (**Figure 2**) [**43, 44**]. Analysis of the ternary complex reveals that the binding of the CsA-CypA complex to calcineurin increases the stability of the complex, and the complexed proteins remain resistant to proteolytic cleavage [**39**]. Upon binding of CsA to CypA, the charges and hydrophobic surfaces of the drug-protein complex becomes more complementary to the binding site on calcineurin. The CsA-CypA complex binds at the interface

between the catalytic and regulatory subunits of calcineurin (**Figure 2**) thus inhibiting the regulation of the T cell signal transduction pathway.

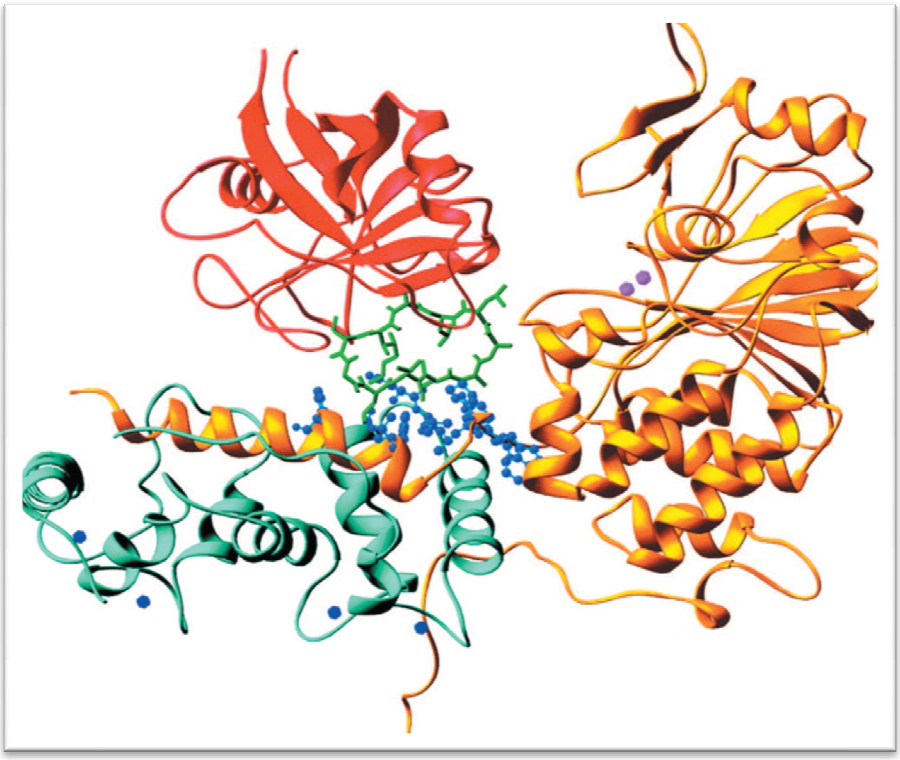


Figure 2: The structure of the ternary complex between the drug cyclosporin A (CsA), human cyclophilin A (CypA) and human calcineurin [44]. The CsA-CypA binary complex lies at the base of the helical arm (shown in ribbon representation in red with CsA shown in green in stick mode) of the catalytic subunit of calcineurin (CnA) that binds the regulatory subunit calcineurin (CnB); it nestles in a hydrophobic groove in intimate contact with both subunits, at a region unique to calcineurin and not found in other phosphatases, and this intimate contact gives the interaction high specificity

 # Adapted from Q. Huai et al. *Proc Natl Acad Sci USA*, 99:12037-12042.

Several protein-folding processes depend on the catalytic and/or chaperone-like activities of cyclophilins. For example, CypA regulates (promotes) both the formation and the infectivity of virions of the human immunodeficiency virus (HIV)-1 [**45-51**]. Capsid (CA)-specific restrictions are determinants of retroviral tropism in mammalian cells. Restriction factors from humans are postulated to target the capsid of human immunodeficiency virus type 1 (HIV-1) resulting in decreased accumulation of viral cDNA, if not bound to cyclophilin A**.** Furthermore, CypA

interacts with HIV accessory proteins, such as the viral protein R (Vpr) and Nef to facilitate a step in the viral life cycle between penetration and reverse transcription [**52, 53**].

Cyclophilins can also act as modulators of protein function. The mammalian cyclophilin Cyp40 is part of the steroid receptor complex and can form a dimeric complex with the heat-shock protein Hsp90, a process not affected by CsA [**54, 55**]. In yeast, the Cyp40 homologs Cpr6 and Cpr7 also associate with Hsp90 homologs and have analogous functions [**56**]. A mammalian Cyp40 has been shown to regulate the activity of the transcription factor c-Myb [**57**], whereas CypA has been associated with YY1, a zinc-finger suppressor of gene transcription **[58],** and Zpr1, an essential zinc-finger protein [**59**]. In addition, the ER-specific cyclophilin CypB can form a complex with the peptide hormone prolactin to induce transcription of a range of genes [**60**].

The involvement of mitochondrial permeability transition pore (mPTP) in Aβ-induced mitochondrial dysfunction plays a siginificant role in the patho-physiology of Alzheimer's disease including perturbation of intracellular calcium regulation, ROS (Reactive Oxygen Species) generation, release of pro-apoptotic factors and changes in mitochondrial morphology. Calcium and ROS are strong inducers of the mPTP formation which lead to increased ROS generation, decreased ATP production and apoptogenic substance release accompanied by mitochondrial swelling [**61-64**]. For instance, the absence of cyclophilin D, a key component of mPTP, protects against Aβ-mediated mitochondrial, neuronal and synaptic dysfunction. Studies in animal models have shown that the mPTP formation can be efficiently blocked by the addition of a cyclophilin D inhibitor, cyclosporine A (CsA) or by ablation of Cyp D [**64-66**].

Further regulatory roles of cyclophilins can be inferred from the fact that inactive aggregates of adenosine kinase from *Leishmania donovani* (LdAdk) used as a model substrate,

was reactivated by cyclophilin A (LdCyP) from the same source in an isomerase independent manner *in* v*itro* by disaggregating its inactive oligomers [**67**]. Besides disrupting preformed aggregates, LdCyP also prevented re-aggregation of the newly formed active protein that is generated after productive refolding from a urea-denatured state [**68**].

1.1.3 Crystal Structures of Cyclophilin A

The first crystal structure of cyclophilin A was from human (hCypA) in 1991 and was published by two independent groups using X-ray crystallography **[69]** and NMR spectroscopy **[70]**. Since then several crystal and NMR structures of CyPs from a variety of sources have been solved **[71-76].** Most of these protein structures exhibit a fold very similar to human hCypA. In addition, several CyP structures are now available with the protein complexed with substrate (ala-ala-pro-phe-p-nitroanilide) or the immunosuppressive drug CsA. The structure of hCypA consists of a β-barrel (**Figure 1**) composed of eight antiparallel β strands, with two α-helices enclosing the barrel from either side. Fifteen hydrophobic residues within the β-barrel constitute the unique hydrophobic core of the molecule (**Table 1)**. The active site of the protein is located primarily on the face of the barrel and is constituted of residues given in **Table 1**.

Table 1: Residues of hCypA constituting the hydrophobic core, the cyclosporine binding site and the native substrate binding active site. The residues common to the cyclosporine binding site and the active site are marked in bold.

Residues of hCypA in the hydrophobic core	Residues of hCypA in the CsA binding site	Residues involved in substrate binding in hcypA
Val6	**Arg55**	**Arg55**
Phe8	**Phe60**	**Phe60**
Val20	**Met 61**	**Met61**
Phe22	**Gln63**	**Gln63**
Leu24	Gly72	**Ala101**
Phe36	**Ala101**	**Asn102**
Leu39	**Asn102**	**Gln111**
Tyr48	Ala103	**Phe113**
Phe53	**Gln111**	**Trp121**
Ile56	**Phe113**	**Leu122**
Leu98	**Trp121**	**His126**
Met100	**Leu122**	
Phe112	**His126**	
Ile114		
Phe129		

There is considerable overlap of residues between the site of the protein responsible for PPiase activity and the cyclosporin A (CsA) binding site (**Table 1**). The overall fold of hCypA is conserved in hCypB (**Figure 3**), the main difference being in the two loop regions (residues 19-24 and 152-164; as numbered in hCyPA), the amino and carboxyl termini [**77**]. On the other hand, hCyp40 consists of a 'cyclophilin like domain' (CLD) similar to the fold of hCypA, linked to tetratricopeptide repeats (TPR: generally found in proteins implicated in stress response), consisting of seven helices of variable lengths. Crystal structure of cyclophilin A from *plasmodium falciparum* (PfCyp19) with 162 amino acid residues [**78**] suggest that it is a

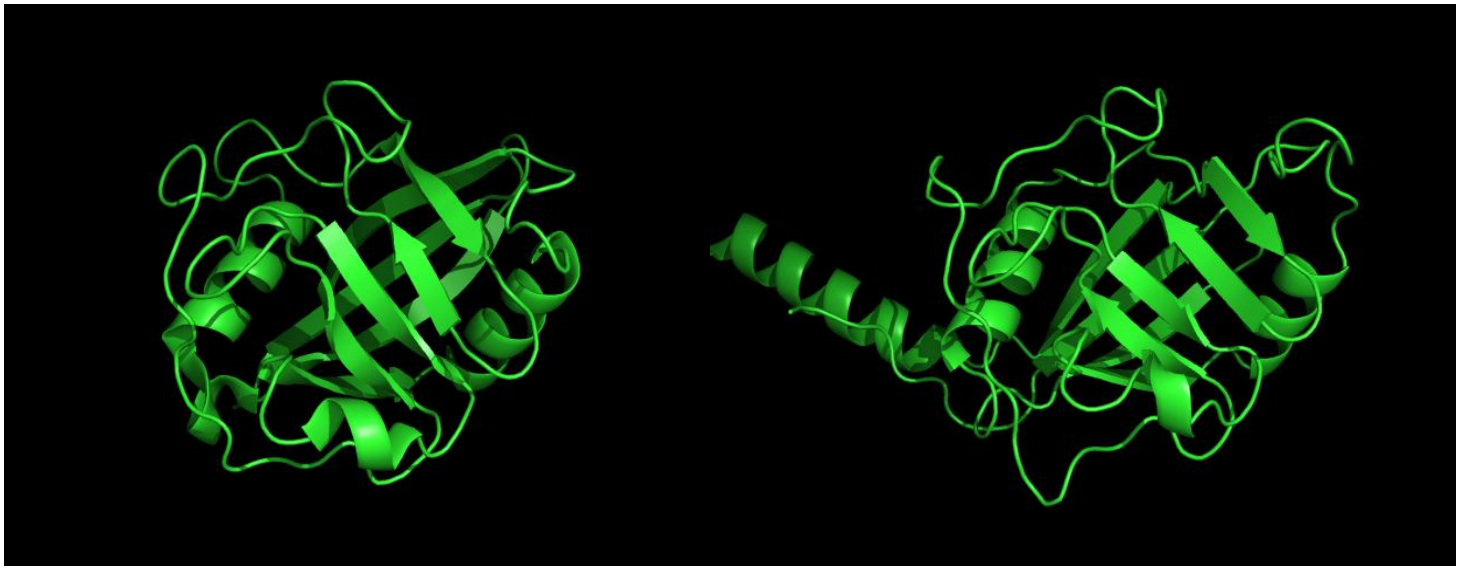


Figure 3: The overall conservation of folds in human cyclophilin A(hCypA: 1CPL.pdb) to the left and human cyclophilin B (hCypB:1ZXC.pdb) to the right

globular molecule with a fold dominated by an eight-stranded anti-parallel β-barrel capped at either end by two α-helices similar to hCypA. There is a short segment of 3_{10} helix and five loops linking the secondary structural elements. A crater-like active site is located on one side of the protein which is entirely conserved with respect to hCypA. With respect to hCypA, a relatively high degree of variation in sequence identity is observed in *E.Coli* cyclophilin with a sequence identity of 34% (**Figure 4**), In addition, a tryptophan residue in the active site of the protein (conserved in most of the cyclophilins) crucial for binding of CsA is converted to phenyl-alanine, possibly justifying the protein's inability to bind to CsA, without however hindering it's isomerase activity. Both the sequence identity and structure of cyclophilin A from *leishmania donovani* (2HAQ) and *trypanosoma cruzi* (1X07) show a high degree of conservation amongst themselves (**Figure 5**).

```
Human           --MVNPTVFFDIAVDGEPLGRVSFELFADKVPKTAENFRALSTGE------KGFGYKGSC 52
L. donovani     EPEVTAKVYFDVMIDSEPLGRITIGLFGKDAPLTTENFRQLCTGE------HGFGYKDSI 75
T. cruzi        MPVVTDKVYFDITIGDEPVGRVVIGLFGNDVPKTVENFKQLASGE------NGFGYKGSI 54
Yeast           --MS--QVYFDVEADGQPIGRVVFKLYNDIVPKTAENFRALCTGE------KGFGYAGSP 50
Plasmodium      --SKRSKVFFDISIDNSNAGRIIFELFSDITPRTCENFRALCTGEKIGSRGKNLHYKNSI 58
E. coli         ----------MVTFHTNHGDIVIKTFDDKAPETVKNFLDYCREG---------FYNNTI 40

                  *    *  *         *                                ***        *
Human           FHRIIPGFMCQGGDFTRHNGTGGKSIYGEKFEDENFILKHTGPGILSMANAGPNTNGSQF 112
L. donovani     FHRVIQNFMIQGGDFTNFDGTGGKSIYGEKFADENLNVKH-FVGALSMANAGPNTNGSQF 134
T. cruzi        FHRVIRNFMIQGGDFTNFDGTGGKSIYGTRFDDENLKIKH-FVGAVSMANAGPNSNGSQF 113
Yeast           FHRVIPDFMLQGGDFTAGNGTGGKSIYGGKFPDENFKKHHDRPGLLSMANAGPNTNGSQF 110
Plasmodium      FHRIIPQFMCQGGDITNGNGSGGESIYGRSFTDENFNMKHDQPGLLSMANAGPNTNSSQF 118
E. coli         FHRVINGFMIQGGGFEP--GMKQKATKEPIKNEANNGLKNTRGTLAMARTQAPHSATAQF 98

                *       **           *                                *
Human           FICTAKTEWLDG--------KHVVFGKVKEGMNIVEAMERFGS-----RNGKTSKKITIA 159
L. donovani     FITTAPTPWLDG--------RHVVFGKVLDGMDVVLRIEKTKTN----SHDRPVKPVKIV 182
T. cruzi        FVTTAPTPWLDG--------RHVVFGKVVEGMDVVKKVENTKTG----LNDKPKKAVKIN 161
Yeast           FITTVPCPWLDG--------KHVVFGEVVDGYDIVKKVESLGS-----PSGATKARIVVA 157
Plasmodium      FITLVPCPWLDG--------KHVVFGKVIEGMNVVREMEKEGA-----KSGYVKRSVVIT 165
E. coli         FINVVDNDFLNFSGESLQGWGYCVFAEVVDGMDVVDKIKGVATGRSGMHQDVPKEDVIIE 158

Human           DCGQLE 165
L. donovani     ASGEL- 187
T. cruzi        DCGVL- 166
Yeast           KSGEL- 162
Plasmodium      DCGEL- 170
E. coli         SVTVSE 164
```

Figure 4: Sequence alignments of cyclophilins from human (2CPL.pdb), *L. donovani* (2HAQ.pdb), *T. cruzi* (1XO7.pdb), Yeast (1VDN.pdb), *Plasmodium falciparum* (1QNG.pdb) and *E. coli* (2NUL.pdb). The cyclosporin binding site is marked with stars.

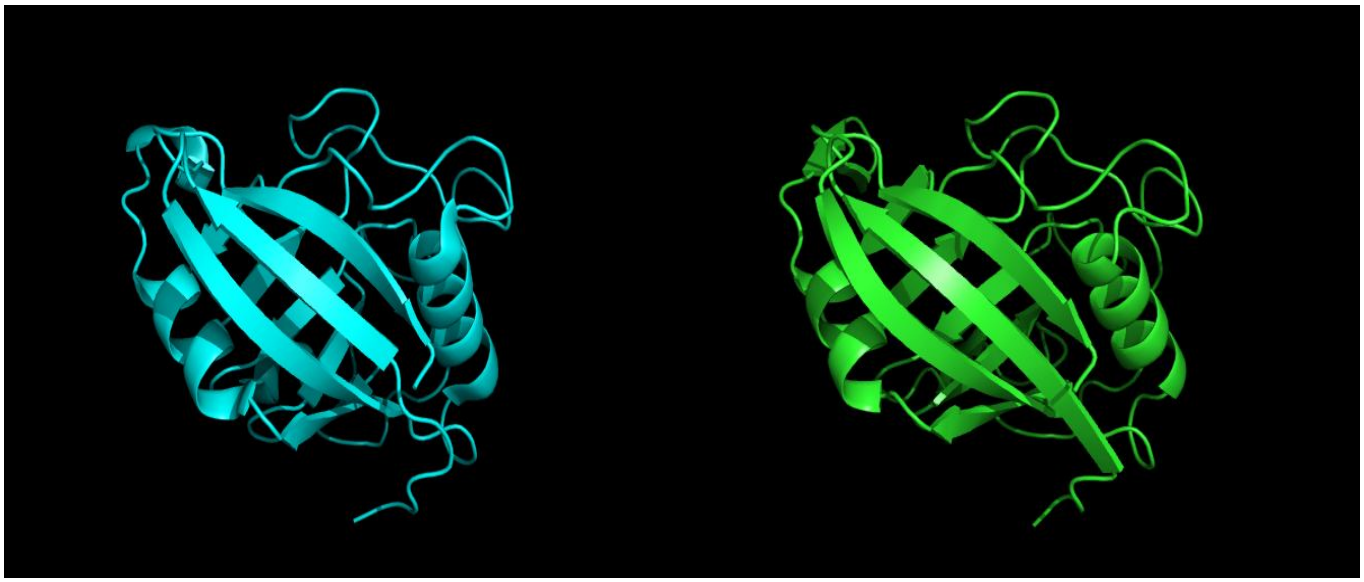


Figure 5: Crystal structures of cyclophilins from a a kinetoplastid *leishmania donovani* (LdCyp: 2HAQ.pdb) to the left (in cyan) and from a trypanosomatid *trypanosoma cruzi* (1X07.pdb) to the right (in green)

1.1.4 Cyclosporine A and Other Non-immunosuppressive Analogs

CyclosproineA (CsA) the immunosuppressive drug binds strongly (**Figure 6**) with most of the cyclophilins apart from CypA of *E.Coli*. CsA is a cyclic undecapeptide composed of unnatural and methylated amino acid side chains (**Figure 6**).

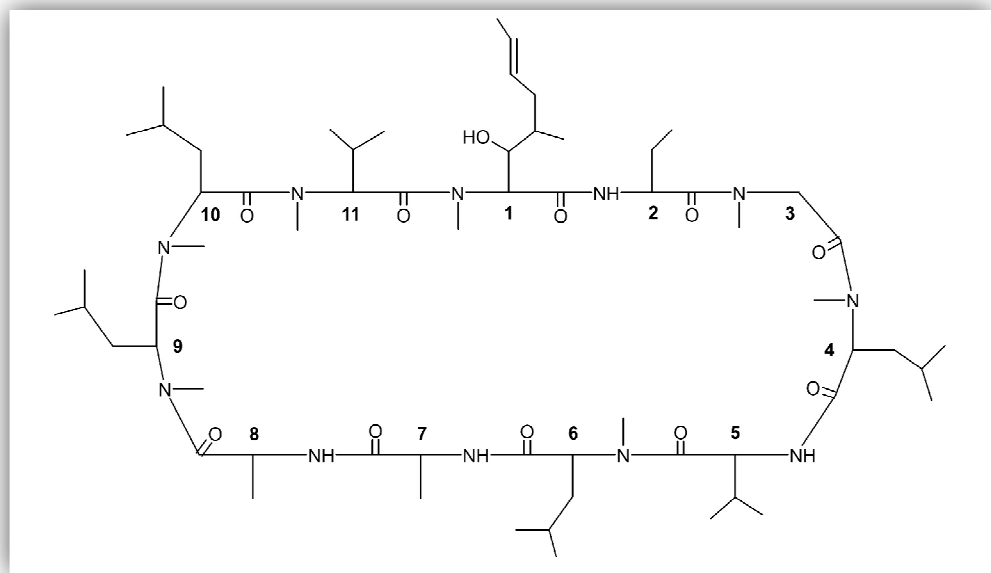


Figure 6: Immunosuppressant drug Cyclosporin A derived from fungi *tolypocladum inflatum* with unnatural and methylated amino acid side chains.

#Adapted from Bua .et.al (2002) *Parasitol*. 22, 340-351.

CsA is often discussed using a two-domain description. The CyP-binding domain comprises residues 1, 2, 3, 9, 10 and 11 whilst an exposed "effector domain" is composed of residues 4, 5, 6, 7, and 8 [**79**]. The effector domain forms a composite "effector surface" with residues from the enzyme, which binds to and inhibits calcineurin, thereby blocking T cell signal transduction, finally leading to immunosupression. Chemical modifications of the side chains of CsA at residue positions of 4,5,6,7 and 8 have resulted in certain CsA analogs that possess potent anti-

trypanocidal/anti-parasitc and anti-fungal activity besides being non-immunosuppressive. A substitution in residue 4 from isoleucine to methylated isoleucine (MeIle4) results in the CsA analog SDZ NIM 811 with anti HIV-1 activity [**80**]. On the other hand, in *T. cruzi* two CsA derivatives H-7-94 & F-7-62 (**Figure 7**) with the substitution of a methyl group in native CsA by an aromatic side chain in residue 1 were found to exhibit potent anti-parasitic activity, without being toxic to mammalian cells [**81,82**].

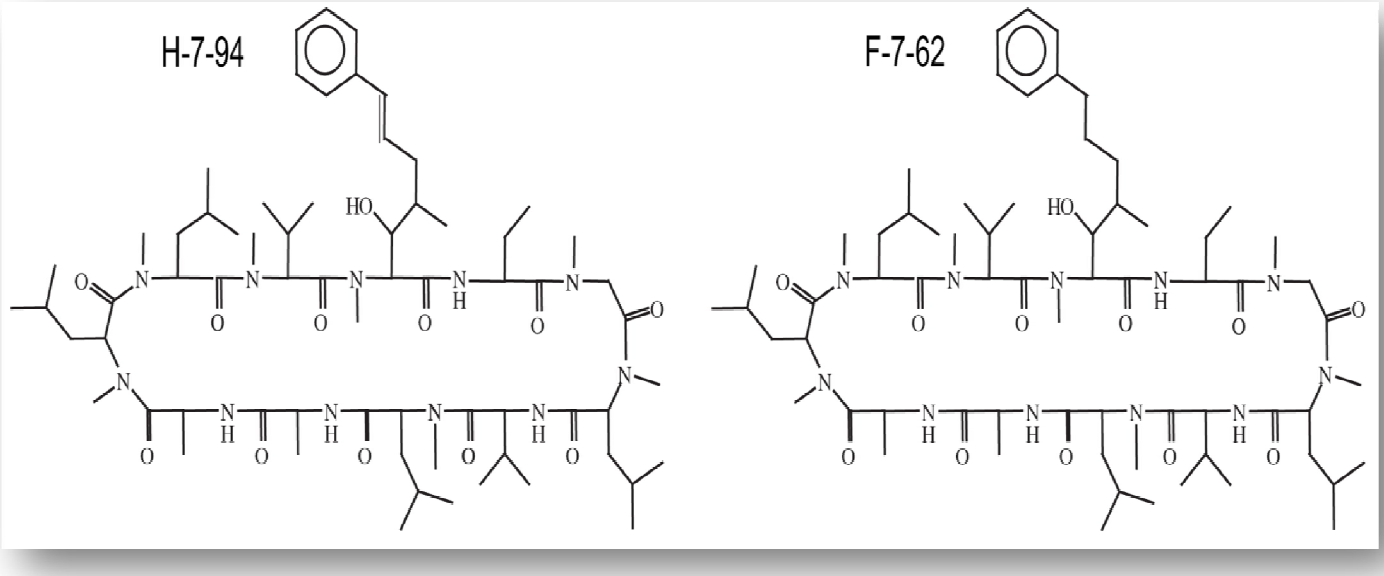


Figure 7: Two of the non-immunosuppressive analogs of CsA (H-7-94 and F-7-62) that have been found to exhibit potent anti-parasitic activity against *T. cruzi*. The modifications are addition of a benzene ring in H-7-94 and a 'double to single bond' in the residue Bmt 1 of CsA in F-7-62.

 # Adapted from Bua et al. (2008) *Parasitology*, 135, 217-228.

Other CsA analogs targeting residue 1 included SDZ PSC-833(3' keto MeBmt1) and SDZ 033-243 (8'-O-Me-dihydroMeBmt1) were found to inhibit the growth of a wide range of parasites (*C. parvum*, *P. palcifarum*, *P. vivax* and *T. gondi*) with little or no immunosuppressive effect in

humans [**83**]. Recently another CsA derivative (Debio 025) was reported to interfere strongly in the hepatitis C viral life cycle.

1.2 Unfolding of Cyclophilin from *leishmania donovani* (LdCyp)

1.2.1 Introduction to Protein Folding

Protein folding is usually a complex process. Proteins structure themselves into specific three-dimensional structures, through diverse conformational changes. The classical outlook of protein folding describes the process as a practically sequential series of discrete intermediate conformations. In contrast, the energy landscape theory of folding considers folding as the progressive organization of an ensemble of partially folded structures through which the protein moves forward on its way to be the natively folded structure. In terms of evolution, proteins are believed to possess rugged funnel-like energy landscape influenced toward the native structure. The common view is that proteins fold rapidly because of random thermal motions that cause conformational changes that leads to energetically downhill processes towards the native structure, a principle that is inherently embedded in the funnel-shaped energy landscapes.

In vitro protein unfolding has been studied either thermally or by the application of chaotropic denaturing agents such as urea or guanidinium chloride [**84-92**]. Broadly, the experimental approaches to characterize protein unfolding can be either kinetic in time scales of 100μs to few minutes, wherein the progress of unfolding is monitored instantaneously through biophysical/spectroscopic methods or equilibrium (30 mins – 24h) where the protein is treated typically with denaturants and measurements are made subsequent to equilibration of the system. Spectroscopic, biophysical equipments and chemicals, such as a thermostated fluorimeter and common denaturants, are adequate to determine the conformational stability of a protein. In this regard, the most important experiments consist of the preliminary ones done to

establish the minimum number of species (conformations) accumulating in the equilibrium. In case of proteins with non-functional equilibrium intermediates, ascertaining the relevant stability of the protein (the free energy difference between the native conformation and the intermediate) is most important, and it elucidates very valuable structural information on the intermediate(s) to be derived when protein variants are compared to wild type using equilibrium ϕ value analysis. It is general that in the folding/unfolding equilibrium of small proteins the only conformations that significantly populate to measurable molar fractions are the native folded state and the unfolded state (N⇌U) [**93, 94**]. However, for most proteins, additional intermediate conformations (less folded than the native state but more folded than the denatured state) accumulate as the solution conditions become more destabilizing of the native conformation, nevertheless not fully stabilizing of the denatured state [**95**]. The number of species that will appear as the solution conditions are shifted from native to denaturing will depend on the specific shape of the folding landscape in native conditions and on the changes that the specific denaturing agent used to probe the equilibrium will induce in that shape. The more complex the landscape, the more likely a higher number of intermediates will be observed. On the other hand, a particular denaturing agent (say, urea) may not divulge the presence of a potential intermediate while other (say, temperature) may do so [**96**]. For proteins displaying equilibrium intermediates, the global equilibrium between the native and the unfolded states can be divided into partial equilibria between the different species, each one governed by an equilibrium constant and, therefore, by a Gibb's free energy difference. For the simplest, non two-state equilibrium scenario with a single intermediate accumulating at moderate denaturing conditions (N⇌I⇌U), the global conformational stability can be partitioned into the relevant conformational stability (the free energy difference between the native and intermediate conformation, ΔG_{NI}) and the residual

conformational stability of the intermediate (its free energy difference with the denatured state, ΔG_{NU}) [**97,98**]. The terms (relevant and residual) chosen to qualify these stability fractions adding up to the global conformational stability is assuming that the intermediate will be devoid of biological activity. While in some cases [**99**] the existence of intermediate(s) is clear with a characteristic plateau of the spectroscopic/optical signal (in 3 state pH denaturation of sperm-whale apomyoglobin), in others the existence of such intermediate is only intuitive and is usually judged by the non-coincident unfolding curves derived from two or more sets of independent experiments involving different spectroscopic probes. There is considerable interest regarding the nature and structure of these intermediate states which lie between the native and the unfolded states.

The existence of stable or metastable intermediates in equilibrium unfolding was often accompanied by hyper fluorescence of the tryptophan residue with increasing denaturant concentrations. One such example was that of apomyoglobin [**100]** where changes in the protein tryptophan (Trp) fluorescence spectrum that accompany increasing urea concentration unambiguously reveal accumulation of some stable or metastable intermediate states (**Figure 8**) whose intrinsic tryptophan is relatively elevated as compared to both the native (N) and the unfolded states (U). Besides, this change in tryptophan fluorescence spectrum is characterized by two isosbestic points at 315 and 370 nm that also confirm the occurrence of at least two transitions detectable by tryptophan fluorescence. Application of salts and pH variations were often employed to extract otherwise elusive intermediate states from an apparent two state transition [**101, 102**]. Conformational transitions in equine cytochrome C [**101**] seemed to involve only the native and the unfolded states but on increasing the salt concentration the two transitions were separated by slightly decreasing the stability of the native state and significantly

increasing the stability of the "molten globule" state. This results in the two-step unfolding transition involving the molten globule intermediate state to be approximated by a 3 state mechanism. Application of salt (Na_2SO4) played a significant role in regulating the stabilities of both the native as well as the intermediate states of SH3 domain of PI3 kinase [**102**] where a four state equation was employed to explain the protein's folding with two intermediates in the kinetic experiments. Another approach to untangle the "intermediate state(s)" from an apparent two state transition involves the application of multiple spectroscopic probes to the system under investigation and the non-superposition of these unfolding curves lead to the existence of equilibrium intermediates [**103-105**].

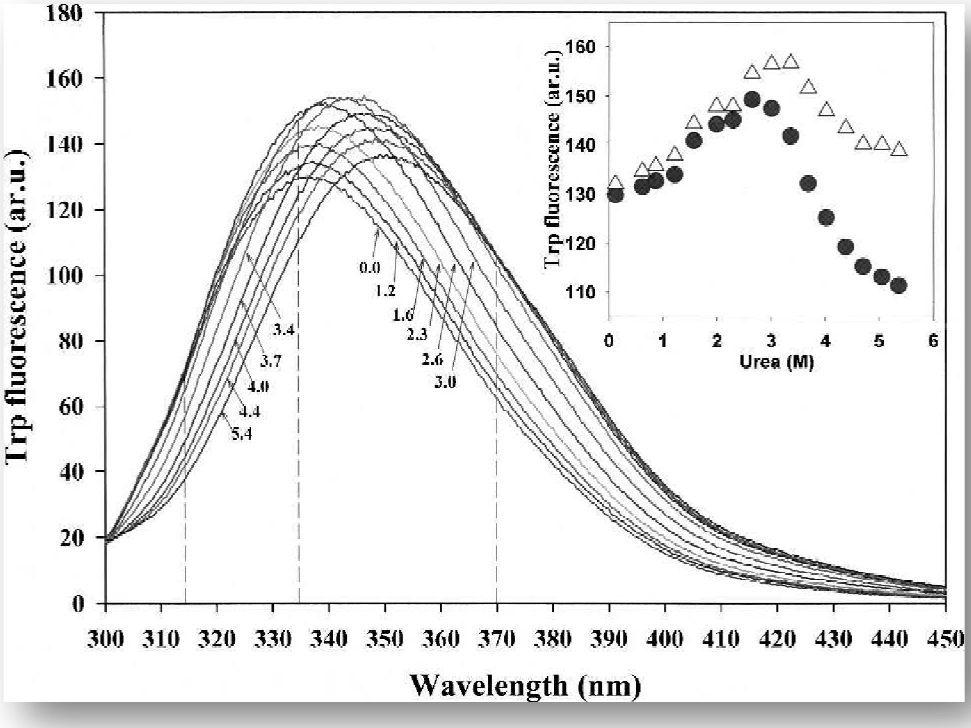


Figure:8 Trp fluorescence spectra of apomyoglobin at various urea concentrations (shown by arrows and numbers near curves). Dashed lines indicate 335 nm wavelength and the pseudoisosbestic points at 315 and 370 nm. (Inset) Dependence of protein Trp fluorescence intensities at 335 nm (filled circles) and the spectrum maximum intensity (open triangles) on urea concentration.

#Adapted from Baryshnikova et al. *Protein Sci* (2005), 14:2658–2667.

Identification and characterization of intermediate state(s) are important because folding/unfolding intermediates serve as "signposts" on folding pathways. In the absence of such a signpost, it is difficult to define a folding pathway, and to distinguish one pathway from an alternative pathway that may be used in different folding conditions [**103**]. Structural investigation into these intermediate state(s) shows most of them to be of "molten globule" type. Common properties of well characterized stable equilibrium intermediates of α-lactalbumin, carbonic anhydrase B and bovine growth hormone [**106-117**] led to the coinage of a unique physical state of the protein molecule the "molten globule". A typical molten globule is a partially folded/unfolded state that differs substantially from the native state topology, fold and tertiary interactions whilst conserving most of the secondary structural content of the native state (**Figure 9**).

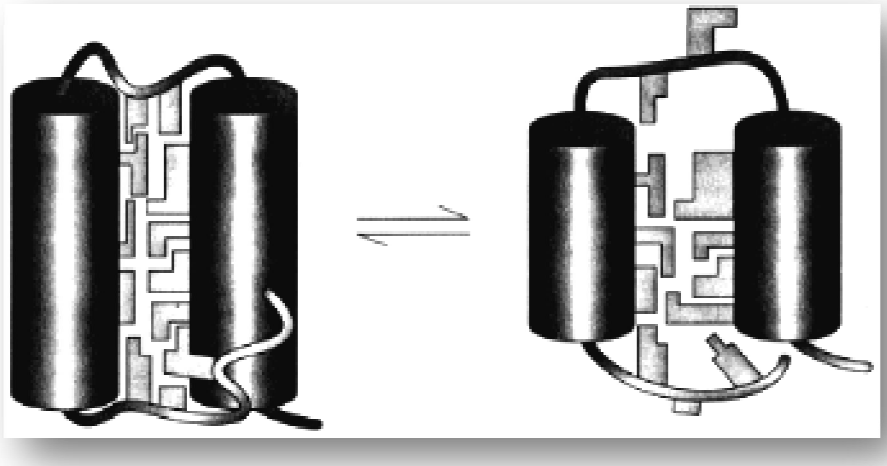


Figure 9: A schematic diagram depicting a close packed native folded state of the protein (left) and its reversible transition into a "molten globule" state (right) with disrupted tertiary interactions with conservation in the secondary structural elements. The diagram is adapted from http://:www.protein.bio.msu.ru.

A number of globular proteins undergo acid and alkaline transitions which produce partially unfolded states analogous to that of α-lactalbumin. Best characterized examples include the acid state of bovine and human carbonic anhydrase B [**118-120**] and the acid and alkaline states of β-lactamase from *Bacillus cereus* [**121**]. The compact intermediate conformational states of the β-lactamase is, however, stable only at high ionic strength at acid or alkaline pH, and the protein is close to fully unfolded at extreme pH's when the ionic strength is too low. This behavior is analogous to the salt-induced formation of the secondary structure in disulfide-reduced α-lactalbumin. Molten globules are classified into "wet" and "dry" molten globules. A dry molten globule is distinguished from its wet counterpart in the extent of perturbation in the native tertiary structure **[122].** In DMG's although there is disruption of the tertiary interactions with subsequent alterations in the secondary structure; it still prevents water from penetrating into the molecular interior whereas in a WMG the molecular interior is solvated **[123-125]**. Experimentally, DMGs can be unambiguously distinguished from WMGs, by the inability of the fluorescent dye ANS to penetrate and bind to the core of a DMG and also the protection of its backbone NH hydrogens from exchange **[126].**

Although the stability and folding of two-state proteins are understood far better than those of more complex ones, a consensus view is still lacking [**127**]. Understanding the structure and energetic of protein folding intermediates (either transient or at equilibrium) is most important to illuminate the folding reaction and the relationships between protein dynamics, function and misfolding [**128–130**]. Till date, majority of our current knowledge on protein stability has been derived using water soluble proteins. The stabilities of membrane proteins are not so well deciphered and more work in this field is anticipated [**131**]. On the other hand, thermal unfolding coupled with chemical unfolding, has supplied most of our stability data. Work on pressure

unfolding is complementing our knowledge of the phenomenon [**132**], and mechanical stability will receive much attention [**133**]. Recently the focus is primarily shifted to single molecule experiments that can also be used to determine conformational stabilities [**134**] and might even be useful to define the characteristics of partially unfolded intermediates.

1.2.2Protein Folding/Unfolding Studies on Cyclophilin A

There are very few reports of equilibrium protein unfolding studies on cyclophilin A, one such report is equilibrium unfolding studies of cyclophilin A from *mycobacterium tuberculosis* (MtCypA) [**135**]. MtCypA lacks a tryptophan in the active site of the molecule and hence a mutation of isoleucine to tryptophan residue was engineered in the hydrophobic core of the molecule for spectroscopic investigations. In this report (MtCypA) sharing a sequence identity of 48% with LdCyP under the application of urea reversibly unfolded following a 2 state mechanism as monitored by the far-UV circular dichroism spectroscopy while the tertiary interactions seemed to follow a 3 state unfolding curve as monitored by tryptophan fluorescence. Dimeric mip-like peptidyl prolyl cis-trans isomerase from *E.Coli* [**136**] similar to cyclophilin A unlike MtCypA seemed to follow a two state transition as suggested both by far-UV unfolding CD and tryptophan fluorescence data.

Keeping in mind the importance to characterize the intermediate states of protein during folding/unfolding equilibrium unfolding of cyclophilin from Leishmania donovani (LdCyp) has been chosen to study equilibrium protein unfolding and to characterize the intermediate state(s) if any, discussed extensively in Chapter2. The unfolding process of LdCyp could be best approximated by a 3-state mechanism involving at least one intermediate state bearing stark resemblances to the properties of a wet molten globule.

1.3 Network Based Metrics to Characterize Unfolding of LdCyP Through Molecular Dynamics Simulation.

1.3.1 Introduction

The process by which a linear amino acid chain folds into a three-dimensional functionally active protein remains an unsolved problem. Various protein folding models have been proposed such as the framework model [**137-139**] which postulates that the first event in folding is the formation of secondary structural elements. Secondary structural elements later diffuse and pack to form the stable native protein [**140-142**]. On the other hand, in the hydrophobic collapse model [**143-145**] the initial rapid formation of the hydrophobic core(s) precedes the formation of secondary and tertiary structure. A common feature of these models is that there exists a well-defined order of events for folding, that is, there is a single or a small number of folding pathways. In contrast, the jigsaw puzzle model [**146**] assumes that there can be a large number of parallel pathways and each individual protein may have its own order of events for folding. None of the above models is capable of comprehensively explaining all the experimental observations. The major hindrance to definitively characterizing folding mechanisms is the difficulty in determining the structure of the protein at each step of the folding/unfolding process.

1.3.2. Molecular Dynamics Simulations to Study Protein Folding/Unfolding.

In most cases the sheer speed of protein folding eludes attempts to experimentally characterize the intermediate states. However, the advent of powerful computers coupled to the development of molecular dynamics simulation (MD) has provided important insights into the phenomenon [**147,148**], which samples a possible set of dynamical conformational states whilst unfolding.

Such a description of dynamics at the atomic level contributes in clarifying many aspects of bio-molecular function. MD simulations have been employed to study a wide spectrum of problems, ranging from folding/unfolding [**149-151**], catalytic activity [**152-155**], network fluctuations, domain motions [**156**] and protein stability [**157**]. The timescales of molecular dynamics simulations primarily to study protein unfolding has ranged from few picoseconds in earlier studies [**158,159**] to few milliseconds [**160-166**] in the recent past. Several recent developments increasing both the temporal resolution of experimental techniques and the time scales of MD simulations promises to expand the scope of biological systems studied by these methods. These developments also allows for the systematic validation of the models and protocols underlying MD simulations [**167**].

The first protein molecule to undergo MD simulation was BPTI (Bovine Pancreatic Trypsin Inhibitor) (**Figure10**) [**168**]. This early MD simulation gave estimates of the magnitude, correlations and decay of fluctuations about the average structure.

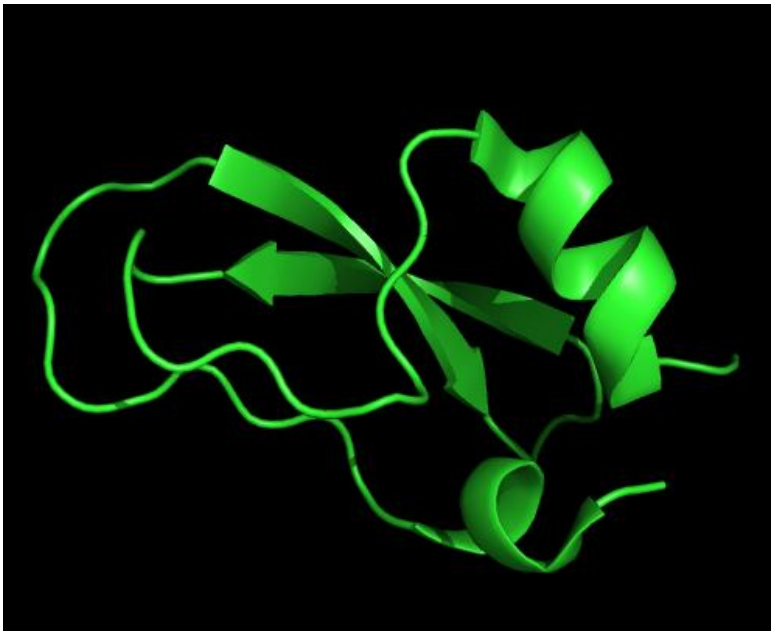


Figure 10: Crystal structure of BPTI (pdb id: 5PTI.pdb) with two 2 ß-sheets, a α-helix and a 3_{10} helical turn. The figure is drawn using Pymol (LLC Schrodinger)

The results from this MD simulation suggested that the interior of BPTI is fluid like and that the atomic motions propagated throughout the molecule resembles diffusion effects. Since then, protein unfolding via molecular dynamics simulation has been studied in vacuo/water [**169**] or with chaotropic denaturants such as urea [**170-172**] , guanidium chloride [**173**] and other solvents [**174**] to determine and characterize the unfolding pathway(s), transition states, intermediates and mapping the interactions within the protein molecule [**175-178**]. The majority of the unfolding simulations of proteins are in the presence of water at abnormally elevated temperatures (from 400-600K) to hasten the unfolding process and save computer time, under the assumption that the elevated temperature does not significantly distort the unfolding pathway(s). Few such examples include, MD simulations of 'barnase' [**179**] in which the final unfolded state(s) were a non-random coil structure. Further, MD simulations revealed the existence of "putative intermediates" in its unfolding pathway [**180**] whose structure comprised of 3-hydrophobic cores and an α-helix consistent with earlier experimental findings on this system [**181**]. In another study, unfolding MD simulations at relatively lower temperatures of 375K for two SH3 domains of α-spectrin and src-SH3 [**182**] with sequence identity of only 34% but with similar native topology seemed to follow the similar unfolding pathway(s). The above result together with earlier established kinetic data [**182**] indicated that for small ß-sheet proteins, the native topology determines the folding mechanism; on the other hand precise interactions guide the statistically significant folding path. Unfolding MD simulations at high temperatures on the thermally stable α-lytic protease was performed in order to reconcile previous experimental data and to elucidate a complete folding mechanism [**157**]. Experimentally the protein exhibited a high degree of cooperative unfolding with negligible partially unfolded states. A domain interface was identified to be integral for its cooperative unfolding. MD simulations of

α-lytic protease along with its thermodynamically stable homolog trypsin demonstrated their respective differences in unfolding. While trypsin unfolded progressively, α-lytic protease unfolded cooperatively with a characteristic disruption of the domain interface.

1.3.3 The Transition State during Folding/Unfolding

The transition state of a protein folding/unfolding differs from an ordinary chemical reaction in many ways [**183**]. In the first instance, low energy non-bonded interactions are broken in the transition state, instead of covalent bonds as in a chemical reaction. On the other hand, there may be many energy barriers along a single unfolding pathway. Usually the steepest of the energy barriers is the rate-limiting step and is assigned the status of the transition state of the overall reaction [**184**]. A crucial step in analyzing unfolding simulations is accurately elucidating the transition state ensemble (TSE) from the large number of generated conformations. Since, the TSE is experimentally accessible through a molecule's folding and unfolding kinetics, its computational identification can be used for both explanatory and predictive purposes. Various methods for identifying the TSE have been used in the past, partitioned into conformational clustering and landscape methods [**185-191**]. Conformational clustering relies on all-versus-all comparisons of conformations, often by Cα RMSD, whereas landscape methods separate native structures from the unfolded ones by utilizing properties of the conformations, such as the fraction of native contacts or secondary structure. In this regard, two pioneering works employing unfolding molecular dynamics to characterize the transition state(s) of a protein include, the characterization of transition states of chymotrypsin inhibitor-2 (CI2) [**185**] and FKBP12 [**192**]. In the first case, based on "conformational analysis" the transition state of CI2 was identified to be closer to the native state as compared to the unfolded state with a relatively

loosely packed hydrophobic core (considered to be the rate limiting step in this case), extensive disruption in the native tertiary and secondary structures with concurrent increase in the solvent accessible surface area of the molecule. On the other hand, unfolding of FKBP12 was extensively dissected in order ascertain its transition state of folding/unfolding by protein engineering (mutational analysis), kinetic studies and lastly by unfolding molecular dynamics simulation. The major transition state in consensus with experiments and simulations was identified to have an overall increase of 30% in the native solvent accessible surface area. Native tertiary and secondary interactions were significantly weaker in the transition state and the deeply buried residues in the core, for example Val 101 retained approximately 60% of the native contacts.

1.3.4 Protein Contact Networks

Given the cooperative nature of the unfolding process, viewing protein interiors as a network of interactions has provided additional insights into the phenomenon. Several attempts have been made to view protein structures as contact networks [**193-201**] wherein the amino acid residues have been designated as nodes and their mutual non-covalent interactions as edges. Besides providing insights into protein structures, these networks have been used to identify residues implicated in folding nuclei **[200]** and transition states **[194]**, identifying functional residues involved in the active site **[196]**, hubs stabilizing the packing of secondary structural elements**[198]**,rationalization of the difference in protein stabilities between thermophilic and mesophilic organisms **[198]** and estimation of folding rates **[193,197].**

Although the transition state ensemble of two proteins, chymotrypsin inhibitor -2 and C-Src SH3 domain have been subjected to extensive experimentation, there still existed

some ambiguity with regard to the transition state ensemble of these proteins. Conventional methods utilizing Cα RMSD or the radius of gyration failed to differentiate between the pre and the post-transition conformations [**202**]. Investigations, utilizing the small-world properties of the protein contact networks was successful in determining the transition state ensemble for folding of these enzymes [**202**]. Pre and post-transitional state ensembles during folding were distinguished based on the topological difference between the conformations of these states. The post-transitional states were found to be more small-world like with high local connections along with a number of long range interactions than the pre-transitional conformations.

In another study, specific hydrophobic core residues of acyl-phosphatase (AcP) crucial for the structure of the transition state during folding was identified [**194**]. Residues Y11, P54 and F94 are part of the hydrophobic core of AcP which make a large number of contacts with other residues (10, 7 and 14 contacts, respectively). These contacts, create a `contact network' in the native state. Monte-Carlo sampling technique was used to determine the transition state ensemble of the protein, which primarily conserved the native state topology with root mean square fluctuations within 6 Å from the native structure. The structure of the transition state essentially was composed of 20 amino acid residues with small spatial alterations with the contact network of these three key residues being conserved, which in turn supported the nucleation model of protein folding where a small number of specific key residues can drive a polypeptide chain to fold to a unique structure.

1.3.5 Surface Contact Networks

In chapter 3 of this thesis, the concept behind point atom contact networks has been extended to surfaces. Briefly, the entire polypeptide chain of the molecule is considered to be one continuous

surface rather than covalently linked discrete atoms. In such a representation the surface fit or surface complementarity (S_m) of two interacting residues can be estimated along with their respective surface patches (overlap, O_v) **[203]** buried upon association with each other. Intra – molecular close packing has been represented as a network with each node representing an amino acid side chain and their mutual association as links **[204-207].** As association between residues was defined in terms of (based on the surface representation of the protein) S_m (≥ 0.40) and Ov ($\geq$0.08), discussed in detail in Chapter3, the resulting graphs were expected to consist primarily of those interactions which strongly condition inter-residue geometry, rather resembling a three dimensional jigsaw puzzle [**203**].

Earlier studies utilizing such 'surface contact networks' based on the criteria of surface complementarity (S_m) and overlap (O_v) **[204]** to study the packing of side chains in the interior of the proteins, showed the preferred occurrence of specific "network topologies" which were designated as packing motifs. Study of the distribution of these packing motifs revealed the presence of typical smaller graphs which often condense to form larger networks, much like the nucleation-condensation model of protein folding. One of these packing motifs is the closely associated, mutually interacting three residue cluster which forms a " clique". These cliques were found typically in the regions of dense packing in the core and were thus envisaged as units of clustering. Cliques were distinctly preferred in smaller networks that formed the 'basis sets' in the construction of larger networks. Other topological analyses and measures based on "surface contact network" also proved to be fruitful in distinguishing the native fold from a set of decoys.

There are few reports of molecular dynamics simulation on cyclophilin [**152,154**] but all of them focus primarily on the structural basis of its catalytic activity. In chapter 3, unfolding

molecular dynamics simulation of cyclophilin induced by temperature has been studied extensively. Surface contact based novel metrics (Disnet) has also been designed to gain insights into the patterns of unfolding in cyclophilin. Attempts were also made to identify the transition state ensemble and the wet or dry characteristics of the molten globule (if any) during the unfolding of the molecule. Multidimensional scaling with metric Disnet clearly indicated the presence of transition states at 450K simulation temperature. The metric Disnet (other than the conventional metrics of Cα RMSD, radius of gyration etc.) was extremely sensitive to fluctuations in the surface contact network of the protein and gave unambiguous signals as to the perturbations of the tertiary interactions within the protein core.

1.4 Characterization of the Thermal and Conformational Stability of Cyclophilin Mutants

1.4.1 Introduction

Thermally stable proteins are of interest for several reasons. They can be used to improve the efficiency of many industrial processes and provide insight into the general mechanisms of protein folding and stabilization. Despite arduous efforts through protein engineering methods, theoretical calculations and simulations, a detailed and comprehensive understanding of protein stability is still desirable. A major bottleneck in this regard is the small difference in free energy between the folded and unfolded states of a protein and the large number of partly antagonistic factors, which are difficult to assess on an energetic scale. Most thermostable proteins have close mesostable relatives that conserve the fold and display significant levels sequence similarity.

1.4.2 Factors Contributing to the Thermal Stability of Proteins

Thermostable proteins are characteristically functional at high temperatures in excess of 40° C. Generally, high degrees of thermostability are found in the class of metalloproteinases. Thermolysin catalyzing the hydrolysis of peptide bonds, is the most stable member of the metalloproteinase family from *Bacillus thermoproteolyticus* with a T_m of 86.9°C. Historically, crystal structure of thermolysin solved in 1974 **[208]** formed a basis for the structural explanation of the thermostability of the enzyme. Since then several researchers have attempted to explain thermal stability of proteins at a molecular level. The major causes of thermostability of proteins were attributed to increased hydrophobicity [**209**], superior packing, deletion or shortening of the loop regions [**210**], smaller and lesser number of cavities, increased solvent accessible surface area getting buried as a result of oligomerization [**211**], amino acid replacement within and outside secondary structures [**212,208,213**] and greater propensity of proline residues [**208,214,215**]. Other reasons could include lesser number of thermally unstable residues [**210**], increased helical content, increased polar surface area [**209**,**216,217**], increased hydrogen bonding [**216,217**], higher number salt bridges [**218,219,209,210,213,220,221,222**] and enhanced side chain interactions between buried amino acid residues [**223,224**]

To explore the role of the extent of hydrophobic effect in proteins sequence comparisons of highly related archaeal adenylate kinases (AKs) from the mesophilic *Methanococcus voltae,* the moderate thermophile *Methanococcus thermolithotrophicus,* and two extreme thermophiles *Methanococcus igneus* and *Methanococcus jannaschii,* led to the identification of atomic interactions responsible for the large variation in temperatures for optimal catalytic activity and thermostabilities observed in these proteins [**209**]. The structure of these proteins were modeled

by first threading the sequences against high resolution crystal structures of eukaryotic , eubacterial and mitochondrial AK's based on a energy-based sequence structure threading procedure. The final model structures revealed possible active site interactions with a higher propensity of proline residues near the active site. Based on this study, the enhanced thermostability in the thermophiles were primarily attributed to relatively larger cores with a corresponding increase in the number of aliphatic amino acids.

In another study [**216**], 16 families of proteins with different thermal stability were theoretically examined by comparing their respective fractional polar atom surface areas and the number and type of hydrogen bonds and salt bridges between protein atoms. It was found that in over 80% of the proteins belonging to the same family there existed a strong correlation between thermostability of these proteins and an increase in the number of hydrogen bonds as well as an increase in the fractional polar surface area resulting in added hydrogen bonding density to water. Thus, increased hydrogen bonding density could also be considered as one of the contributing factors. In the same study, the number of ion pairs were also found to be positively correlated with thermal stability in two-thirds of the tested families, though relative to the density of hydrogen bonds, the contribution of salt bridges was considered to be less.

1.4.3 Protein Contact Networks and the Thermal Stability of Proteins

Representation of the interactions between the constituent amino acids in a protein in terms of network representations has also yielded important insights into the stability of protein structures. One such example in this regard is the identification of aromatic clusters of amino acid residues [**223**] that plays an integral role in determining the stability of thermophilic proteins. In this study, the contact network of 24 protein families were studied for which crystal

structures of the thermophilic proteins along with their mesophilic homologs were present. Subsequent analysis of the contact networks of these proteins demonstrated the presence of additional aromatic clusters for 17 thermophilic protein families which were absent in their mesophilic counterparts. These additional aromatic clusters were relatively smaller in size, found on the surface of the proteins and often in close proximity of the active site of these thermophilic enzymes. A T-shaped orthogonal geometry in packing seemed to be favored in these additional clusters. The aromatic residues belonging to these additional clusters were found to be mutated to serine, isoleucine and leucine in their mesophilic homologues.

In another seminal work [**224**] protein contact networks based on the weighted side chain interactions were constructed for globular proteins from a dataset of 200 crystal structures. A salient feature of such contact networks lay in the definition of edges based on the normalized strength of interaction between the amino acid residues in proteins and was referred to as PSG's (Protein Structure Graphs). The PSG's were constructed based on different values of interaction strength. Interestingly, it was found that the network topology of such PSGs relied on the threshold values of the interaction strength between amino acid residues used to define the edges in the graph. For a narrow range of threshold values (cutoffs on the interaction strength) there seemed to be a transition in the largest cluster (larger clusters or interaction giving way to smaller graphs) of the protein contact network, irrespective of the fold or size of the protein, thus signifying a generality in the transition for globular proteins. The method was also successful in the identification of "highly connected" amino acid residues playing a key role in maintaining the integrity of the "protein structure network". These highly inter-connected amino acid residues are designated as hubs in network terminology. It was also observed that the aromatic

residues along with arginine, histidine, and methionine act as strong hubs at high interaction cutoffs, while hydrophobic residues leucine and isoleucine participate in these hubs at low interaction cutoffs, forming weak hubs. Protein networks are less susceptible to random attacks on nodes but are prone to overall disruption following specific attacks on the hubs [**225**], where mutations of these residues participating in such hubs leads to the destabilization of the protein.

Analysis of the native simulation of cyclophilin (LdCyp) in terms of "surface contact network" (discussed in detail in Chapter3, Materials and Methods) led to the identification of highly "persistent" interactions between amino acid residues , which in some cases involved residues which were implicated in several strategic interactions, thus qualifying as 'hubs'. In order to study, the extent of destabilization caused as a result of disrupting these hubs, a set of such amino acid residues in the hydrophobic core, away from the active site of the molecule were mutated to alanine and their corresponding destabilizing effect studied (in Chapter 4) as a function of temperature and chaotropic denaturant guanidinium chloride (GdmCl).

References

1. N. Takahashi, T. Hayano, M. Suzuki (1989). *Nature*, 337, 473-475.
2. G. Fischer, B. Wittmann-Liebold, K. Lang, T. Kiefhaber, F. X. Schmid, (1989). *Nature*, 337, 476-478.
3. E. R. Schoenbrunner, S. Mayer, M. Tropschug, G. Fischer, N. Takahashi, F. X. Schmid, (1991). *J. Biol. Chem*. 266, 3630-3635.
4. A. Rinfret, C. Collins, R. Menard, S.K. Anderson (1994). *Biochemistry*, 33, 1668-1673.
5. G.Kern, D. Kern, F. X. Schmid, C. Fischer (1994). *FEBS Lett*. 348, 1145-1148.
6. A. Galat (2003) *Curr Top Med Chem*, 3:1315-1347.

7. P.C. Waldmeier, K. Zimmerman, T. Qian, M. Tintelnot-Blomley, J.J Lemasters (2003) *Curr Med Chem*, 10:1485-1506.

8. Z. He, L. Li, S. Luan (2004) *Plant Physiol,* 134**:**1248-1267.

9. M. Arevalo-Rodriguez, X. Wu, S.D. Hanes, J. Heitman (2004) *Front Biosci*, 9**:**2420-2446.

10. B.E. Bierer, P.K. Somers, T.J. Wandless, S.J. Burakoff, S.L. Schreiber,(1990). *Science*, 250, 556-559.

11. H.M. Ke, L.D. Zydowsky, J. Liu, C.T. Walsh,(1991). *Proc. Natl. Acad. Sci. U. S. A*. 88, 9483-9487.

12. J. Kallen, C. Spitzfaden, M. Zurini, G. Wider, H. Widmer, K. Wuthrich, M.D. Walkinshaw (1991). *Nature (London)*, 353, 276–279.

13. J. Dornan, P. Taylor, M.D. Walkinshaw (2003) *Curr Top Med Che* , 3**:**1392-1409.

14. L. Andreeva , R. Heads, C.J. Green (1999), *Int J Exp Pathol* 80:305-315.

15. G.S. Hamilton, J.P. Steiner JP (1998) *J Med Chem*, 41:5119-5143.

16. H. Mi, O. Kops, E. Zimmermann, A. Jaschke, M. Tropschug (1996) *FEBS Lett*, 398:201-205.

17. L.J. Kieffer, T.W. Seng, W. Li, D.G. Osterman, R.E. Handschumacher, R.M. Bayney (1993) *J Biol Chem*, 268:12303-12310.

18. S.K. Anderson, S. Gallinger, J. Roder, J.Frey, H.A. Young, J.R. Ortaldo. (1993) *Proc Natl Acad Sci USA*, 90:542-546.

19. A. Rinfret, C. Collins, R. Menard, S.K. Anderson. (1994) *Biochemistry*, 33:1668-1673.

20. B. Haendler, R. Keller, P.C. Hiestand, H.P. Kocher, G. Wegmann, N.R. Movva (1989):. *Gene*, 83:39-46.

21. M. Arevalo-Rodriguez, J. Heitman (2005) *Eukaryot Cell*, 4:17-29.

22. P.L. Koser, D. Sylvester, G.P. Livi, D.J. Bergsma (1990) *Nucleic Acids Res*, 18:1643.

23. P.L. Koser, D.J. Bergsma, R. Cafferkey, W.K. Eng, M.M. McLaughlin, A. Ferrara, C. Silverman, K. Kasyan, M.J. Bossard, R.K. Johnson.(1991) *Gene*, 108:73-80.

24. M.M McLaughlin, M.J. Bossard, P.L. Koser, R. Cafferkey, R.A. Morris, L.M. Miles, J. Strickler , D.J. Bergsma, M.A. Levy, G.P. Livi. (1992)*Gene*, 111:85-92.

25. G. Frigerio, H.R.Pelham. (1993) *J Mol Biol*, 233:183-188.

26. W.K. Huh, J.V. Falvo, L.C. Gerke, A.S. Carroll, R.W. Howson, J.S. Weissman , E.K. O'Shea (2003) *Nature*, 425:686-691.

27. A. A Duina, J.A. Marsh, R.F. Gaber. (1996) *Yeast*, 12:943-952.

28. J. Luban, K.L. Bossolt, E.K. Franke, G.V. Kalpana S.P. Goff, (1993) *Cell* 73, 1067-1078.

29. N. Raghavan, I. Ghosh, W.S. Eisinger, D. Pastrana, A.L. Scott (1999) *Mol. Biochem. Parasitol.* 104, 233-246.

30. M.F. Wiser (2003) *Parasitol. Res*. 90, 166-170.

31. G.R. Reddy, (1995) *Mol. Biochem. Parasitol*. 73, 111-121.

32. D.L. Looker, R.L. Berens, J.J. Marr, (1983) *Mol. Biochem. Parasitol.* 9, 15-28.

33. A. Chakraborty, I. Das, R. Datta, B. Sen, D. Bhattacharyya, C. Mandal, A.K. Datta, (2002) *J. Biol. Chem*. 277, 47451-47460.

34. J.R. Glover, S.Lindquist, (1998) *Cell* 94, 73-82.

35. A.L. Fink, (1999), *Physiol. Rev*. 79, 425-449.

36. M. Thali A. Bukovsky , E. Kondo,B. Rosenwirth, C.T. Walsh, J. Sodroski, H.G. Göttlinger . (1994)*Nature*. 24(372)363-365.

37. H. Du, S.S Yan (2010) *Biochimica et Biophysica Acta*, 1802, 198-204.

38. A. Chakraborty, B. Sen, R. Datta, A.K. Datta (2004) *Biochemistry*, *43,* 11862-11872.

39. T. Hornbogen, R. Pieper, K. Hoffmann, H. Kleinkauf, R. Zocher (1992): *Biochem Biophys Res Commun*, 187:791-796.

40. J. Liu, J.D. Farmer Jr, W.S. Lane, J. Friedman, I. Weissman, S.L. Schreiber (1991) *Cell*, 66:807-815.

41. F. Foor, S.A.Parent, N. Morin, A.M. Dahl, N. Ramadan, G. Chrebet, K.A. Bostian, J.B. Nielsen (1992) *Nature* , 360:682-684.

42. J. Liu, M.W. Albers, T.J. Wandless, S. Luan, D.G. Alberg, P.J. Belshaw, P. Cohen, C. MacKintosh, C.B. Klee, S.L. Schreiber(1992) *Biochemistry*, 31: 3896-3901.

43. L. Jin, S.C. Harrison (2002) *Proc Natl Acad Sci USA*, 99:13522-13526.

.44. Q. Huai, H.Y. Kim, Y. Liu, Y. Zhao, A. Mondragon, J.O. Liu, H. Ke (2002) *Proc Natl Acad Sci USA*, 99:12037-12042.

45. J. Luban, K.L. Bossolt, E. K. Franke, G.V. Kalpana, S.P. Goff(1993) *Cell*, 73:1067-1078.

46. J. Luban (1996) *Cell*, 87:1157-1159.

47. D. A. Bosco, E.Z. Eisenmesser, S. Pochapsky, W.I. Sundquist, D. Kern (2002). *Proc Natl Acad Sci USA*, 99:5247-5252.

48. D. Braaten, C. Aberham, E.K. Franke, L. Yin, W. Phares, J. Luban: (1996) *J Virol*, 70:5170-5176.

49. D. Braaten, E.K. Franke, J. Luban(1996) *J Virol* , 70 , 3551-3560.

50. D. Braaten, J. Luban, (2001) *EMBO J*, 20:1300-1309.

51. E. Sokolskaja, D.M. Sayah, J. Luban , (2004) *J Virol*, 78:12800-12808.

52. K.Zander, M.P. Sherman, U. Tessmer, K. Bruns, V. Wray, A.T. Prechtel (2003). *J Biol Chem* 2003; 278: 43202–43213.

53. M. Qi, C. Aiken. (2008) *Virology*; 373: 287–297.

54. T. Ratajczak, A. Carrello, P.J. Mark, B.J. Warner, R.J. Simpson, R.L. Moritz, A.K. House,(1993) *J Biol Chem* 1993, 268:13187-13192.

55. H.C. Chang, S. Lindquist: (1994) *J Biol Chem*, 269:24983-24988.

56. M. Arevalo-Rodriguez, X. Wu, S.D. Hanes, J. Heitman(2004) *Front Biosci*, 9:2420-2446.

57. J.D. Leverson, S.A. Ness (1998) *Mol Cell*, 1:203-211.

58. W.M. Yang, C.J. Inouye, E. Seto. (1995), *J.Biol.Chem*. 270:15187-15193.

59. Ansari H, Greco G, Luban J., *Mol Cell Biol* 2002, 22:6993-7003.

.60. Rycyzyn MA, Clevenger CV., *Proc Natl Acad Sci USA* 2002, 99:6790-6795.

61. G. Petrosillo, F.M. Ruggiero, M. Pistolese, G.Paradies. (2004) *J Biol Chem*. 2004;

279:53103–53108.

62. T.R. Rosenstock, A.C. Carvalho, A. Jurkiewicz, R. Frussa-Filho, S.S. Smaili (2008) *J*

Neurochem. 88:1220–1228.

63. M.F. Galindo, J. Jordan, C. Gonzalez-Garcia, V. Cena(2003) *Br. J Pharmacol*.139:797–804.

64. C.P. Baines, R.A. Kaiser, N.H. Purcell, N.S. Blair, H. Osinska, M.A. Hambleton, E.W.

Brunskill, M.R. Sayen, R.A. Gottlieb, G.W. Dorn, J. Robbins, J.D. Molkentin (2005)

Nature.434:658–662.

65. A.P. Halestrap, C.P. Connern, E.J. Griffiths, P.M. Kerr(1997) *Mol Cell Biochem*,174:167–172.

66. L. Khaspekov, H. Friberg, A. Halestrap, I. Viktorov, T. Wieloch (1999) *Eur J Neurosci*,11:3194–3198.

67. A. Chakraborty, I.Das , R. Datta,B.Sen,D. Bhattacharyya,C. Mandal, Chhabinath , A.K. Datta (2002) *J.Biol.Chem* 277 (49) 47451-47460.

68. A. Chakraborty, B. Sen, R. Datta, A.K. Datta (2004) *Biochemistry*, *43,* 11862-11872

69. H. Ke, M. Zydowsky, L. D. Liu, C.T. Walsh,(1991). *Proc. Natl. Acad. Sci. U.S.A*. 88, 9483-9487.

70. J. Kallen, C.Spitzfaden, M. Zurini, G.Wider, H. Widmer, K. Wuthrich, M.D. Walkinshaw, (1991). *Nature (London)*, 353, 276–279.

71. H. Ke, (1992). *J. Mol. Biol*. 228, 539-550.

72. J. Fejzo, F.A. Etzkorn, R.T. Clubb, Y. Shi, C.T. Walsh, G. Wagner, G. (1994). *Biochemistry*, 33, 5711-5720.

73 .J. Kallen, V. Mikol, P. Taylor, M.D. Walkinshaw (1998). *J. Mol. Biol*. 283, 435-449.

74. P.J. Ellis, C.K. Carlow, D. Ma, P. Kuhn (2000). *Biochemistry*, 39, 592-598.

75.M.R. Peterson, D.R. Hall, M. Berriman, J.A. Nunes, G.A. Leonard, A.H. Fairlamb, W.N. Hunter, (2000). *J. Mol. Biol*. 298, 123-133.

76. V. Venugopal, B. Sen, A.K. Datta, R. Banerjee (2007). *Acta Crystallogr. F*, 63, 60-64.

77. V. Mikol, J. Kallen , M.D. Walkinshaw (1994) *Proc Natl Acad Sci USA* , 91**:**5183-5186.

78. G. Frigerio, H.R. Pelham (1993) *J Mol Biol*, 233:183-188.

79. J. Liu, J.D.Farmer Jr, W.S. Lane, J. Friedman, I. Weissman, S.L. Schreiber (1991) *Cell*, 66:807-815.

80. A. Billich, F.H. Schmid,P. Peichl, R. Wenger, G. Zenke, V. Quesniaux, B. Rosenwirth (1995) , *J.Virol* 69, 2451-2461.

81. J.Bua, A.M. Ruiz, M. Potenza, L.E. Fichera, (2004). *Bioorg. Med. Chem. Lett*. 14, 4633-4637.

82. J. Bua, L.E. Fichera, A.G. Fuchs, M. Potenza, M, M. Dubin, R.O. Wenger, G. Moretti, C.M. Scabone, A.M. Ruiz (2008). *Parasitology*, 135, 217-228.

83. M.E. Perkins, T.W. Wu, S.M. LeBlancq, (1998). *Antimicrob. Agents. Chemother*. 42, 843-848.

84. O.I. Povarova, I. M. Kuznetsova, K.K Turoverov , *Plos One* , 5(11).

85. B.Yélamos, E. Nunez , J Gomez-Gutierrez , C Delgado , B Pacheco , D L Peterson , F Gavilanes. *Biochim. Biophys Acta*. (2001) 1546(1):87-97.

86. B. I Molero, J.M. Sanchez-Ruiz, *Biochemistry*, (1997), 36, 9616-9624

87. R.F. Latypov, H Cheng, N.A. Roder, J.Zhang, H.Roder (2006) *J. Mol. Biol*. 357, 1009–1025

88. F. Khan, A. Ahmad, M. I. Khan (2007) IUBMB Life, 59(1), 34 – 43.

89. M.M. Patel, F. Tzul, G. I. Makhatadze, (2011) *Biophys. Chem*. 159, 58-65.

90. H. A Sathish , M. Cusan, C. Aisenbrey, B. Bechinger (2002) 41, 5340-5347.

91. O.D. Monera, C.M. Kay, R.S. Hodges (1994) *Prot. Sci*., 3:1984-1991.

92. Z. Lai, J.McCulloch, H.A. Lashuel, J.W. Kelly (1997) *Biochemistry*, 36, 10230-10239.

93. C.N. Pace, B.A Shirley, J.A. Thomson, Measuring the conformational stability of a protein. in: T.E. Creighton (Ed.), Protein Structure: A Practical Approach, IRL Press at Oxford University Press, Oxford, New York, Tokio, (1989), 311–330

94. T.O. Street, N. Courtemanche, D. Barrick, (2008) Protein Folding and Stability Using Denaturants in Methods in Cell Biology, vol. 84, *Elsevier Inc*., Chapter 11

95. B.S Yang (2011). The Stability of a Three-State Unfolding Protein, Thermodynamics – Physical Chemistry of Aqueous Systems, Juan Carlos Moreno-Piraja´n (Ed.), ISBN: 978-953

96. M.P. Irún, M.M. Garcia-Mira, J.M. Sanchez-Ruiz, J. Sancho, (2006) *J. Mol. Biol*. 306 877–888.

97. J. Sancho, M. Bueno, L.A. Campos, J. Fernández-Recio, M.P. Irún, J. López, C. Machicado, I. Pedroso, M. Toja, (2002) *Sci. World J*. 2, 1209–1215.

98. L.A. Campos, M.M. Garcia-Mira, R. Godoy-Ruiz, J.M. Sanchez-Ruiz, J. Sancho (2004) *J. Mol. Biol*. 344, 223–237.

99. D.Barrick, R.L. Baldwin (1993) *Biochemistry* 32, 3790-3796.

100. E.N. Baryshnikova, B.S. Melnik, A.V. Finkelstein, G.V.Semisotnov,V. E. Bychkova (2005) Prot. Sci. 14:2658–2667

101. Y. Goto, Y. Hagihara, D. Hamada, M. Hoshino, I.Nishii (1993) *Biochemistry, 32,* 11878-1 1885

102. A. Dasgupta, J. B. Udgaonkar(2012) *Biochemistry* doi:http://dx.doi.org/10.1021/bi300223b.

103. A.H. Wani, J.B. Udgaonkar. *J. Mol. Biol*. 387, (2009), 348–362.

104. A.K. Mandal, S. Samaddar, R. Banerjee, S. Lahiri, A. Bhattacharyya, S. Roy, *J. Biol. Chem.* 278 , (2003), 36077-84.

105. R.F Latypov, H. Cheng , N. A. Roder , J. Zhang , H. Roder*, J. Mol. Biol*. 357, (2006), 1009-1025.

106. K. Kuwaiima, K. Nitta, M. Yonevama, S. Suzai. (1976) *J. Mol. Biol*. 106:359-373.

107.K. Kuwaiima (1977) *J. Mol. Biol*. 114:241-258.

108. M. Ikeguchi, K. Kuwajima, S. Sugai (1986) *J. Biochem*. (Tokyo) 99:1191-1201

109. M. Ikeguchi, K. Kuwajima, M. Mitani, S. Sugai (1986) *Biochemistry* 25:6965-6972

110. K. Kuwajima, M. Mitani, S. Sugai. (1989) *J. Mol. Biol*. 206:547-561

111. D.A. Dolgikh, R.I. Gilmanshin, E.V. Brazhnikov, V.E. Bychkova, G.V. Semisotnov, S. Venyaminov, S. Yu, O.B. Ptitsyn (1981) *FEBS Lett*. 136:311-315

112. D.A. Dolgikh, A.P. Kolomiets, I.A. Bolotina, O.B. Ptitsyn (1984) *FEBS Lett*. 165:88-92

113. G.V. Semisotnov, N.A. Rodionova, V.P. Kutyshenko, B. Ebert, J. Blanck, O.B. Ptitsvn (1987*) FEBS Lett.*224:9-13

114. O.B. Ptitsyn (1987) *J. Protein Chem*. 81: 6972-6993

115. K.P. Wong, C. Tanford (1973) *J. Biol. Chem*. 248:8518-8523.

116. L.A. Holladay, R.G. Hammonds Jr, D. Puett (1974) *Biochemistry* 13:1653-1661

117. D.N. Brems, S.M. Plaisted, H.A. Havel, E.W. Kauffman, J.D. Stodola, L.C. Eaton, R.D. White (1985) *Biochemistry* 24:7662-7668, 1985.

118. K.P. Wong, L.M. Hamlin (1974) *Biochemistry* 13:2678-2683.

119. E.V. Brazhnikov, V. Chirgadze, N. Yu, D.A**.** Dolgikh, O.B. Ptitsyn (1985) *Biopolymers* 24: 1899-1907

.

120. M.V. Jagannadham, D. Balasubramanian (1985) *FEBS Lett.* 188:326-330.

121. Y. Goto, A.L. Fink (1989*) Biochemistry* 28:945-952

122**.** P.Malhotra, J.B. Udgaonkar (2014) *Biochemistry* 53 (22):3608-20.

123. S.S. Sarkar, J.B. Udgaonkar, G. Krishnamoorthy(2013). *Biophys J* 105(2392-2402).

124. B.R. Rami, J.B. Udgaonkar (2002) *Biochemistry* 41, 1710-1716.

125. R.L. Baldwin, C. Frieden, G.D. Rose (2013) *Proteins: Struct. Func. Bion*. 78: 2725–2737.

126. T. Kiefhaber, R.L. Baldwin (1995) *Proc Natl Acad Sci USA*; 92:2657–2661.

127. T.R. Sosnick, D. Barrick, *Curr. Opin. Struct. Biol.* 21 (2011) 12–24.

128. I. Bahar, T.R. Lezon, L.W. Yang, E. Eyal, *Annual Review of Biophysics* 39 (2010) 23–42
.
129. V.E. Angarica, J. Sancho, Protein dynamics governed by interfaces of high polarity and low packing density. Plos One, in press, http://dx.doi.org/10.1371/ journal.pone.0048212

130. A. Pastore, P.A. Temussi, *Curr. Opin. Struct. Biol*. 22 (2012) 30–37.

131. H. Hong, N.H. Joh, J.U. Bowie, L.K. Tamm, *Methods Enzymol*. 455 (2009) 213–236.
132. J.B. Rouget, T. Aksel, J. Roche, J.L. Saldana, A.E. Garcia, D. Barrick, C.A. Royer, J. *Am. Chem. Soc*. 133 (2011) 6020–6027.

133. D.K. West, D.J. Brockwell, P.D. Olmsted, S.E. Radford, E. Paci, Biophys. J. 90 (2006) 287–297.

134. J. Liphardt, S. Dumont, S.B. Smith, I. Tinoco Jr., C. Bustamante, Science 296 (2002) 1832–1835.

135. D. Mitra, S. Mukherjee, A.K. Das (2006) FEBS Lett.580, 6846-6860.

136. B. Jana, A. Bandhu, R. Mondal, A. Biswas, K. Sau, S. Sau (2012) *Biochemistry*, 51(6):1223-37.

137. O.B. Ptitsyn, A.A. Rashin (1975). *Biophys. Chem*. 3, 1-20.

138. P.S. Kim, R.L. Baldwin (1982) *Annu. Rev. Biochem*. 51, 459-489.

139. P.S. Kim, R.L. Baldwin (1990) *Annu. Rev. Biochem*.59, 631-660.

140. M. Karplus, D.L. Weaver (1976). *Nature,* 260, 404-406.

141. M. Karplus, D.L. Weaver (1994). *Prot. Sci*. 3, 650-668.

142. O.B. Ptitsyn (1994). *Protein Eng*. 7, 593-596.

143. O.B. Ptitsyn (1987). *J. Protein Chem*. 6, 273-293.

144. R.L. Baldwin (1989) *Trends. Biochem. Sci*. 14, 291-294

145. K.A. Dill (1990) *Biochemistry*, 29, 7133-7155.

146. S.C. Harrison, R. Durbin (1985). *Proc. Natl Acad. Sci. USA*, 82, 4028-4030

147. M. Levitt (1983) *J. Mol. Biol*. 168, 595-620

148. W.F van Gunsteren, H.J Berendsen, J.Hermans, W.G Hol, J.P. Postma (1983) *Proc. Natl Acad. Sci. USA*, 1983,80(14), 4315–4319.

149. O.V. Galzitskaya, J.Higo, A.V.Finkelstein (2002) *Curr Protein Pept. Sci*. 3, 191–200

150. J. Gsponer, A. Caflisch (2002) *Proc. Natl Acad. Sci. USA*, 99, 6719-6724.

151. P.L. Freddolino, F.Liu, M.Gruebele, K. Schulten (2008), 94,doi:

http://dx.doi.org/10.1529/biophysj.108.131565

152. D. Trezesniak, W. F. Van Gunsteren, (2006), *Prot Sci* 15:2544–2551

153. M.Orozco. J. Tirado-Rives, W. L. Jorgensen *Biochemistry* (1993), 32, 12864-12874

154. P.K. Agarwal, A. Geist, A. Gorin (2004) *Biochemistry,* 43(33) 10605-1061.

155. M. Nagaraju, L .C. McGowan, D. Hamelberg (2013) J. Chem. Inf. Model., 53, 403−410

156. B.L. de Groot, S. Hayward, D.M.F van Aalten, A. Amadei, H.J.C. Berendsen (1998) Proteins: Structure, Function, and Genetics 31:116–127.

157. N.L Salimi, B. Ho, D.A Agard,(2010) Plos Comp. Biol. doi: 10.1371/journal.pcbi.1000689

158. I. Tsigelny,J. P. Greenberg,S. Cox, W .L. Nichols, S.S. Taylor, L. F. Ten Eyck (1999) *Biopolymers* , 50, 513–524

159. L.S. Caves, J. D. Evanseck, M. Karplus (1998) *Protein Sci*. 7(3): 649–666.

160. D.L. Ensign, P.M. Kasson, V.S. Pande, (2007) *J Mol. Biol*. 374, 806–816.

161. A. Pérez, F.J. Luque, M. Orozco, (2007) *J Am Chem Soc*, 129, 14739–14745.

162. M.M. Seibert, A. Patriksson, B. Hess, D. van der Spoel (2005) *J Mol Biol*, 354,173–183

163. P. Maragakis, K. Lindorff-Larsen, M.P. Eastwood, R.O. Dror, J.L. Klepeis, I.T. Arkin, M.O. Jensen, H. Xu, N. Trbovic, R.A. Friesner, (2008).*J Phys Chem B*, 112, 6155–6158.

164. K. Martinez-Mayorga, M.C. Pitman, A. Grossfield, S.E. Feller, M.F. Brown(2006) *J Am Chem Soc,* 128, 16502–16503

165. A. Grossfield, M.C. Pitman, S.E. Feller, O. Soubias, K. Gawrisch , (2008) *J Mol Biol*, 381, 478–486

166. R.O. Dror, D.H. Arlow, D.W. Borhani, Jensen MO, S. Piana, D.E. Shaw (2009) *Proc Natl Acad Sci U S A,* 106, 4689–4694.

167. R.D. Schaeffer, A. Fersht, V. Daggett, (2008) *Curr Opin Struct Biol*, 18, 4–9.

168. McCammon JA, Gelin BR, Karplus M (1977) *Nature* 267, 585-590.

169. G.A Arteca, C.T. Reimann, O. Tapia (2001) *Mass Spectrom Rev*. 20(6):402-22.

170. J.T Rives, M. Orozco, W. L. Jorgensen (1997) *Biochemistry* 36, 7313 - 7329.

171. A Caflisch, M Karplus (1999) *Structure* 7, 477-488.

172. A. Wallqvist and D. G. Covell (1998) *J. Am. Chem. Soc*. 120, 427-428.

173. C. Camilloni, A. G. Rocco, I. Eberini, E. Gianazza, R. A. Broglia, G. Tiana (2008) *Biophys J* , 94, 4654–4661.

174. C. M.Soares, V. H. Teixeira, A.M. Baptista (2003) *Biophys J*, 84(3): 1628–1641.

175. A. Li, V. Daggett, (1996) *J. Mol. Biol*. 257, 412–429.

176. A. Li, V. Daggett, (1998) *J. Mol. Biol.* 275, 677- 694.

177. A. Li, V. Daggett (1994) *Proc. Natl. Acad. Sci. USA* 91, 10430-10434

178. V. Daggett, M. Levitt (1992) *Proc. Natl. Acad. Sci*. USA 89, 5142-5146

179. C.L.Bond, K.B. Wong, J.Clarke, A.R. Fersht, V. Dagget (1997) *Proc. Natl. Acad. Sci. USA*, 94,13409-13413.

180. A. Li A, V.Daggett (1998) *J.Mol.Biol*. 275, 677-694.

181. A.R. Fersht (1993) *FEBS lett*. 325, 5-16.

182. J. Gsponer, A. Caflisch (2001) *J.Mol.Biol*. , 309, 285- 298.

183. A.R. Fersht (1995) *Curr. Opin. Struct. Biol.* 5, 79–84

184. A. Matouschek, L. Serrano, A.R. Fersht (1993) *Nature*, 340, 122–126.

185. A. Li, V. Daggett (1994) *Proc Natl Acad Sci USA* 91: 10430–10434.

186. R. Day, V. Daggett (2005) *Proc Natl Acad Sci USA* 102: 13445–13450.

187. T. Lazaridis, M. Karplus (1998) *Science* 278: 1928–1931.

188. K.A. Scott, L.G. Randles, S.J. Moran, V. Daggett, J. Clarke (2006) *J Mol Biol* 359: 159–173

189. S.L. Kazmirski, A. Li, V. Daggett (1999) *J Mol Biol* 290: 283–304.

190. E.M. Boczko, C.L. Brooks (1995) *Science* 269: 393–396

191. J.E. Shea, C.L. Brooks (2001). *Annu Rev Phys Chem* 52: 499–535

192. K.F. Fulton, E.R. Main, V. Daggett, S.E. Jackson (1999) *J Mol Biol*. 291(2):445-61.

193. K.W. Plaxco, K.T. Simons, D. Baker (1998) *J. Mol. Biol.* 277: 985-994

194. M. Vendruscolo, E. Paci, C.M. Dobson, M. Karplus (2001) *Nature.* 409: 642-645

195. L.H. Greene, V.A. Higman (2003) *J. Mol. Biol.* 334: 781-791

196. G. Amitai, A. Shemesh, E. Sitbon, M. Shklar, D. Netanely, I. Venger, S. Pietrokovski (2004) *J. Mol. Biol.*344: 1135-1146.

197. M. Punta, B. Rost (2005) *J. Mol. Biol.* 348: 507-512.

198. K.V. Brinda, S. Vishveshwara (2005) *Biophys J* 89: 4159-4170

199. M. Aftabuddin, S. Kundu (2007) *Biophys. J.*, 93: 225–231

200. J. Li, J.Wang, W. Wang W (2007) *Proteins.* 71: 1899-1907

201. G. Bagler, S. Sinha S (2007) *Bioinformatics.* 23:1760–1767

202. N.V. Dokholyan, L. Li, F. Ding, E. I. Shakhnovich (2002) *Proc Natl Acad Sci U S A*. 99(13): 8637–8641

203. R. Banerjee, M. Sen, D. Bhattacharyya, P. Saha (2003) *J.Mol.Biol* 333, 211–226.

204 S. Basu, D. Bhattacharyya, R. Banerjee (2011) *BMC Bioinformatics* (2011) 12:195.

205. L.H. Greene, V.A. Higman (2003) *J.Mol.Biol.* 334, 781-791.

206. M. Vendruscolo, N.V. Dokholyan, E. Paci, M.. Karplus (2001) *Phys Rev E*. 65 (6).

207. A.R. Atilgan, P. Akan, C. Baysal (2004) *Biophys J.* 86, 85-91.

208. B.W. Matthews, L.H. Weaver, W.H. Kester (1974) *J.Biol.Chem* , 249, 8030-8044.

209. P. Haney, J. Konisky, K.K. Koretke, Z. Luthey-Schulten, P.G. Wolynes (1997) *Proteins* 28, 117-130.

210. R.J.M Russell, J.M.C Fergusson, D.W. Haugh, M.J. Danson, G.L. Taylor (1997) *Biochemistry* 36, 9983-9994.

211. T. Salminen, A. Teplyakov, J. Kankare, B.S. Cooperman, R. Lahti, A. Goldman (1995) *Prot. Sci.* 5, 1014-1025.

212. H. Zuber (1988*) Biophys Chem*. 29, 170-179.

213. R.J.M Russell, U. Gerike, M.J. Danson, D.W. Hough, G.L. Taylor (1998) *Structure*, 6, 351-361

214. K. Watanabe, Y. Hata, H. Kizaki, Y. Katsube, Y. Suzuki (1997) *J. Mol. Biol*., 269, 142-153.

215. Y. Bogin, M. Peretz, Y. Hacham, Y. Korkhin, F. Frolow, A.J. Kalb, Y. Burstein (1998) *Prot. Sci*. 7, 1156-1163.

216. G. Vogt, P. Argos (1997) *Fold Des*. 2, S40-S46.

217. G. Vogt, S. Woell, P. Argos (1997) *J.Mol.Biol*. , 269, 631-643.

218. K.S.P. Yip, (1995) Structure, 3, 1147-1158.

219. K.S.P. Yip, K.L. Britton, T.J. Stillman, J. Lebbink, W.M. De Vos, F.T. Robb, C. Vetriani, D. Maeder, D.W. Rice (1998) *Eur. J. Biochem., 255, 5759-5765.*

220. A.H. Elcock (1998) *J. Mol. Biol*. 284, 489-502.

221. L. Xiao, B. Honig (1999) *J. Mol. Biol,* 289, 1435-1444.

222. S. Kumar, B. Ma, C.J. Tsai, R. Nussinov (2000) *Proteins,* 38, 363-383.

223. N. Kannan, S.Vishveshwara (2000) *Prot. Eng.* 13(11), 753-761.

224. K.V. Brinda, S. Vishveshwara (2005) *Biophys J*, 89, 4159-4170.

225. A.L. Barabasi (2002) Linked: The New Science of Networks. Persues Publishing, Cambridge, MA

Chapter 2

Equilibrium unfolding of cyclophilin from *leishmania donovani* (LdCyp): Characterization of intermediate state(s) biophysically and through molecular dynamics simulation.

2.1 Introduction

Cyclophilin from *leishmania donovani* (LdCyp) belongs to the ubiquitous class of peptidyl-prolyl cis-trans isomerases (PPIases), also known to be the intracellular receptor of the immunosuppressive drug cyclosporine A (CsA), a cyclic undecapeptide constituted of non-natural amino acids. CsA derivatives formed by specific modification of its side chains have been shown to lack immunosuppressive activity though retaining its anti-trypanosomatid and anti-parasitic character **[1, 2]**. In addition, cyclophilins have also been implicated in signal transduction, cell division, cell surface recognition, chaperone activity and heat shock response [**3**]. Other studies show Cyp's to be involved in the maintenance of mitochondrial transition pores [**4**] and regulation of HIV-1 infectivity by functional association with HIV-1 virions in humans [**5**].

Several crystal and NMR structures of CyPs in both ligated / unligated forms are currently available **[6-7]**. Cyclophilin (LdCyp) has been used as the model system in the present study of equilibrium protein unfolding under the influence of chaotropic denaturant guanidium chloride (GdmCl). The three dimensional structure of LdCyp (2HAQ: 1.97Å) **[8]** consists of an 8 stranded ß-barrel with two α helices located at either end. The location of the helices with respect to the barrel effectively blocks the solvent accessibility of the only hydrophobic core of the molecule, located in the interior of the barrel. The single domain molecule consists of a single cysteine

residue and its enzymatic activity is hindered by the binding of CsA as there is considerable overlap between the native active site of the enzyme and the binding site for CsA, both being located on the face of the barrel. LdCyp consists of a single tryptophan (Trp 143 also involved in its active site) which is situated on a flexible 3_{10} helix (residues 141-145) and is partially exposed to the solvent (**Figure 1**).

Given the intrinsic interest of Cyp's in general and specifically in *leishmania*, LdCyp has been selected for unfolding studies, both thermally and with denaturant. There are reports of thermal denaturation studies on cyclophilin from *mycobacterium tuberculosis* both in the presence and absence of CsA. CsA binding did not appear to impart any extra stabilization to the protein, though facilitated the formation of secondary structural elements at lower temperatures (25°C). The peptidyl - prolyl isomerase (cyclophilin A from *mycobacterium tuberculosis*) tended to aggregate beyond 70°C and unfolding induced via denaturant demonstrated that the loss in tertiary structure preceded the loss in secondary structure both in the presence and absence of CsA [**9**].

In this chapter we have studied the unfolding of LdCyp by guanidium chloride (GdmCl) by means of fluorescence, far and near UV- circular dichroism (CD) and differential scanning calorimetry (DSC). The results from the studies indicate that most probably the unfolding of LdCyp proceeds via at least one equilibrium intermediate. We have made an attempt to structurally characterize this possible intermediate(s) using multiple spectroscopic tools and molecular dynamics simulations at temperatures of 310,400 and 450K.

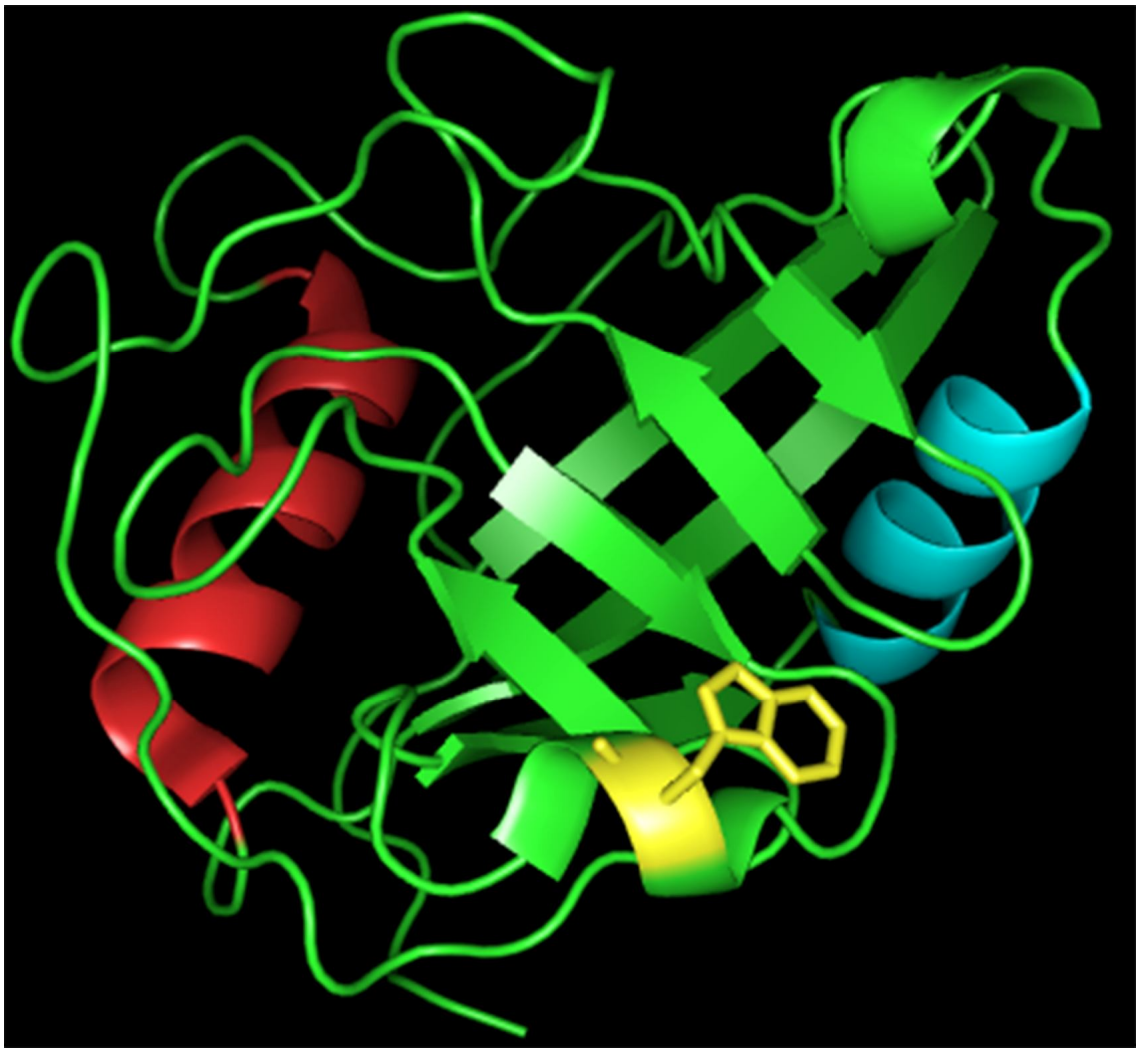


Figure 1: Native crystal structure of cyclophilin from *Leishmania donovani* (LdCyp) in green (cartoon representation) with the partially solvent exposed lone tryptophan residue Trp 143 (in yellow, ball and stick mode). Helix H1 is colored red and Helix H2 colored cyan. The Figure is generated using Pymol (Schrodinger LLC).

2.2. Materials and Methods:

2.2.1. Reagents and Chemicals

Guanidium chloride (GdmCl), bis-anilino-8-napthalenesulphonate (bis-ANS) were obtained from Sigma-Aldrich (USA). Potassium dihydrogen orthophosphate (KH_2PO4) and anhydrous dibasic potassium phosphate (K_2HPO4) were purchased from Sisco Research Laboratories (SRL) and Nickel–nitrilotriacetic acid-agarose, superflow (Ni-NTA) from Qiagen (USA). All other chemicals used were of analytical grade. All the requisite buffers were prepared in deionized Mili-Q water (Millipore, USA).

2.2.2. Protein Purification

The expression vector pQE32 (qiagen, USA) was transformed into *E.Coli M15* cells, which contained the LdCyp gene. The first 21 amino acids of the protein constitutes a signal sequence that gets cleaved post-translationally in the native protein, so the native protein consists of residues from 22-187 fused into a N-terminal 6x-His tag. 500mL of Luria broth supplemented by 100µg/ml of ampicillin and 50µg/ml of kanamycin was inoculated with 2mL of log-phase culture and was allowed to grow till OD_{600} reached 0.6. To induce the expression of LdCyp, IPTG was added and the culture was allowed to grow further for 6-8 hours at 37°C. Cell harvesting was carried out by centrifugation at 8000g for 10 min at 15°C. The cells were then washed twice with PBS buffer, pH 7.4. Cell pellets were then re-suspended in lysis buffer that comprised of 50mM sodium phosphate buffer pH 7.5, 300mM NaCl and 10mM imidazole. Resuspended cells were then centrifuged at 18000g for 40 min, after which the lysate was passed through approximately 3-4 times through a nickel-nitrilotriacetic acid (Ni-NTA) column pre-

equilibrated with lysis buffer. The Ni-NTA column was then thoroughly washed with wash buffer containing 50mM sodium phosphate buffer pH 7.5, 300mM NaCl and 20mM imidazole. LdCyp bound to the column was eluted using a elution buffer containing 50mM sodium phosphate buffer pH 7.5, 300mM NaCl and 250mM imidazole. The eluted fractions containing the protein was pooled and then dialyzed extensively against a buffer containing 20mM Tris.HCl pH 8.5 and 0.02% NaN_3. Purity of the dialyzed protein was judged from a single band on SDS-PAGE. The protein was re-dialyzed extensively in 25mM potassium phosphate buffer pH 7.5, prior to all spectroscopic and calorimetric experiments. The concentration of the protein sample was estimated using a molar extinction coefficient of 9230 $mol^{-1}cm^{-1}$ at 278nm.

2.2.3. Fluorescence Measurements

Intrinsic tryptophan fluorescence of the protein was measured at 25°C on a Perkin Elmer LS55 spectrofluorimeter (Perkin Elmer Ltd., UK). LdCyp at 4 μM concentration in 25mM potassium phosphate buffer (pH 7.5) was excited at 295nm and the emission spectra was recorded in the range 310nm to 450nm with both excitation and emission slit widths kept at 5nm. The spectra recorded at a scan speed of 200nm/min were averaged over 4 scans. Appropriate controls for the background emission were subtracted in each case.

. Quenching experiments with the neutral quencher acrylamide was performed for LdCyp (4μM) pre-incubated with increasing concentrations of chaotropic denaturant guanidium chloride (GdmCl). The acrylamide concentration in the protein sample was varied from (10-50mM) and fluorescence spectra measured ranging from 310-450nm, with excitation at 295nm. Stern-Volmer quenching constant was calculated as the slope of the plot of F_0/F values at 360nm against the input concentration of quencher, acrylamide from the Stern-Volmer equation:

$$F_0/F = 1 + K_{SV}[Q] \qquad (1)$$

where Ksv is the Stern-Volmer constant and [Q] is the concentration of the quencher. All quenching experiments were repeated thrice.

Binding of hydrophobic fluorescent dye, bis-ANS (bis-anilino-8-napthalenesulphonate), with LdCyp was carried out to probe the intermediate conformational states of LdCyp (4µM, 25 mM potassium phosphate buffer , pH 7.5) in varying concentrations of the denaturant (GdmCl, 0.0-3.0 M). The protein pre-incubated overnight with the denaturant was equilibrated with bis-ANS for 20 minutes before recording the spectra. Ensuring complete saturation of the dye binding sites in the native protein the concentration of bis-ANS in the samples was kept fixed at 10µM. Bis-ANS emission spectra was recorded from 410 to 675 nm, with excitation at 395 nm and both emission and excitation slit widths were kept at 5nm. Appropriate buffer corrections were done.

2.2.4. Circular Dichroism Measurements

The CD spectra of LdCyp were measured on a JASCO J715 (Jasco Corporation , Tokyo , Japan) spectropolarimeter in a rectangular quartz cell of path length 1mm. Far UV measurements of LdCyp were recorded in the wavelength range of 200-250 nm. For all CD experiments in 200-250nm region, spectra were obtained by averaging over 4 scans, utilizing a slit width of 1nm. For protein unfolding with chemical denaturants far UV spectra were recorded from 210-250nm in case of GdmCl concentrations exceeding 1.0 M. Near UV spectra were recorded in the region 250-300nm to probe the denaturant induced alteration of tertiary structure of the protein. Buffer contributions alone or with denaturants were subtracted from all protein spectra.

The observed values of ellipticity (after appropriate buffer correction) were converted to mean residue ellipticity (MRE) in $degcm^2dmol^{-1}$ defined as

$$\mathrm{MRE} = \frac{\mathrm{M}\theta_\lambda}{10\mathrm{dcr}} \qquad (2)$$

where M is the molecular weight of the protein (in Daltons) , θ_λ is the CD in millidegree, d is the path length in cm, c is the protein concentration in mg/ml and r the number of amino acid residues in the protein polypeptide chain.

2.2.5. Differential Scanning Calorimetry

10μM of LdCyp was scanned on a VP-DSC Microcalorimeter (Microcal LLC, Northampton, MA, USA) from 10°C to 70°C with a scan rate of 30°C/h and at approximately 28 psi pressure. All samples (buffer and protein) were extensively degassed prior to loading. Initially, the buffer (25mM potassium phosphate, pH 7.5) or buffer combined with denaturant (GdmCl, 0.2-1.4M) were repeatedly scanned to ensure a stable baseline. The buffer baseline was subtracted from the protein thermogram using Microcal Origin version7.0 software provided with the instrument. The data was then normalized by the protein concentration and a non-linear curve fitting algorithm was employed to obtain the thermodynamic parameters of the transition. The reversibility of unfolding was checked at different scan rates, buffer conditions and temperature ranges. All DSC experiments were repeated at least three times.

2.2.6 Molecular Dynamics Simulations

Unfolding molecular dynamics simulations were performed on cyclophilin (coordinates obtained from: 2HAQ) at temperatures 310,400 and 450 K. Each simulation run was for a duration of 50ns. The simulation at 310K could be considered to be 'native'. Initially, the protein was

solvated in a cuboidal box of dimensions 78.683, 68.897 and 78.137 Å by the addition of 11088 waters following the TIP3P model and charge neutralization of the system was accomplished by the addition of Na^+ ion utilizing the xleap module of AMBER **[10].** The structure was then energy minimized (AMBER 2002 force field) for 200 steps of steepest descent followed by 19800 steps of ABNR incorporated in the SANDER module. All simulations were performed using NAMD 2.0 **[11]**. For each 50ns run the targeted temperature was attained in increments of 10K/ps for a NPT ensemble with the Langevin piston set to the targeted temperature (310,400,450 K etc.) and the pressure fixed at 1.032 bar. Bond lengths were constrained at a tolerance level of 0.005 Å by the SHAKE **[12]** algorithm. Visual molecular dynamics program (VMD 1.9.1) **[13]** was used to view and write the trajectories in PDB format obtained from NAMD. The time step (δt) for the velocity - verlet algorithm **[14]** was kept fixed at 1fs. Snapshots were sampled at intervals of 10 ps. Secondary structural content for the snapshots were calculated using STRIDE **[15]**.

2.3. Results and Discussions

2.3.1. Unfolding of LdCyp with GdmCl

As mentioned previously, LdCyp contains a single tryptophan residue (**Figure 1**). The maximum intensity of intrinsic fluorescence, λ_{max}, was observed at 348nm for the native protein suggestive of the partial solvent exposure of the lone tryptophan. The emission spectrum changed with the variation of denaturant concentration till 1.4M GdmCl. Further increase in the concentration of GdmCl led to a slight decrease in the intensity (**Figure 2**).

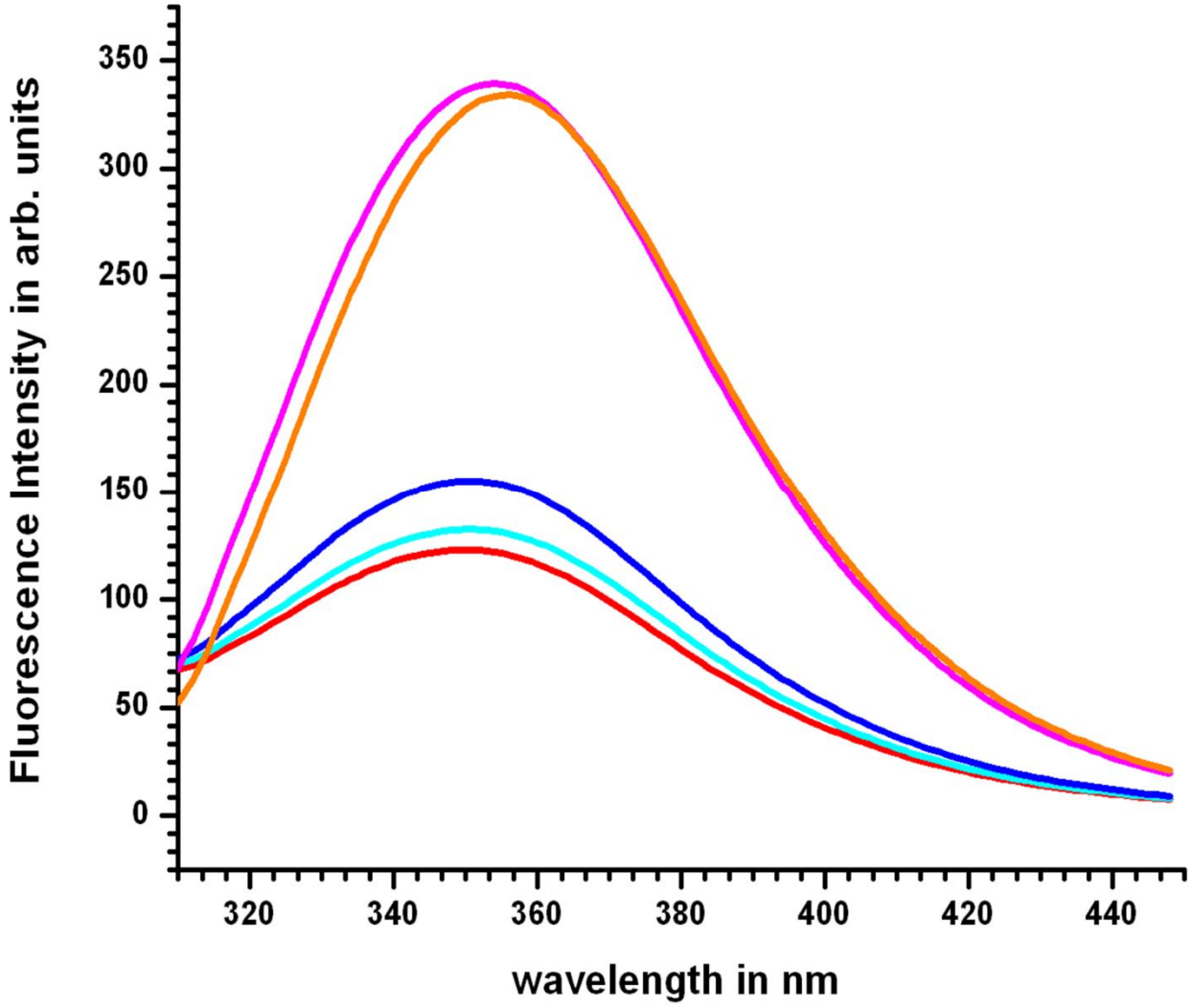


Figure2. Fluorescence unfolding spectra of native cyclophilin at various GdmCl concentrations ranging from (0-3) M a) (—) 4µM LdCyp b) (—) 4µM LdCyp+0.8 M GdmCl c) (—) 4µM LdCyp + 1.0 M GdmCl d) (—) 4µM LdCyp +1.4M GdmCl e) (—) 4µM LdCyp+3.0M GdmCl.

The intensity at 360nm was plotted as a function of GdmCl concentration (**Figure 3**). With increase in GdmCl concentration unfolding was initiated around 1.0M and completed by 2.0M, as clearly indicated by the saturation in fluorescence intensity.

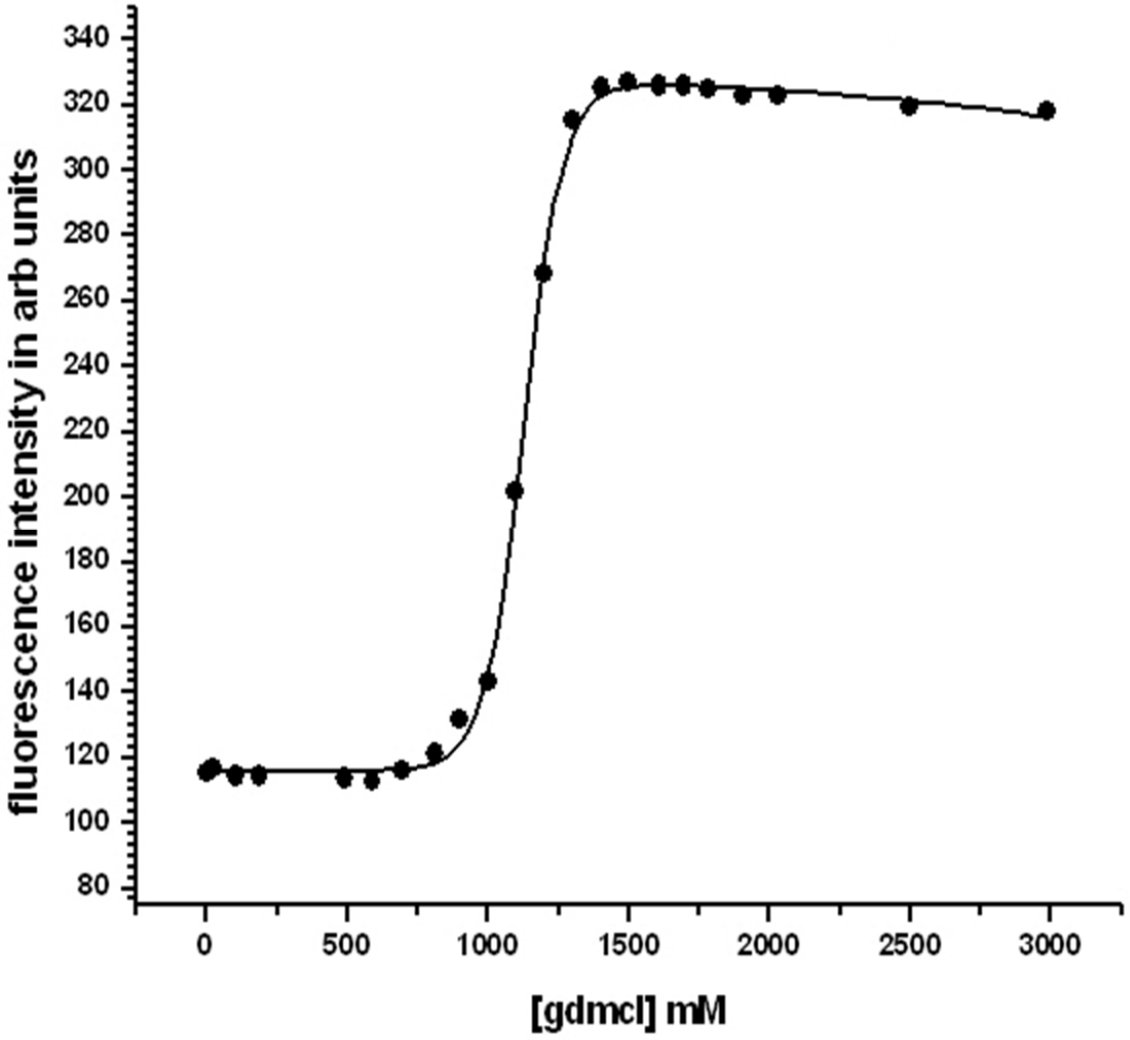


Figure3. Unfolding curve of LdCyp induced by GdmCl as monitored by intrinsic tryptophan fluorescence with emission maxima points at 360nm, fitted to a 3-state equation. The black circles (●) indicate the fluorescence intensity in arbitrary units at the corresponding denaturant concentrations (in mM) and (—) is the fitted line to the datapoints.

The unfolding curve of LdCyp fitted best to a three state equation originating from the following equilibrium: N⇌I⇌U

$$S_{obs} = \frac{S_N + S_I e^{\frac{-\Delta G_{NI}}{RT}} + S_U e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (3)$$

where S_{obs} is the signal at any denaturant concentration , S_N is the signal due to the native state ,S_I the signal due to the intermediate state and S_U the signal due to the unfolded state. ΔG_{NI} and ΔG_{NU} are the corresponding free energies of the transitions from N⇌I and N⇌U states [**16**] where ΔG_{NI} and ΔG_{NU} are assumed to have linear dependence with the denaturant concentration [D], resulting in the following equations,

$$\Delta G_{NI} = \Delta G_{NI}(H_2O) + m_{NI}[D] \qquad (4)$$

$$\Delta G_{NU} = \Delta G_{NU}(H_2O) + m_{NU}[D] \qquad (5)$$

where m_{NI} and m_{NU} are the corresponding slopes of transitions from the native to the intermediate and native to the unfolded states respectively. C_m for any unfolding curve is defined as the midpoint of the transition from the native to the unfolded state (C_m(N↔U) : for a 2 – state transition) or between the native to intermediate and native to the unfolded state respectively , (that is C_m(N↔I), C_m(N↔U) for a three state transition) obtained from equations (4) and (5) when the corresponding free energies of transition from the native to the intermediate state ΔG_{NI} and from the native to the unfolded state ΔG_{NU} becomes zero. The above fit yielded $\Delta G_{NI}(H_2O)$ of 12.13 kcal mol^{-1} with m_{NI} of -8.166 kcal mol^{-1}M^{-1} and $\Delta G_{NU}(H_2O)$ of 6.72 kcal mol^{-1} and m_{NU} of 5.89 kcal mol^{-1} M^{-1} (**Table 1**) which suggests that the unfolding of N to I

exposes more solvent accessible surface area than the unfolding of N to U, which is also clear from other experiments like (acrylamide quenching and bis-ANS fluorescence studies).

Table 1

Equilibrium thermodynamic parameters of unfolding for LdCyp mediated by GdmCl.

Spectroscopic Tools employed	Thermodynamic parameters					
	$\Delta G_{NI}(H_2O)$ kcal mol^{-1}	m_{NI} Kcal mol^{-1} M^{-1}	$\Delta G_{NU}(H_2O)$ kcal mol^{-1}	m_{NU} Kcal mol^{-1} M^{-1}	$C_m(N\leftrightarrow I)$ M	$C_m(N\leftrightarrow U)$ M
Intrinsic tryptophan fluorescence	12.13 (± 0.38)	-8.17 (± 0.92)	6.72 (± 0.54)	-5.89(±0.65)	1.47±0.06	1.16 ±0.04
Far UV Circular Dichroism	2.12 (±0.13)	-2.18 (± 0.29)	5.24 (±0.45)	-3.47 (± 0.43)	0.99 ± 0.02	1.51 ± 0.04

The thermodynamic parameters in Table 1 were obtained from the 3-state equation, equation (3) by employing a non-linear least square algorithm. ΔG_{NI} (H_2O) and ΔG_{NU} (H_2O) refer to the free energy of transition from the native to the intermediate state and native to the unfolded state, respectively at zero denaturant concentration with ΔG_{NI} and ΔG_{NU} going to 0. m_{NI} and m_{NU} are the corresponding slopes of the transition from the native to the intermediate and native to the unfolded states, $C_m(N\leftrightarrow I)$ and $C_m(N\leftrightarrow U)$ are the midpoints of transition (denaturant concentration) obtained from equations (4) and (5), where ΔG_{NI} and ΔG_{NU} goes to 0.

The far UV CD of the native protein exhibited a pronounced minimum at 225nm characteristic of a protein with prevalently β-sheets. With the addition of GdmCl, MRE value at 222nm progressively reduced in the concentration range of 0.8-2.0 M, saturating thereafter. This was also accompanied by a gradual and concomitant reduction in the overall secondary structural

content till 1.2 M GdmCl, after which there was an abrupt loss of all secondary structural features in the protein (**Figure 4**). Thus the CD signature of secondary structural features was observed till 1.2 M GdmCl.

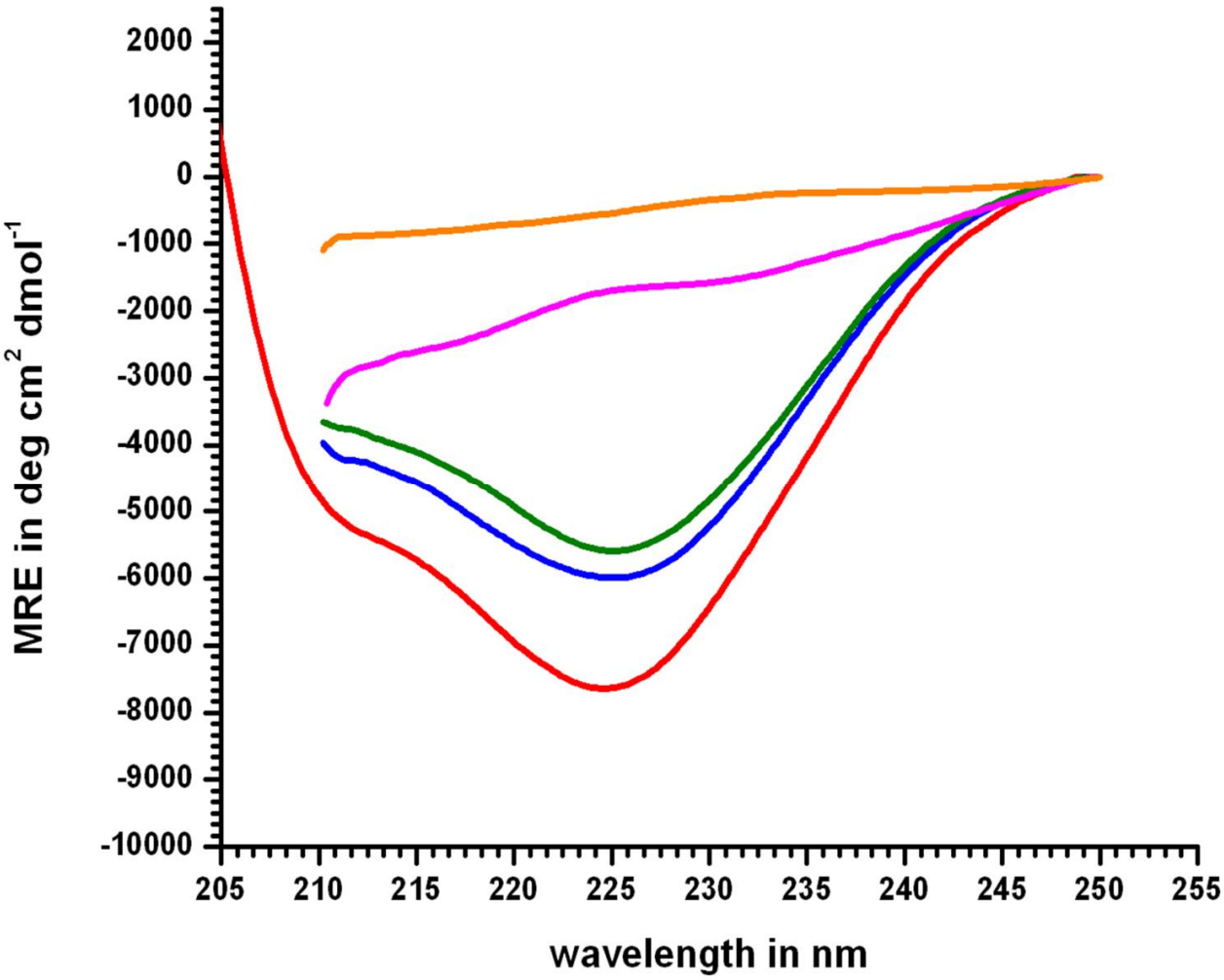


Figure4. Far-UV CD spectra of LdCyp (native) along with different denaturant concentra tions a) (—) 4µM b) (—)4µ M+ 1M GdmCl c) (—)4µM+ 1.2 M GdmCl d)(—) 4µ M+1.4 M GdmCl e) (—) 4µM+ 3.0 M GdmCl.

Similar to the feature obtained from fluorescence titration profile (**Figure 3**), the unfolding curve from CD value at 222 nm was best fitted to the above three state equation yielding ΔG_{NI} (H_2O) and C_m1 of 2.12 (±0.13) kcal mol^{-1} and 0.97 (±0.11) M and ΔG_{NU} (H_2O) of 5.24 (±0.16) kcal mol^{-1} and C_m2 of 1.51 (±0.2) M respectively (**Figure 5).**

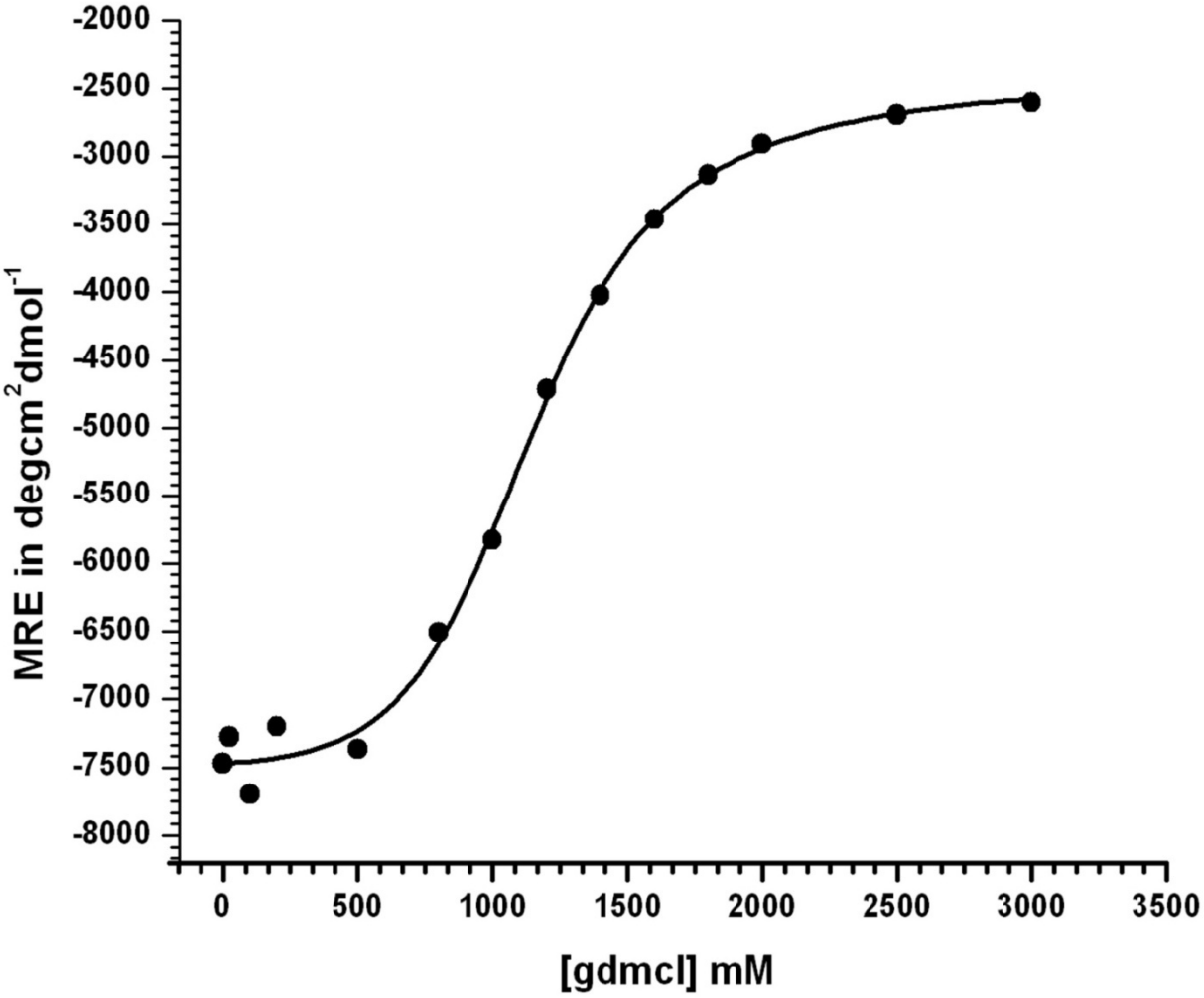


Figure5. MRE_{222} [in black circles (●)] values obtained from far-UV CD spectra plotted as a function of the denaturant and fitted to a 3-state equation shown as a black (—) line.

In addition, CD unfolding curves at MRE_{225} and MRE_{215} (**Figure 6, Table 2a,b,c**) gave similar C_m values though some deviation in the curves were observed between MRE_{222}, MRE_{225} on one hand and MRE_{215} on the other.

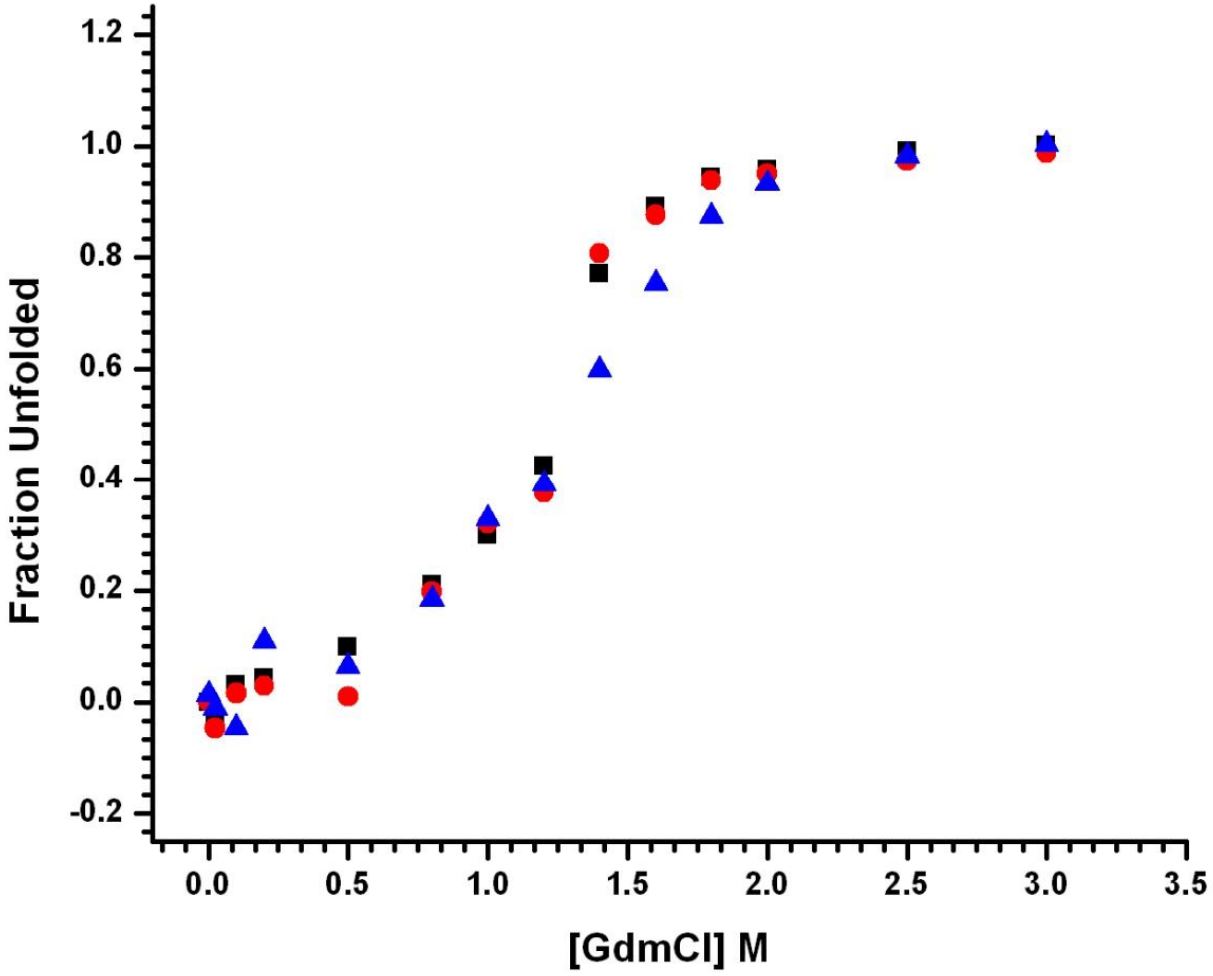


Figure6. Far-UV CD unfolding curves of LdCyp monitored at MRE_{222} (■) black squares, MRE_{225} (●) red circle and at MRE_{215} (▲).

Table 2a: Circular Dichroism data at (MRE_{222}), fitted to a 3-state model. The experiment was performed in triplicate and the standard deviations are given in parentheses.

S.No.	C_m (N↔I) M	C_m (N↔U) M	Adj R_{sq}	Mean C_m (N↔I) M	Mean C_m (N↔U) M
1.	0.97	1.51	0.995	0.99 (.02)	1.51 (.04)
2.	1.02	1.56	0.990		
3.	0.993	1.47	0.992		

Table 2b: Circular Dichroism data at (MRE_{225}), fitted to a 3-state model. The experiment was performed in triplicate and the standard deviations are given in parentheses.

S.No.	C_m (N↔I) M	C_m (N↔U) M	Adj R_{sq}	Mean C_m (N↔I) M	Mean C_m (N↔U) M
1.	1.15	1.43	0.992	1.09 (.05)	1.44 (.02)
2.	1.02	1.46	0.993		
3.	1.09	1.42	0.992		

Table 2c: Circular Dichroism data at (MRE_{215}), fitted to a 3-state model. The experiment was performed in triplicate and the standard deviations are given in parentheses.

S.No.	C_m (N↔I) M	C_m (N↔U) M	Adj R_{sq}	Mean C_m (N↔I) M	Mean C_m (N↔U)
1.	1.05	1.53	0.982	1.09 (.05)	1.44 (.02)
2.	0.99	1.61	0.980		
3.	1.16	1.56	0.993		

$C_{m(}$N↔I) and C_m(N↔U) are the midpoint of transitions from the native to the intermediate and native to the unfolded states (denaturant concentrations) obtained from equations (4) and (5), where ΔG_{NI} and ΔG_{NU} the corresponding free energies of transitions goes to zero, a non-linear curve fitting algorithm was employed to fit the 3-state model , equation(3).

The lack of superposition of the unfolding curves obtained by two independent probes CD and fluorescence and the consequent non-equivalence of their respective C_m values (**Table 1**) appears to be supportive of a non – two state transition and thereby indicates the presence of a possible intermediate in the unfolding of LdCyp: N⇌I⇌U (**Figure 7**).

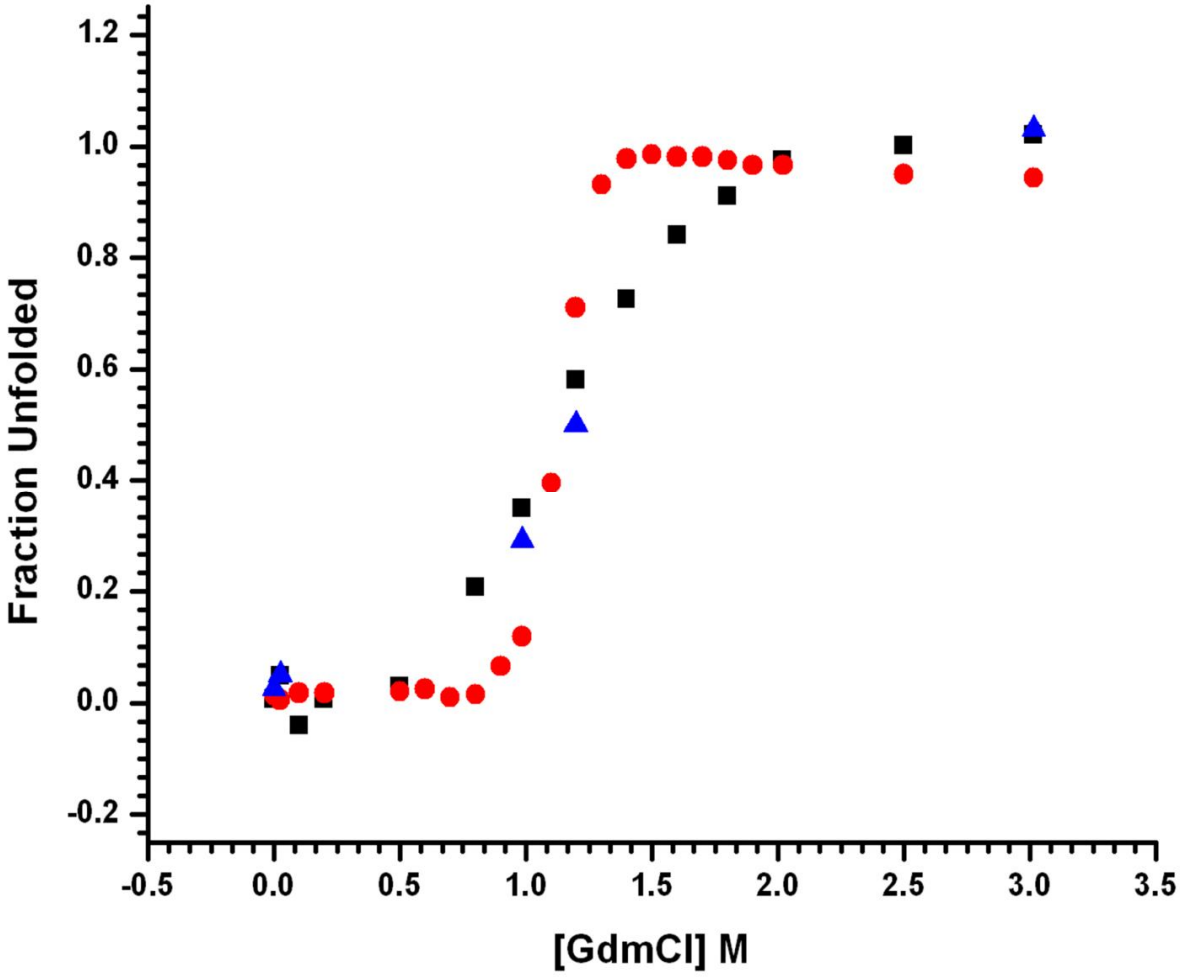


Figure7. The fraction unfolded versus GdmCl curve with independent probes of intrinsic tryptophan fluorescence monitored at 360nm, shown in filled red circles(●) , far UV CD monitored at 222nm in black squares (■) and near UV CD monitored at a maxima of 260nm shown in blue triangles(▲).

There are numerous instances in the literature where non-coincidence of C_m values in equilibrium unfolding experiments [**16-18**] via two independent probes has been cited as evidence for the presence of equilibrium intermediate. The adoption of a three state model improved the fitting in the protein unfolding curve monitored by fluorescence, when compared to a 2 state model **(Figure 8).**

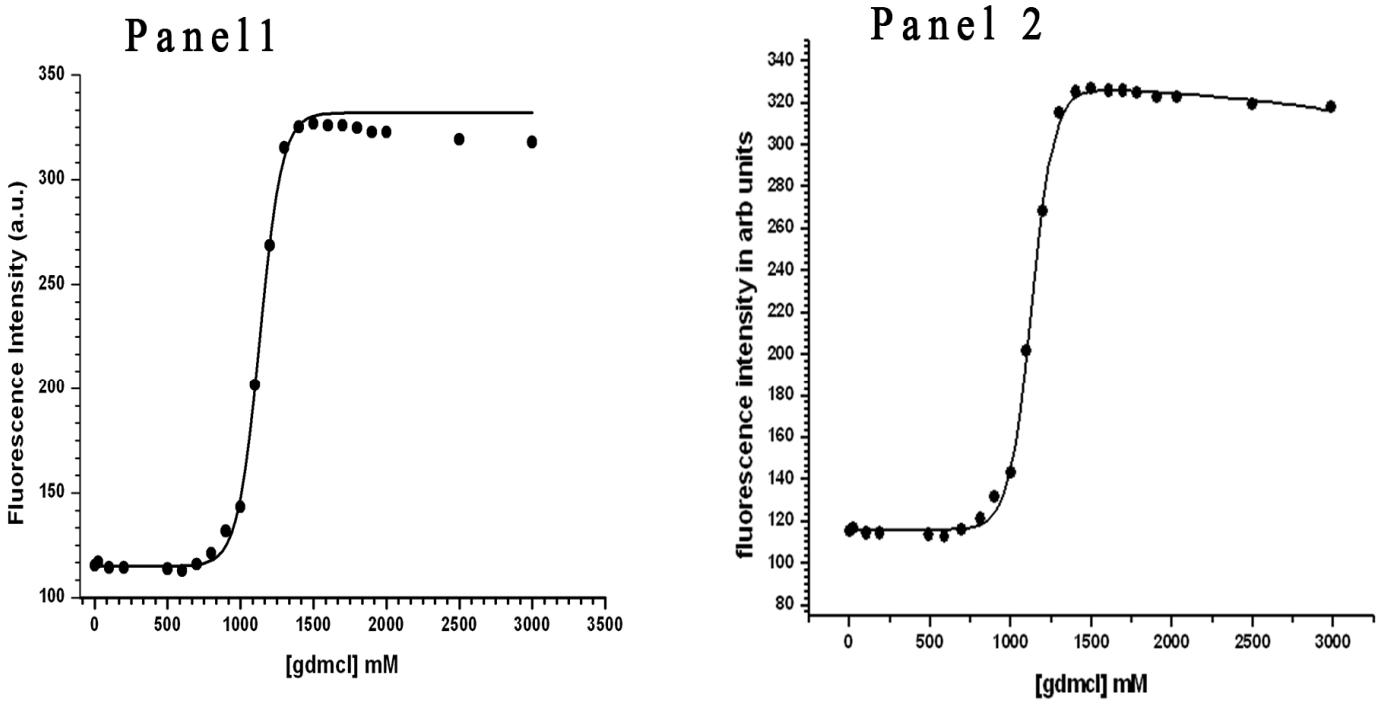


Figure8. Fluorescence unfolding curve of LdCyp monitored at 360 nm and fitted to a) two state model in Panel1 and b) three state model in Panel 2.

However, for unfolding curves monitored by CD (MRE_{222}) the improvement was only marginal (3-state compared to 2-state), **(Figure 9).** Notably, a 2-state transition could also be considered

adequate in protein unfolding probed by CD.

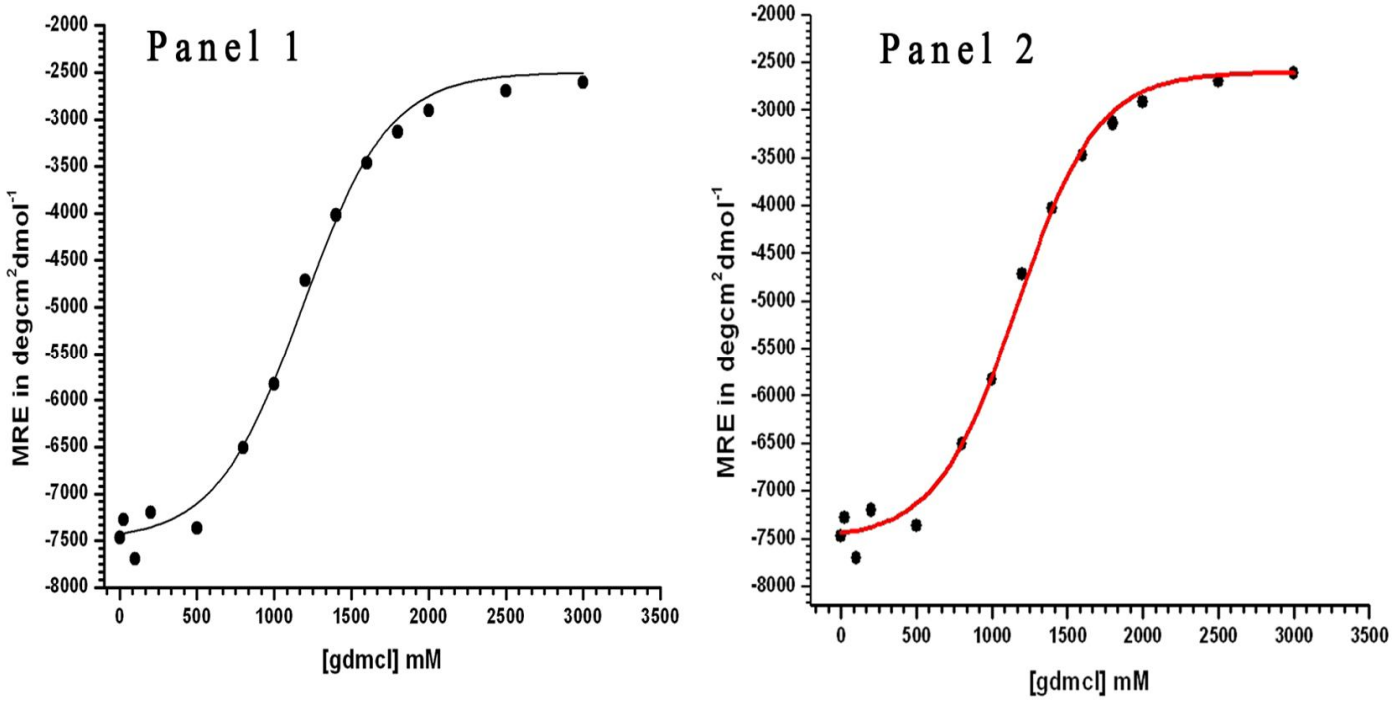


Figure9. Far UV CD unfolding curve of LdCyp monitored at 222nm being fitted to a) 2-state process in panel 1 b) and 3-state process in panel 2.

In any case for both fluorescence, CD the improved fitting (3-state versus 2-state) was only mildly reflected in their respective R^2 values. However, three state fittings were retained for both fluorescence and CD for consistency with bis-ANS fluorescence intensity (recorded at 488nm) curves as function of GdmCl concentration and also because of the lack of an isochromatic point in the far UV CD data (ranging from 210 – 250 nm) , generally indicative of a non – two state transition. Near UV CD (250-300nm) spectra of native LdCyp show a broad shoulder in the region of 260-285nm. CD spectra of LdCyp incubated with low concentration of GdmCl (0.2M) indicate that the native tertiary interactions are relatively preserved. In contrast

LdCyp incubated with 1.0 M and 1.2M of GdmCl shows considerable perturbation in the native tertiary contacts as manifested by the reduced intensity of the broad shouders (**Figure. 10**). Absence of tertiary interaction at high GdmCl concentration is evident from the featureless spectra with a significant reduction in band intensity.

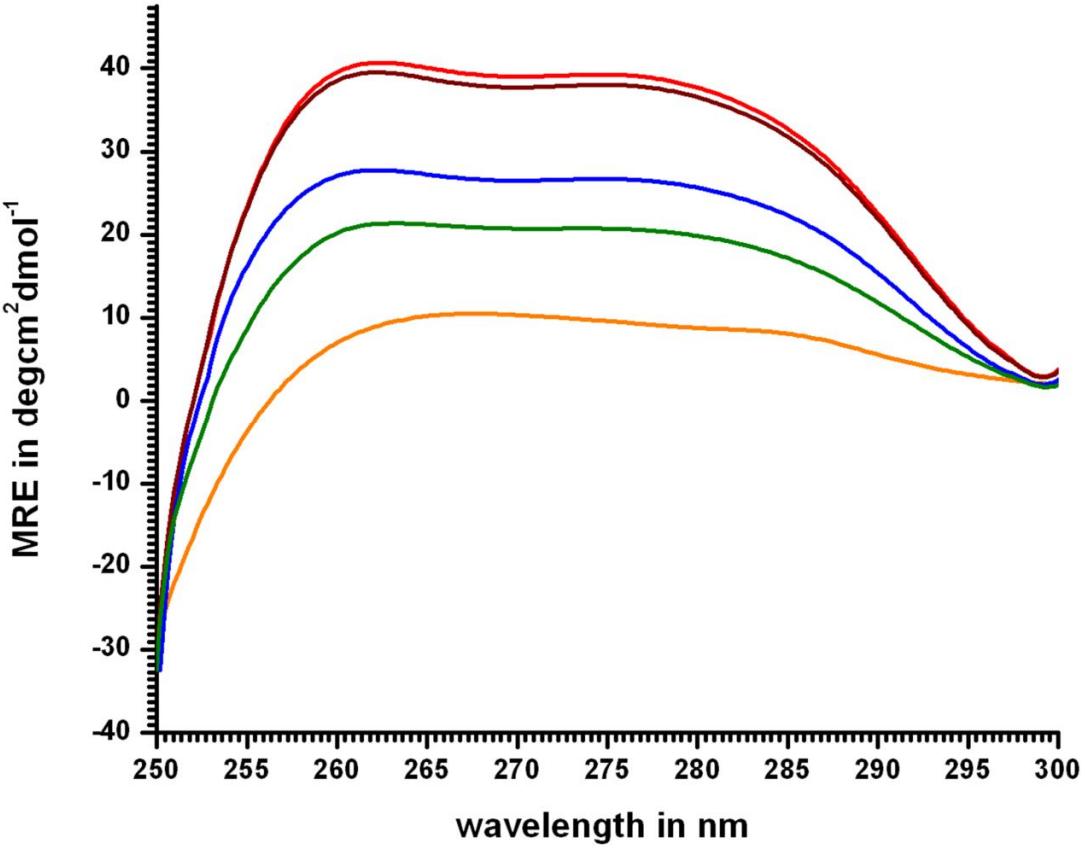


Figure 10: Near UV CD spectra of LdCyp for a) native protein LdCyp (—) 40 µM b) LdCyp + 0.2M GdmCl (—) maintaining a characteristic broad shoulder which considerably flattens out for c) LdCyp+ 1.0M GdmCl (—) d) LdCyp + 1.2M GdmCl (—) e) LdCyp + 3.0 M GdmCl (—), thereby depicting the loss of tertiary interactions with increasing denaturant concentrations.

Summing up, the results from CD studies suggest that LdCyp incubated with 1.0 and 1.2 M of GdmCl shows considerable disruption in the original tertiary interactions whilst conserving their secondary structure.

2.3.2. Bis-ANS fluorescence experiment

Bis-ANS fluorescence studies have been widely used to identify intermediate states in protein un folding [**20, 21**]. Natively folded or completely unfolded proteins are unable to bind to bis-ANS, as in the former case there are no exposed hydrophobic patches and in the latter there are no orde red clusters of exposed hydrophobic patches. Thus the partial exposure of erstwhile buried hydro phobic residues characteristic of "molten globules" are ideal for bis-ANS interaction. For LdCyp, bis-ANS fluorescence intensity (recorded at 488nm) initially rose with GdmCl concentration upt o 1.2 M and declined thereafter with a minima at 3.0M (**Figure 11, Figure 12**). This observation strongly suggests that with increasing GdmCl concentrations the buried hydrophobic residues are solvent exposed till 1.2M, after which there is a collapse in the ordered structure of the exposed hydrophobic patches which denies bis-ANS of its binding site thus leading to the decline of the b is-ANS fluorescence intensity, reaching a basal level at 3.0M GdmCl. Both of these observations are indicative of a molten globule like intermediate in the equilibrium unfolding of LdCyp.

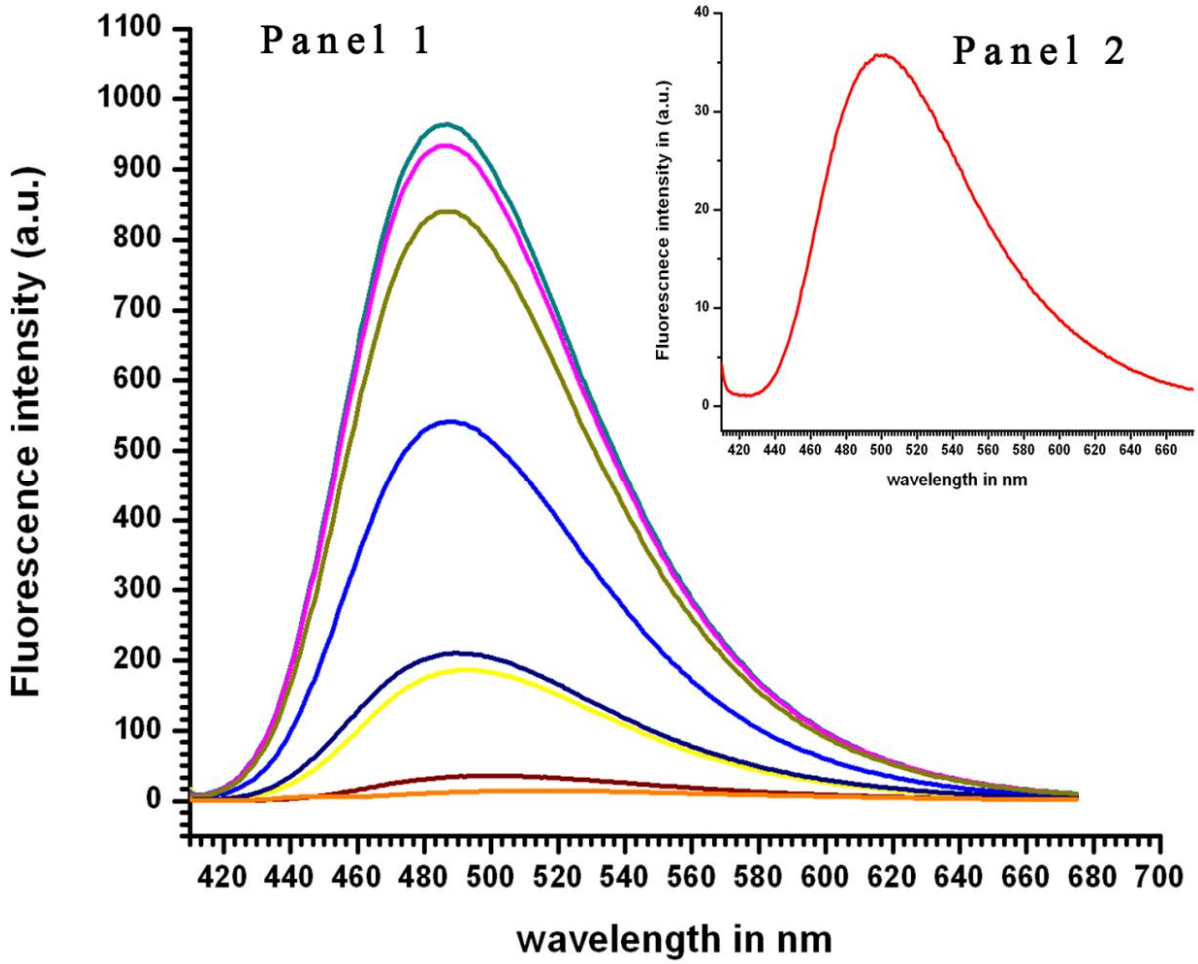


Figure11. Panel1: Bis–ANS fluorescence (10 µM) spectra for LdCyp (4µM) incubated with various GdmCl concentrations, Panel 1 : a) (—) 4 µM LdCyp+0.2M GdmCl b) c) (—) 4 µM LdCyp + 0.7M GdmCl d) (—) 4 µM LdCyp + 1.0 M GdmCl e) (—) 4 µM LdCyp + 1.2 M GdmCl f) (—) 4 µM LdCyp + 1.4 M GdmCl g) (—) 4 µM LdCyp + 1.5 M GdmCl h) (—) 4 µM LdCyp + 1.9 M GdmCl i) (—) 4 µM LdCyp + 3.0M GdmCl. The native spectra (—) 4 µM LdCyp is shown in Panel 2.

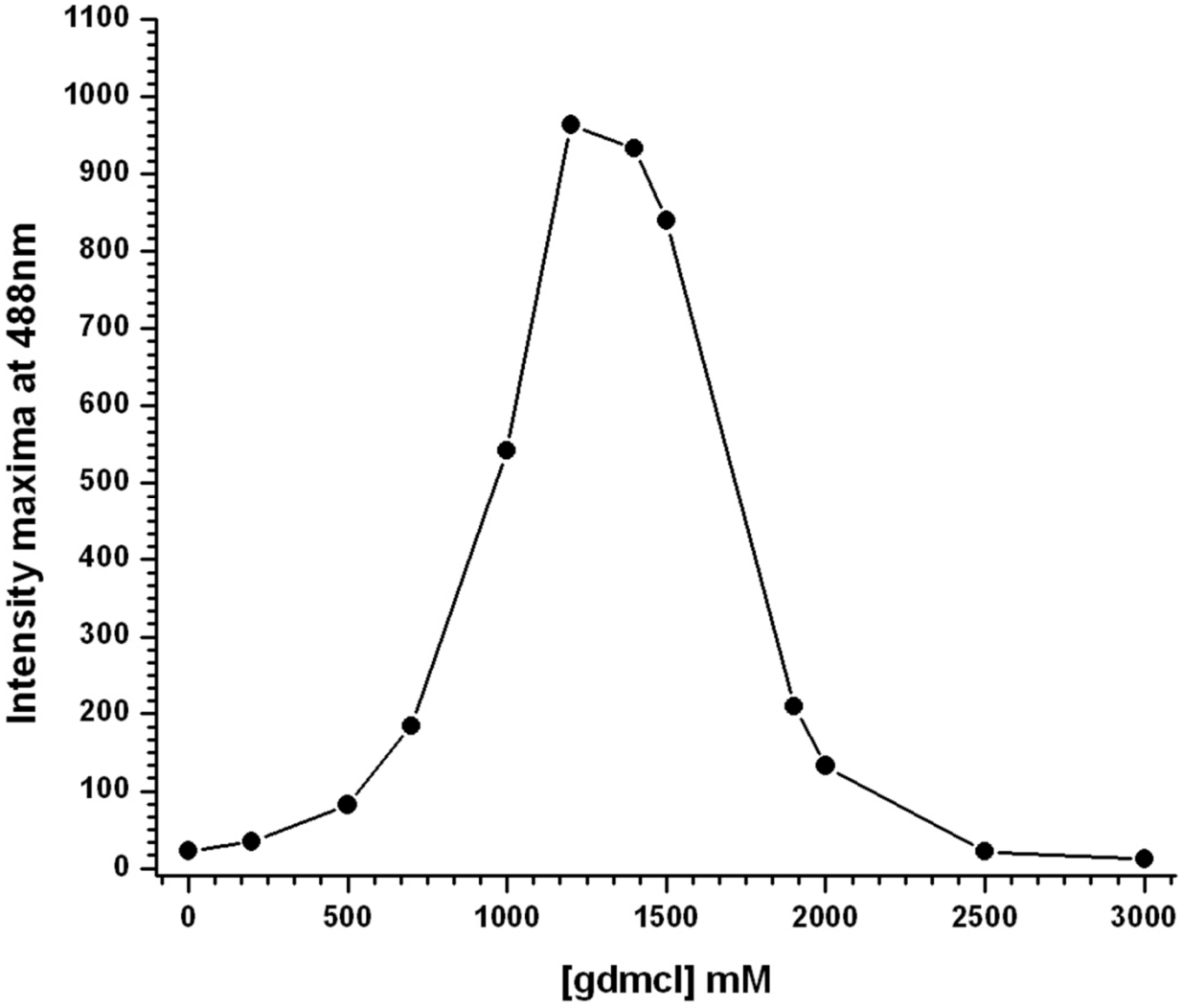


Figure12. Plot of Bis –ANS fluorescence (10 μM) intensity maximum in black circles (●) at maximum blue shifted wavelength of 488nm plotted as a function of GdmCl concentration, with maximum peak intensity obtained for protein sample (4 μM) incubated with 1.2M of GdmCl

2.3.3. Acrylamide Quenching

Stern-Volmer coefficients estimated for different denaturant concentrations (See **Materials** and **Methods, Section 2.3**) yielded a Ksv value of 7.64 M^{-1} for the native protein indicating partial so lvent exposure of its lone tryptophan. Ksv increased to 13.18 M^{-1} at 1.4M concentration of Gdm Cl after which the change is not significant (**Table 3**) at higher denaturant concentrations suggest ing maximum exposure of the said tryptophan to occur by 1.4M. The dynamic character of the q uenching was also confirmed by repeating the experiment at a higher temperature at 45 °C , whic h gave stern – volmer coefficients different from those at 25 °C **(Figure 13, Table 4)**

Table 3

Dependence of Ksv values on denaturant (GdmCl) concentrations

Denaturant Conce ntration in M	**Ksv values in M^{-1}**
0.0	**7.64 ± 0.23**
0.2	**8.84 ± 0.87**
1.2	**11.65 ± 0.94**
1.4	**13.18 ± 1.05**
3.0	**12.87 ± 1.11**

The Ksv values in Table 2 were calculated using the Stern-Volmer equation, equation (1) by linear least square fitting.

Table 4: Stern-Volmer constants and F_0/F values for native LdCyp at 25°C and 45 °C

Concentration of Acrylamide (M)	F_0/F (LdCyp at 25°C)	F_0/F (LdCyp at 45°C)
0.0	1.000	1.000
0.01	1.087	1.221
0.02	1.180	1.482
0.03	1.244	1.571
0.04	1.342	1.702
0.05	1.385	1.821
	Ksv = 7.87 M^{-1}	Ksv = 16.09 M^{-1}

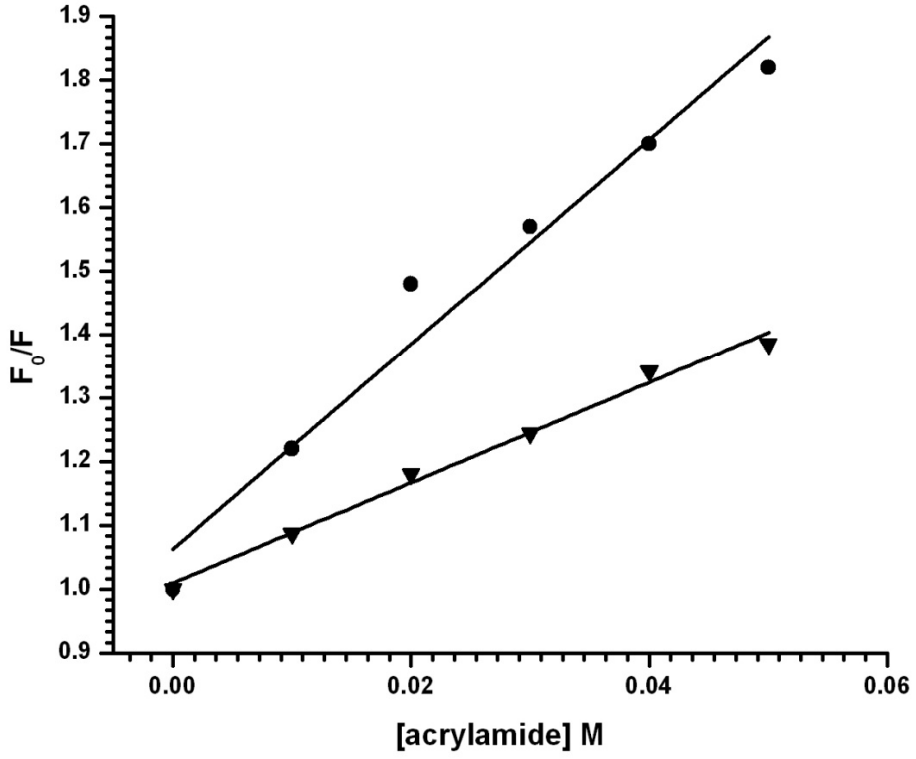


Figure13. F_0/F values plotted for both LdCyp at 25°C shown in inverted triangles in black (▼) and at 45°C, shown in filled black circles (●) as a function of acrylamide concentration.

2.3.4. Differential Scanning Calorimetry

Differential scanning calorimetry was performed with the native protein (LdCyp) and also in the presence of different denaturant (GdmCl) concentrations. For the native protein reversibility was observed only till 54.6°C, despite being tested on a wide range of buffering conditions and scan rates. The DSC thermogram of native LdCyp could be best fitted to a non-two state process with two T_m's at 49.6°C and 54.6°C. Different scan rates (at 20 and 60 °C/hour) did not significantly alter the position of the peak in Cp(T), the shape of the curve in the thermogram or the value of the two T_m's confirming that the shoulder observed in the thermograms were unlikely due to kinetic effects.' The apparent ΔH of the transitions were 100 kcal mol^{-1} and 117kcal mol^{-1}, respectively. DSC measurements under identical conditions with the addition of 0.2 M GdmCl gave a similar thermogram, but with reduction in both T_m's by about 5°C (44.9 and 50.6°C) and a concomitant decrease in their ΔH (75.2 and 104.7 kcal mol^{-1}), respectively. Although the non-two state character (with two T_m's) of the thermogram was still maintained at 0.6 M GdmCl, yet there was a significant reduction in the apparent enthalpy ΔH of the first transition (T_m = 41.0°C, ΔH=7.7 kcal mol^{-1}) relative to the second (T_m = 46.3°C, ΔH = 98 kcal mol^{-1}), with further decrease in the two T_m's. At 1.0M GdmCl a single transition was observed at T_m = 37.9°C and ΔH = 49.5 kcal mol^{-1} (**Figure. 14**). At still higher concentrations (1.4 M and above) no statistically significant peak was observed in the DSC measurement most probably indicating the initiation of the molten globule like intermediate, mentioned previously.

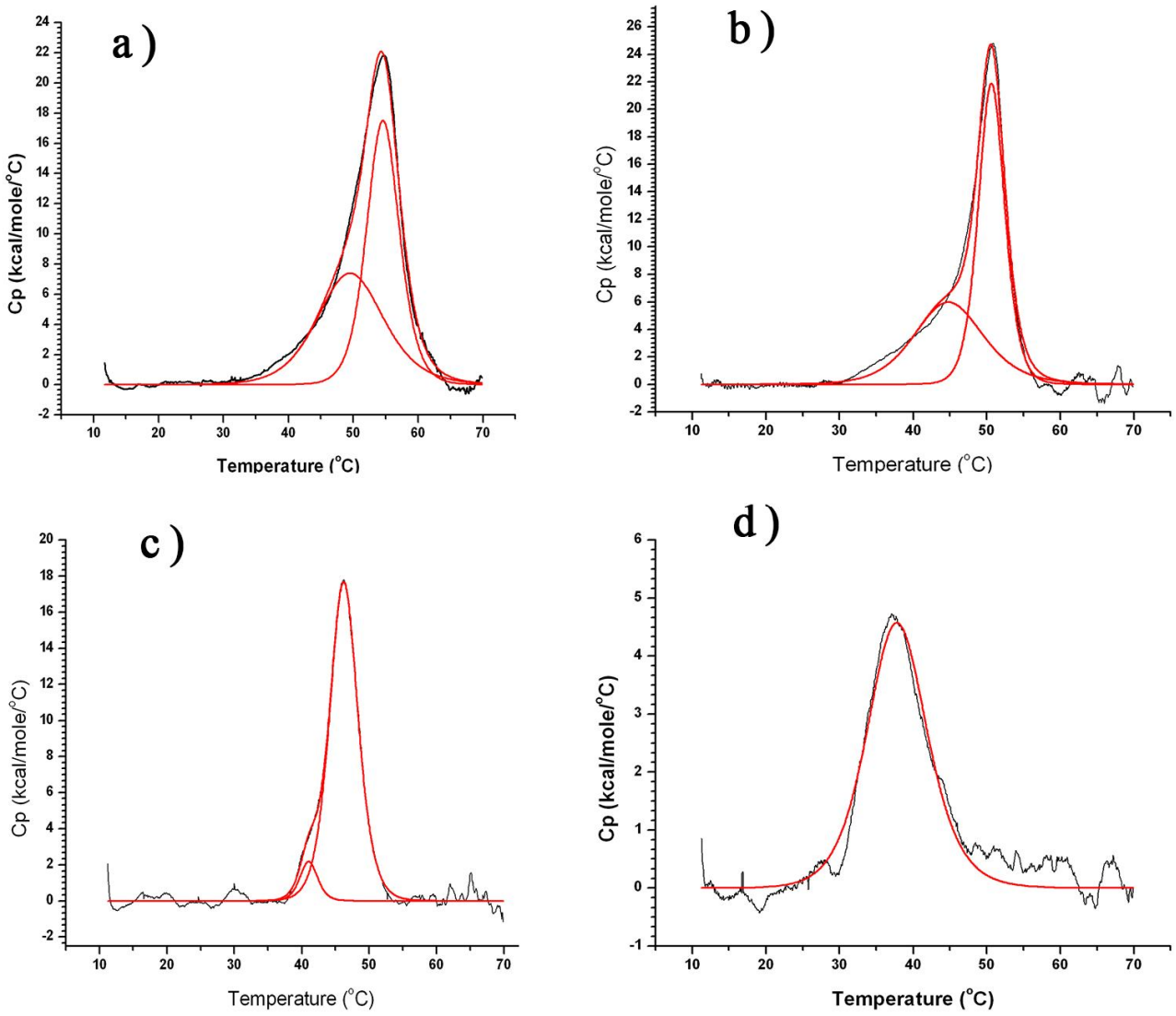


Figure14. Thermal characterization of the intermediate states of LdCyp through DSC a) 10 µM LdCyp b) 10µM LdCyp+0.2MGdmCl c) 10µM LdCyp+0.6M GdmCl d) 10µM LdCyp+ 1.0M GdmCl, two step (two T_m values) to a single step (one T_m value) transition.

The black solid line (—) corresponds to the experimental data points while the solid red lines (—) are the fitted lines.

2.3.5 Molecular Dynamics simulations

LdCyp consists of a single extended hydrophobic core, whose constituent residues are primarily contributed by the beta strands composing the barrel (**Figure. 1**) , the two helices (H1, H2) and a few from surrounding loops. Initially, three sets of residues were constructed; S1: residues constituting helix H1; S2: residues constituting helix H2 and S3: residues of the hydrophobic core excluding those from helices H1, H2 contributed by beta strands and loops (**Table 5**).

Table 5: List of residues comprising sets S1 (Helix H1), S2 (Helix H2), S3 (Residues constituting the core excluding core residues contributed by helices H1, H2). All residues composing the core are in bold.

S1 **Helix H1**	**Ala 52,** Pro 53, Leu 54, **Thr 55,** Thr 56**,** Glu 57**, Asn 58, Phe 59**, Arg 60**,** Gln 61, **Leu 62, Cys 63,** Thr 64**,** Gly 65
S2 **Helix H2**	Met 158, Asp 159, Val 160, **Val 161,**Lue 162, Arg 163, **Ile 164,** Glu 165, Lys 166
S3 **Core excluding H1,H2**	**Val 29, Phe 31, Val 33, Ile 45, Leu 47, Phe 48, Tyr 71, Phe 76, Val 79, Ile 85, Leu 120, Met 122, Phe 134, Phe 151, Val 179, Ile 181**

To study the disposition of the helices with respect to the barrel, (upon raising temperature in an MD simulation; See **Materials & Methods,** Section 2.6) the contacts (**Table 6**) between S1 - S3 and S2 – S3 were first calculated from the crystal structure (2HAQ). 4.0 Å was considered to be the distance cutoff for two atoms to be in contact.

Table 6: List of native contacts between Helix1 (H1) and Core and Helix2 (H2) and Core. For definition of S1, S2 and S3 see Table 5.

Native Contacts Between residues in Helix H1 and Core							Native Contacts Between residues in Helix H2 and Core						
S1			S3			Distance (Å)	S2			S3			Distance (Å)
N	ALA	52	O	PHE	48	3.037	CG1	VAL	161	CD1	ILE	85	3.994
CA	ALA	52	O	PHE	48	3.855	CG2	ILE	164	CG2	VAL	79	3.959
C	ALA	52	O	PHE	48	3.973	CG2	ILE	164	CG2	ILE	85	3.983
CG2	THR	55	CG	MET	122	3.588	CG2	ILE	164	CG1	VAL	179	3.977
CG2	THR	55	CB	PHE	151	3.724	CD1	ILE	164	CG1	VAL	179	3.978
CG2	THR	55	CG	PHE	151	3.677	O	ILE	164	CG1	VAL	79	3.798
CG2	THR	55	CD1	PHE	151	3.739	CA	GLU	165	CG1	VAL	79	3.781
CG2	THR	56	CB	LEU	47	3.848	C	GLU	165	CG1	VAL	79	3.903
CG2	THR	56	N	PHE	48	3.686	O	GLU	165	CG1	VAL	79	3.433
CG2	THR	56	C	PHE	48	3.511							
CG2	THR	56	O	PHE	48	3.378							
CG2	THR	56	O	PHE	151	3.573							
O	THR	56	CD1	LEU	47	3.729							
CB	PHE	59	CE2	PHE	151	3.848							
CG	PHE	59	CE2	PHE	151	3.515							
CD1	PHE	59	CE1	TYR	71	3.577							
CD1	PHE	59	CZ	TYR	71	3.925							
CD1	PHE	59	OH	TYR	71	3.433							
CD1	PHE	59	CZ	PHE	151	3.641							
CD1	PHE	59	CE2	PHE	151	3.532							
CE1	PHE	59	CE1	TYR	71	3.500							
CE1	PHE	59	CZ	TYR	71	3.879							
CE1	PHE	59	OH	TYR	71	3.691							
CE1	PHE	59	CG	PHE	134	3.833							
CE1	PHE	59	CD1	PHE	134	3.722							
CE1	PHE	59	CE1	PHE	134	3.968							
CE1	PHE	59	CZ	PHE	151	3.911							
CZ	PHE	59	CZ	PHE	31	3.594							
CZ	PHE	59	CD1	LEU	120	3.989							
CD2	PHE	59	CD2	LEU	47	3.791							
CD2	PHE	59	CE2	PHE	151	3.972							
CE2	PHE	59	CG2	VAL	29	3.747							
CB	LEU	62	CD1	TYR	71	3.896							
CB	LEU	62	CE1	TYR	71	3.612							
CB	LEU	62	CZ	TYR	71	3.945							
CD1	LEU	62	CZ	TYR	71	3.749							
CD1	LEU	62	OH	TYR	71	3.478							
C	LEU	62	CD1	TYR	71	3.995							
O	LEU	62	N	TYR	71	3.002							
SG	CYS	63	CB	VAL	29	3.851							
SG	CYS	63	CG2	VAL	29	3.943							
SG	CYS	63	CE1	PHE	31	3.831							
SG	CYS	63	CD1	TYR	71	3.910							

The fractional conservation of these native contacts (between the helices and the remaining residues of the core) were then estimated for all snapshots and averaged over 50 ns for each simulation run at temperatures 310, 400 and 450 K (**Table 7**). In addition (average) fractional conservation of secondary structural content (with respect to the native crystal structure) was also calculated for each simulation (50 ns).

Table 7

The fraction of native contacts conserved between the residues constituting the helices (H1, H2) and the remaining residues in the core, averaged over 50 ns for each MD simulation run at 310, 400 and 450 K respectively. The standard deviation is given in parentheses. The fractional conservation of secondary structural content with respect to the native crystal structure is also given.

Temperature (K)	Fraction of Native Contacts Conserved		Fraction of Secondary Structural Content		
	Helix H1-core	Helix H2 - core	Helix H1	Helix H2	Beta Strands
310	0.73(0.16)	0.65(0.23)	0.92(0.02)	0.89(0.01)	0.97(0.02)
400	0.65(0.18)	0.47(0.29)	0.92(0.03)	0.60(0.27)	0.86(0.05)
450	0.48(0.21)	0.37(0.30)	0.70(0.28)	0.51(0.29)	0.81(0.12)

On an average, about 0.65 - 0.75 of the native contacts between helices and core, were conserved in the simulation at 310 K. However, helix H2 appeared to be relatively loosely bound to the core compared to H1, with a sharper decline in the conserved fraction upon elevation of temperature to 400 K (H1 – 0.65(0.18); H2 - 0.47(0.29)). At 450 K, 0.48, 0.37 of the native

contacts were conserved for H1, H2 respectively. Again, H2 also exhibited an increased tendency to unwind as only 40% of its constituent residues (as found in the crystal structure) retained their helical conformation even by 400 K. In general, the strands constituting the barrel appear to have greater stability (0.81 secondary structural content conserved, **Table 7**), in contrast to both the helices. Thus, the simulations appeared to indicate the tendency of both helices to gradually unwind and adopt non – native geometries with respect to the core, quite early in the unfolding process.

2.4. Conclusion

The work reported in this chapter characterizes the equilibrium unfolding of cyclophilin from *leishmania donovani* (LdCyp) mediated by the denaturant GdmCl. Initial evidence for the existence of intermediate states appears with regard to the non-superposition of the unfolding curves monitored by fluorescence and CD. Analysis of the near and far UV CD spectra indicates the pronounced disruption of the native tertiary contacts at 1.2M GdmCl while conserving significant fraction of the native secondary structural content, characteristic of a molten globule [**22-34**]. This was further confirmed by hydrophobic dye-binding studies utilizing bis-ANS where a prominent peak was observed at 1.2M GdmCl. Thermal characterization of the protein incubated with 0.2, 0.6 and 1.0 M of GdmCl by DSC exhibited a gradual shift from a two (involving two T_m's) to a single step transition (with a unique T_m) .

Examination of the native crystal structure and data from MD simulations leads us to hypothesize that the possible equilibrium intermediate could involve disjuncture of the helices from their native geometry with respect to the barrel/core during initial stages of the unfolding process.

References

1. M. Ashutosh, S. Sundar, N. Goyal, (2007), *J. Infect. Dis*. 56, 143–153. doi:10.1099/jmm.0.46841-0.

2. D. Sereno, P. Holzmuller, J.L. Lemesre, (2000), *Acta Trop*. 74, 25-31.

3. J. Bua, L. E. Fichera, A. G. Fuchs, M. Potenza, M. Dubin, R.O. Wenger, G. Morretti, C.M. Scabone, A. M. Ruiz, (2008), *Parasitol*. 135, 217–228.

4. M. E. Perkins, T.W. Wu, S. M. Le Blancq, (1998), Antimicrob. Agents. Chemother. 42, 843–848.

5. P. Wang, J. Heitman, (2005), Genome.Biol. 6, 226-231.

6. H. Du , L. Guo, F. Fang , D. Chen , A.A Sosunov, G.M. McKhann, Y. Yan, C. Wang , H. Zhang , J.D. Molkentin, F.J. Gunn-Moore, J.P. Vonsattel, O. Arancio, J.X. Chen, S.D. Yan, (2008), *Nat. Med*. 14, 1097-1105.

7. Z. Keckesova, L.M. Ylinen, G.J. Towers (2006), *J. Virol.* 80, 4683-4690.

8. J. Kallen, V. Mikol, P. Taylor, M.D. Walkinshaw, (1998), *J. Mol. Biol*. 283, 435-449.

9. H. Ke, *J Mol Biol*. (1992), 228, 539-550**.**

10. V. Venugopal, B. Sen, A.K. Datta , R. Banerjee, (2007), *Acta. Cryst. F*, 63, 60-64.

11. D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. De Bolt , D. Ferguson, G. Siebel, P. Kollman, (1995), *Comput. Phys. Commun*. 91, 1-41.

12. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L . Kale, K. Schulten(2005), *J. Comput. Chem*. 26 (16), 1781-1802.

13. J.P. Ryckaert, G. Ciccotti, H. J. C. Berendsen (1977), *J. Comp. Phys*. 23 , 327-341.

14. W. Humphrey, A. Dalke , K.Schulten, *J Molec Graphics* (1996) "VMD - Visual Molecular Dynamics" 14: 33-38 (http://www.ks.uiuc.edu/Research/vmd/).

15. D. Frishman, P. Argos, (1995), *Proteins: Structure, Function and Genetics*, 23, 566-579.

16. A.H. Wani, J.B. Udgaonkar (2009), *J. Mol. Biol*. 387, 348–362.

17. A.K. Mandal, S. Samaddar, R. Banerjee, S. Lahiri, A. Bhattacharyya, S. Roy, (2003), *J. Biol. Chem*. 278 , 36077-84.

18. R.F Latypov, H. Cheng , N. A. Roder , J. Zhang , H. Roder, (2006), *J. Mol. Biol.* 357, 1009-1025.

19. S. Lindhoud, A. H. Westphal, J. W. Borst, C.R.M van Mierlo, (2012), PLoS ONE. 7(10) doi: 10.1371/journal.pone.0045746.

20. L. Zhu, Y. X Fan, J.M Zhou (2001), *Biochim. Biophys. Acta*. 1544, 320-332.

21. G.V. Semisotnov, N.A. Rodionova, O.I. Razgulyaev, V.N. Uversky, A.F. Gripas, R.I. Gilmanshin, (1991) *Biopolymers*, 31, 119-128.

22. A.L. Fink, L.J. Calciano, Y. Goto, T. Kurotsu, D. R. Palleros (1994), *Biochemistry* 33, 12504-12511.

23. B.K. Das, T. Bhattacharyya, S. Roy (1995), *Biochemistry*. 34, 5242-5247.

24. R. L. Baldwin, C. Frieden, G. D. Rose (2010) *Proteins.* 78, 2725–2737.

25. D. Barrick, R.L. Baldwin (1993), *Prot. Sci.* 2, 869-876.

26. A. Iram, T. Alam, J.M. Khan, T.A. Khan, R. H. Khan, A. Naeem. (2013), *PLoS ONE*. 8(8), doi: 10.1371/journal.pone.0072075.

27. K. Kuwajima, (1989), *Proteins*. 6, 87-103.

28. K. Kuwajima, (1996) *Faseb. J.*, 10, 102-109.

29. M. Ferrer, G. Barany, C. Woodward, (1995) *Nat. Struct. Biol*. 2, 211-217.

30. H. Christensen, R. H. Pain, (1991), *Eur. Biophys. J.* 19, 221-229.

31. S.H. Park, (2004) *J. Biochem. Mol. Biol*. 37, 676-683.

32. S. Sheshadri, G.M. Lingaraju, R. Vardarajan, (1999) *Prot. Sci.* 8, 1689-1695.

33. O.B. Ptitsyn, V.N. Uversky, (1994) *Febs. Lett.* 341, 15-18.

34. S. K. Jha, J. B. Udgaonkar (2009) *Proc.Natl. Acad. Sci. U.S.A.* 106, 12289 –12294.

Chapter 3

Application of network based metrics to study unfolding patterns in cyclophilin by molecular dynamics simulation

3.1 Introduction

Currently, there is considerable interest in the role of molten globules (MG) during the process of protein folding/unfolding **[1].** Two distinct classes of MGs have been distinguished, the dry (DMG) and wet (WMG) molten globules, both forms preserving native–like secondary structural elements, topology and dimensions; with perturbed tertiary contacts **[2-5].** The degree of perturbation in the native tertiary structure of the protein distinguishes DMGs from WMGs **[6].** In DMGs, though there is dislocation in the tight packing of residues within the molecular interior, the consequent fraying of secondary structural elements and molecular expansion is not sufficient to allow for the solvation of the hydrophobic core. On the contrary, water enters into the interior of the protein in case of a WMG **[7-9]**.

Experimentally, DMGs can be unambiguously distinguished from WMGs, by the inability of the fluorescent dye ANS to penetrate and bind to the core of a DMG and also the protection of its backbone NH hydrogens from exchange **[10]**. Another issue with regard to DMGs is their relationship to the highly unstable transition state of unfolding**.** In an earlier theoretical study **[16]**, the DMG was confounded with the transition state species in a two – step unfolding process. The unlocking of packing contacts within the molecular interior in case of a DMG must be a very subtle effect, as in all probability, the dissociated residues continue to maintain close proximity to hinder the solvation of the core.

Previous studies with regard to the packing of amino acid side chains within proteins had established that residues which interdigitate with stereospecific geometry can be identified based on fairly stringent mathematical criteria **[22].** Briefly, the entire polypeptide chain of the molecule is considered to be one continuous surface rather than covalently linked discrete atoms. In such a representation the surface fit or surface complementarity (S_m) of two interacting residues can be estimated along with their respective surface patches (overlap, O_v) **[22]** buried upon association with each other.

In this chapter, intra – molecular close packing has been represented as a network with each node representing an amino acid side chain and their mutual association as links **[23-26].** As association between residues was defined in terms of (based on the surface representation of the protein) S_m (≥ 0.40) and Ov (≥ 0.08), the resulting graphs were expected to consist primarily of those interactions which strongly condition inter-residue geometry, rather resembling a three dimensional jigsaw puzzle **[22].** In order to test these ideas, thermal unfolding MD simulations have been performed on cyclophilin from *L. donovani* (LdCyp) as a model system **(Figure 1).** Cyclophilins (Cyps) are a ubiquitous class of peptidyl prolyl cis-trans isomerases, also known to be the intracellular receptor for the immunosuppressive drug cyclosporine A (CsA) regularly used in organ transplants **[29].** LdCyp is composed of a beta barrel encompassing the unique hydrophobic core of the molecule and is flanked by two helices (H1, H2) on either end of the barrel **[30].** The work reported in this chapter analyzes the time evolution of packing networks (SCN) in the unfolding trajectory of LdCyp to probe temporal patterns in the disjuncture of specific packing interactions within a protein.

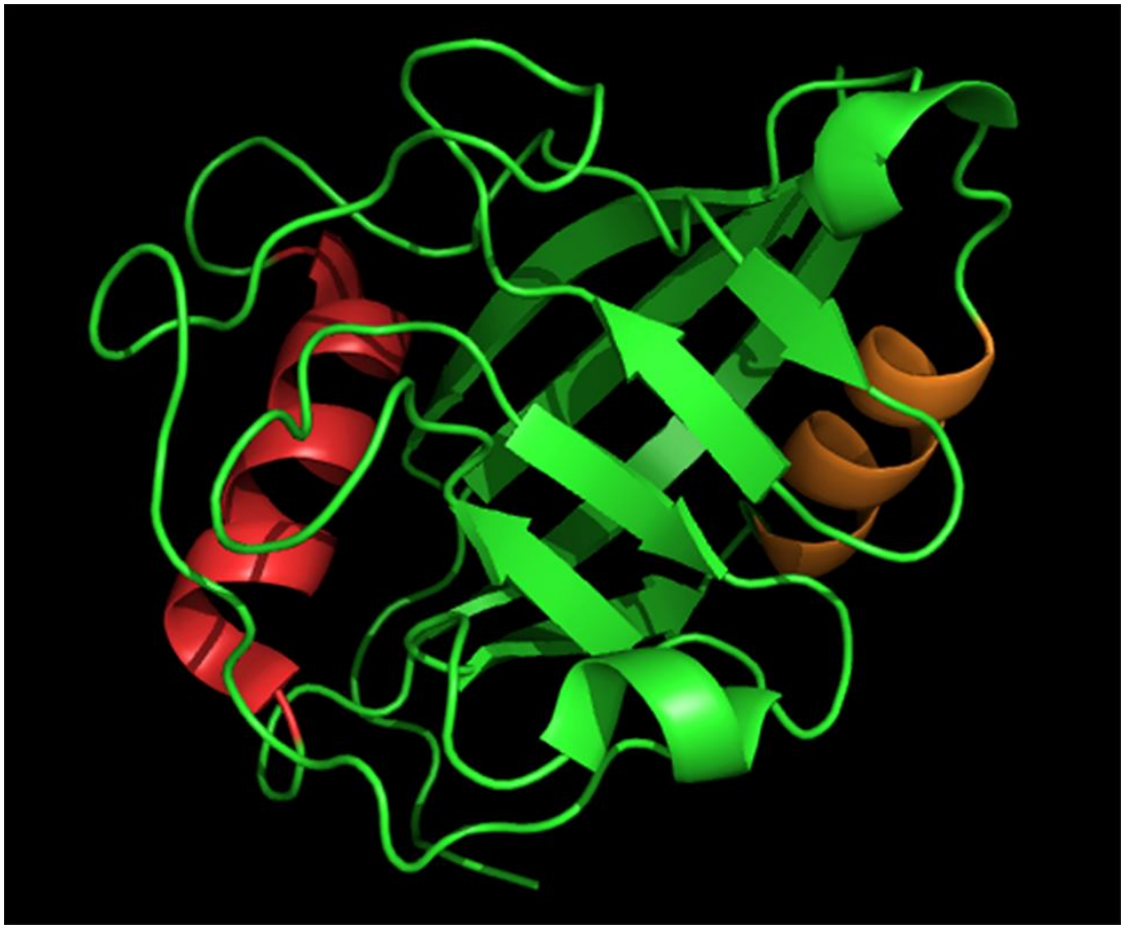


Figure 1: Crystal structure of cyclophilin from *Leishmania donovani* (LdCyp). The helices H1 and H2 are indicated in red and orange respectively (figure generated using Pymol [40])

3.2 Materials and Methods

3.2.1 Molecular Dynamics Simulations

Unfolding molecular dynamics simulations were performed on cyclophilin (coordinates obtained from: 2HAQ) at temperatures 310, 400, 450 and 500K. Each simulation run was for a minimum duration of 50ns and was repeated five times for 500 K, four times for 450 K, and twice for 400

and 310 K with different initial random number seeds. The simulation at 310K could be considered to be 'native'. Initially, the protein was solvated in a cuboidal box of dimensions 78.683, 68.897 and 78.137 Å by the addition of 11088 waters following the TIP3P model and charge neutralization of the system was accomplished by the addition of Na^+ ion utilizing the xleap module of AMBER **[32].** The structure was then energy minimized (AMBER 2002 force field) for 200 steps of steepest descent followed by 19800 steps of ABNR incorporated in the SANDER module. All simulations were performed using NAMD 2.0 **[33].** For each 50ns run the targeted temperature was attained in increments of 10K/ps for a NPT ensemble with the Langevin piston set to the targeted temperature (310,400,450,500 etc.) and the pressure fixed at 1.032 bar. The bond lengths were constrained at a tolerance level of 0.005 Å by the SHAKE algorithm **[34].** Visual molecular dynamics program (VMD 1.9.1) **[35]** was used to view the trajectories obtained from NAMD and also to extract the coordinates for further downstream computations. The time step (δt) for the velocity - verlet algorithm was kept fixed at 1fs.

3.2.2 Surface Generation, Surface Complementarity & Overlap

The van der Waals surface for the entire polypeptide chain was generated, sampled at 10 dots/ $Å^2$. The atomic radii were taken from the all atom molecular mechanics force field **[36]**. The details regarding surface generation have been described in several earlier publications **[22-23].** The entire surface of the polypeptide chain was sampled as an array of discrete dot surface points (10 dots/$Å^2$) where each dot point is an area element characterized by its location(x, y, z) and the direction cosines of its normal (dl, dm, dn). Generally, residues within proteins will participate in a 'network' of interactions with its neighboring amino acids. The entire surface of the side chain can thus be partitioned into sub-surfaces, such that each sub-surface is in close association with the corresponding surface patch of one of its neighbors. Thus, the (previously mentioned)

quantities surface complementarity (S_m) and overlap (O_v) can be used to characterize the association between two side chain surfaces of interacting amino acids **[22].**

Briefly, consider a residue **A**, which consists of a total of N_A side chain surface points. For every side-chain dot point of **A**, its nearest neighbors (within 3.5Å)was identified from the surrounding dot points contributed by the rest of the polypeptide chain (inclusive of both side and main chain atoms). Then, the following expression was computed**:**

$$S_{ab} = \mathbf{n_a}.\mathbf{n_b} \exp(-wd_{ab}^2 \quad (1)$$

where $\mathbf{n_a}$, $\mathbf{n_b}$ are the unit normals of a dot point '**a**' (on residue **A**) and those of its nearest neighbor '**b**' (from residue **B**) respectively, d_{ab} is the distance between them and w is a scaling factor set to 0.5. Thus, for the set of points N_{AB}, (dot points on **A** which have found points on the side chain **B** as nearest neighbors) median of distribution $\{S_{ab}\}$ is defined as the surface complementarity $S_m^{A\rightarrow B}$.

The corresponding overlap between residues A and B is defined as,

$$O_v^{A\rightarrow B} = N_{AB}/N_A \quad (2)$$

Notably, the maximum possible value attained by both S_m and O_v are 1.00.

3.2.3 Protein Surface Contact Networks (SCN)

Generally, two point atoms are said to be in 'contact' based on purely distance criteria. Since the representation of atoms has been expanded to surfaces the criteria for two residues to be in contact based on S_m and O_v is as follows. S_m and O_v are generally non-commutative, that is $S_m^{A\rightarrow B}$ or $O_v^{A\rightarrow B}$ does not necessarily equal $S_m^{B\rightarrow A}$ or $O_v^{B\rightarrow A}$ respectively. Thus, two residues **A**, **B** are said to be in 'contact' when both $S_m^{A\rightarrow B}$, $S_m^{B\rightarrow A}$ and $O_v^{A\rightarrow B}$, $O_v^{B\rightarrow A}$ simultaneously satisfy

the cutoffs (are greater than or equal to) 0.4 and 0.08 respectively. These cutoffs were chosen on the basis of a previous work **[22]** which showed that these criteria (in S_m and O_v) led to highly constrained inter-residue geometries for interacting binary pairs of amino acids. In the context of protein surface contact networks (SCN), a node stands for an amino acid side chain and two nodes are connected by an edge, when they are in surface contact (as define above). Henceforth, these networks will be referred to as protein 'Surface Contact Networks' (SCN). Like all other networks SCN's can be represented in terms of adjacency matrices with elements (a_{ij}) equal to one when two nodes are connected by an edge and zero otherwise **[23].**

3.2.4 Metrics defined on Surface Contact Networks: Disnet, Dlf, Persistence and Persf

SCN's were constructed based on the coordinates of LdCyp (snapshots) at various stages of the unfolding simulation at different temperatures. The SCN for any snapshot was represented as an adjacency matrix, where the elements of the matrix a_{ij} could assume values either 0 or 1 (1 denoting the presence of an interaction between residues i and j as defined above). Only residues constituting the hydrophobic core of the protein were considered in the construction of the SCN adjacency matrices. The distance between any two such adjacency matrices was determined by counting the number of links present in one and absent in other, and normalizing by the dimension of the adjacency matrix (n x n).

$$Disnet(A, A`) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left| A_{ij} - A`_{ij} \right|}{(nxn)} \qquad (3)$$

Where A_{ij} and A'_{ij} are the matrix elements of two adjacency matrices (A, A') corresponding to two SCN's (from their corresponding snapshots) and (n x n) is the dimension of the adjacency matrix.

Based on Disnet, two more related measures, 'dlf' and 'persf' were defined in order to assess the degree of stability or fluctuations in the SCNs. As has been mentioned previously, each simulation run at a specific temperature (310 K, 400K etc.) was for a total of 50 ns. Discarding the first 2 ns, every simulation was partitioned into contiguous intervals of 2ns (henceforth referred to as epochs). Adjacency matrices for the SCNs were constructed pertaining to snapshots sampled every 10 ps, leading to 200 snapshots (SCN adjacency matrices) per epoch. Within an epoch the 'persistence' of a contact between residues i and j was defined as the number of times the surface contact was detected in the snapshots (spanning the epoch) divided by the total number (200) of snapshots. Thus

$$P_{ij} = \sum_{k=1}^{n} A_{ij}/n \qquad (4)$$

where, A_{ij} are elements of the SCN adjacency matrices, the summation being over snapshots (k) and n equals 200. Now a new matrix (P') with a dimension identical to that of SCN adjacency matrices was defined such that ($P'_{ij} = 1$ when $P_{ij} \geq 0.25$) and ($P'_{ij} = 0$ when $P_{ij} < 0.25$). The new matrix P' (constructed applying a predetermined threshold :0.25), referred to as the 'persistence matrix', contains information regarding the relatively stable regions of the network, as only stable surface contacts detected in more than 25 % of the snapshots in an epoch contribute to the adjacency matrix (having been set to unity). The metric Disnet could again be used to estimate the 'distance' between two persistence matrices and in this instance will be referred to as 'persf'. A persistence matrix derived from the entire 2-50 ns block at 310 K (native) was considered to

be the baseline and 'persf' (unless stated otherwise) is the 'distance' (Disnet) between this matrix and persistence matrices derived from every epoch.

In order to estimate the degree of fluctuations associated with SCNs within an epoch, the distance Disnet was estimated for every pair of (SCN) adjacency matrices associated with the snapshots spanning an epoch. Then somewhat analogous to the measure 'alf' **[21]**, the average <Disnet> over all such pairs ([200 x 199]/2!) was defined as 'dlf' and epochs with higher dlf values were expected to be regions of enhanced network volatility.

3.2.5 Solvent Accessible Surface & Burial

The solvent accessibility of amino acid residues in the protein molecule was estimated by the ratio of the solvent accessible areas (SAA : probe radius 1.4 Å) of the residue (X) in the polypeptide chain to that of an identical residue in a Gly – X – Gly peptide fragment **[23**], in a fully extended conformation. Thus, burial (bur) was defined as:

Bur = SAA of residue in polypeptide chain / SAA of residue in Gly – X – Gly fragment

3.2.6 Secondary Structural Content

The secondary structural content for any snapshot was estimated by STRIDE **[37]** and the status of residues constituting the helices (H1,H2) and the 8 ß strands composing the barrel (as found in the native crystal structure, pdb id: 2HAQ) was monitored. A total of 79 residues out of 166 residues of the protein made up the total secondary structural content of native LdCyP crystal

structure (with 57 residues in the ß barrel; H1 : 13 residues , H2 : 9 residues). Thus, fractional secondary structural content for any snapshot was defined as the fraction of these 79 residues found to reside in their native secondary structural elements, either helices or strands. Average fractional secondary structural content within an epoch of 2ns (SSC_{epoch} was determined by averaging fractional secondary structural content over all the snapshots (200) within the epoch.

3.2.7 Classical Multi-Dimensional Scaling

Methods of classical multi-dimensional scaling were adopted to map 'distances' between snapshots based on Disnet to three dimensional coordinates. Initially, the entire 50ns simulation at a specific temperature was sampled in 100 ps intervals leading to 500 snapshots in all. Computing the 'distance' (w.r.t. the metric Disnet) for every pair of snapshots led to the construction of a 500 X 500 matrix (D: elements d_{ij}), where the matrix element d_{ij} gives the metric (Disnet) distance between the i^{th} and j^{th} snapshots. The calculation was also repeated at 200 ps sampling leading to a 250 x 250 matrix. Then following Borg and Groenen **[38]** the eigenvalues and eigen vectors were computed for the matrix

$$B = -1/2\ J\ P^{(2)} J \qquad (5)$$

Where for P^2, every element $P_{ij} = D_{ij}^2$ and $J = I - 1/n\ [I\ I']$ where 'I' is a n x n unit matrix, 'n' is the number of snapshots and [I I'] is a n x n matrix where every element is unity. The eigenvectors corresponding to the three largest eigenvalues were selected and the three dimensional coordinates for each snapshot was obtained upon suitably combining the eigenvectors subsequent to multiplying them by the square root of their respective eigenvalues.

3.3 Results and Discussion

Cyclophilin from *L. donovani* (LdCyp) is a single domain protein (crystal structure: 2HAQ at 1.97Å) consisting of an 8 stranded β-barrel with two helices located at either end **(Figure 1).** The location of the two helices (designated H1 and H2) with respect to the barrel effectively blocks the solvent accessibility of the only hydrophobic core of the molecule, located in the interior of the barrel. The extended core of LdCyp is composed of a total of 24 residues, of which 8 are contributed by flanking helices (H1, H2), while the rest predominantly come from the strands encompassing the core. Thus the full set of surface contacts involving the residues of the hydrophobic core can be partitioned into three subsets, S1: contacts between residues located on the ß strands, S2: between helix H1 and the remaining residues of the core and S3: between H2 and the remaining core residues **(Table1).**

Table 1: Set of contacts of S1, S2 and S3, with the secondary structural elements in parentheses, S-Strands, H1-helix1, H2-helix2 and L-loops.

S1	S2	S3
120LEU(S)-151PHE(S)	59 PHE(H1)-151PHE(S)	85 PHE(S)-164PHE(H2)
29VAL(S)-47LEU(S)	62 LEU(H1)-71 TYR(L)	85 ILE(S)-161VAL(H2)
122 MET(S)-151PHE(S)	29VAL(S) -63CYS(H1)	164 ILE(H2)-179VAL(S)
31PHE (S)-181ILE(S)	55THR(H1)-151PHE(S)	
76 PHE(L)-85 ILE(S)	59 PHE (H1)-71TYR(L)	
31PHE(S)-45ILE(S)	55THR(H1)-122MET(S)	
76 PHE(L)-181 ILE(S)	47LEU(S)- 59PHE(H1)	
71 TYR(L) -134 PHE(S)	59PHE(H1)-134PHE(S)	
31PHE (S)-134PHE(S)	31PHE(S)-59PHE(H1)	
31PHE (S)-71TYR(S)		
33VAL(S -179VAL(S		

3.3.1 Cα RMSD

Each simulation run (for 50ns blocks) was conducted 5,4,2,2 times at temperatures 500,450,400 and 310 K respectively (**Materials and Methods**). The first 2ns for every simulation was not considered for any further calculation. To start with the C^{α} RMSD of all the snapshots (w.r.t 2HAQ) were plotted (with respect to time) for all the simulations using CHARMM **[39].** The simulation at 310K can be considered native with an average C^{α} RMSD of 1.35 Å ± 0.12 (310KSIM1: from the 3rd to the 50th ns), exhibiting minimal fluctuations **(Figure 2).**

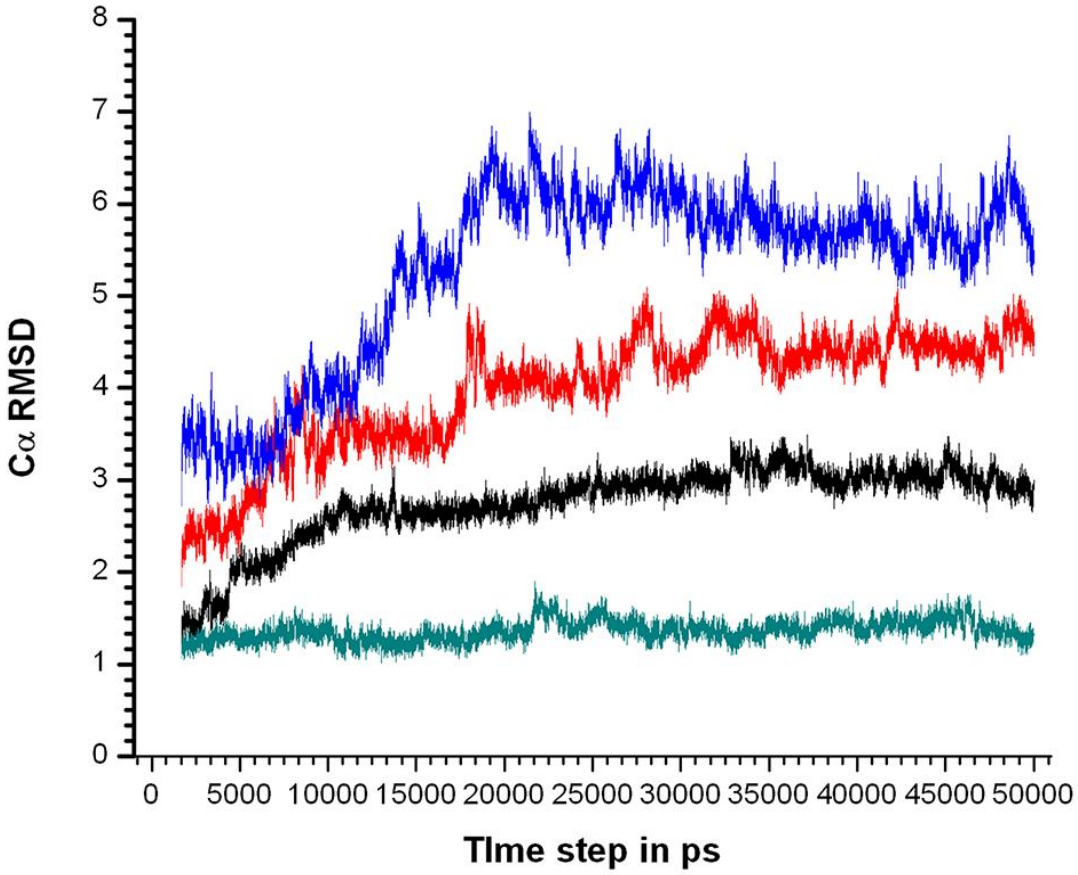


Figure 2: Cα RMSD between snapshots and the native crystal structure (2HAQ) plotted as a function of simulation time at different simulation temperatures indicated by the colors (—) 310K_SIM1 (—) 400K_SIM1 (—) 450K_SIM1 and (—) 500K_SIM1.

Two simulations at 310K (310K_SIM1 and 310K_SIM2) had more or less overlapping C^{α} RMSD's (**Figure 3**) with an average of 1.44 Å ± 0.11 for 310K_SIM2. At 400K_SIM1, an abrupt increase was observed around 4ns, subsequent to which the average RMSD increased to 2.76 ± 0.47 Å **(Figure 2)** (400K_SIM1 from 4-50ns). Careful visual examination of the snapshots around this region revealed the relative detachment of the helix H2 from its native position with respect to the barrel, which was responsible for the abrupt rise in RMSD. While in 400K_SIM2 the surge in C^{α} RMSD was observed as early as 10ns **(Figure 4**)

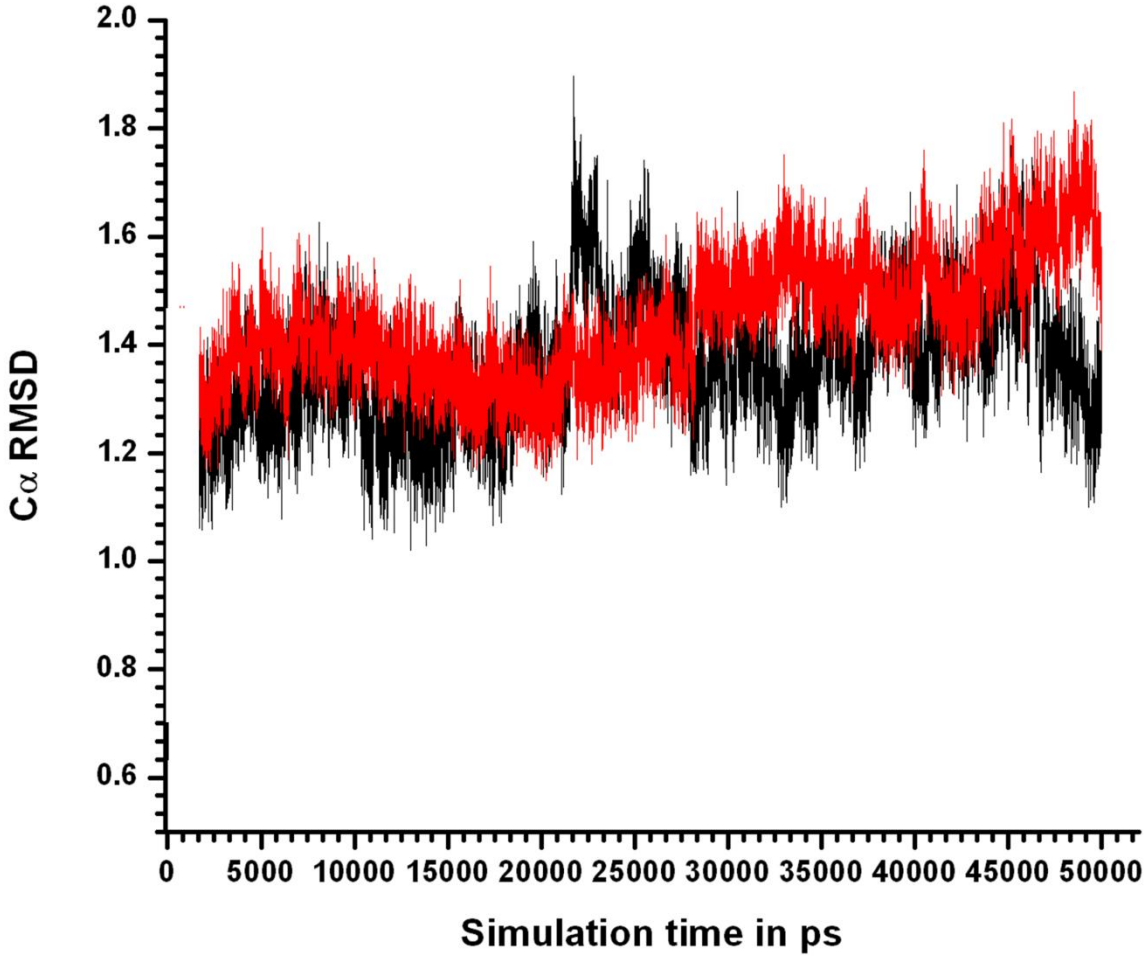


Figure 3: C^{α} RMSD between native simulation snapshots (—) 310K_SIM1, (—) 310K_SIM2 and the native crystal structure of LdCyp (2HAQ).

with slightly lower average of 2.19 ± 0.37 Å. A similar surge was observed around 15ns at 450K_SIM1 rising to 4.0-4.5 Å due to the increased disjuncture of both the helices from the body of the barrel. Both these events were reproducible in the other trajectories (450K_SIM2 – 450K_SIM4). Subsequent to this event at 450K there appeared to be a greater tendency for divergence amongst the other trajectories (450K_SIM1 - 450K_SIM4) at the same temperature (**Figure 5**).

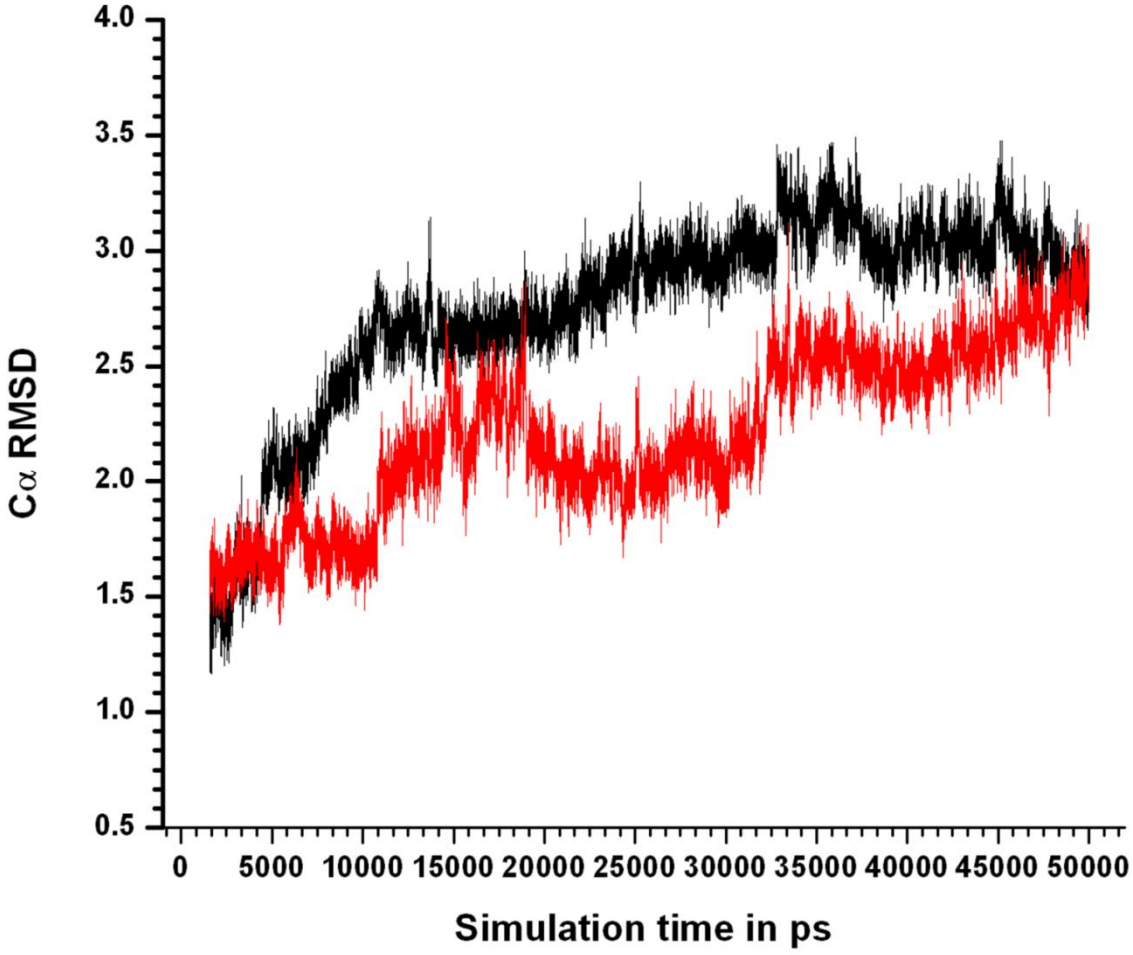


Figure 4: C^{α} RMSD between native simulation snapshots (—) 400K_SIM1, (—) 400K_SIM2 and the native crystal structure of LdCyp (2HAQ).

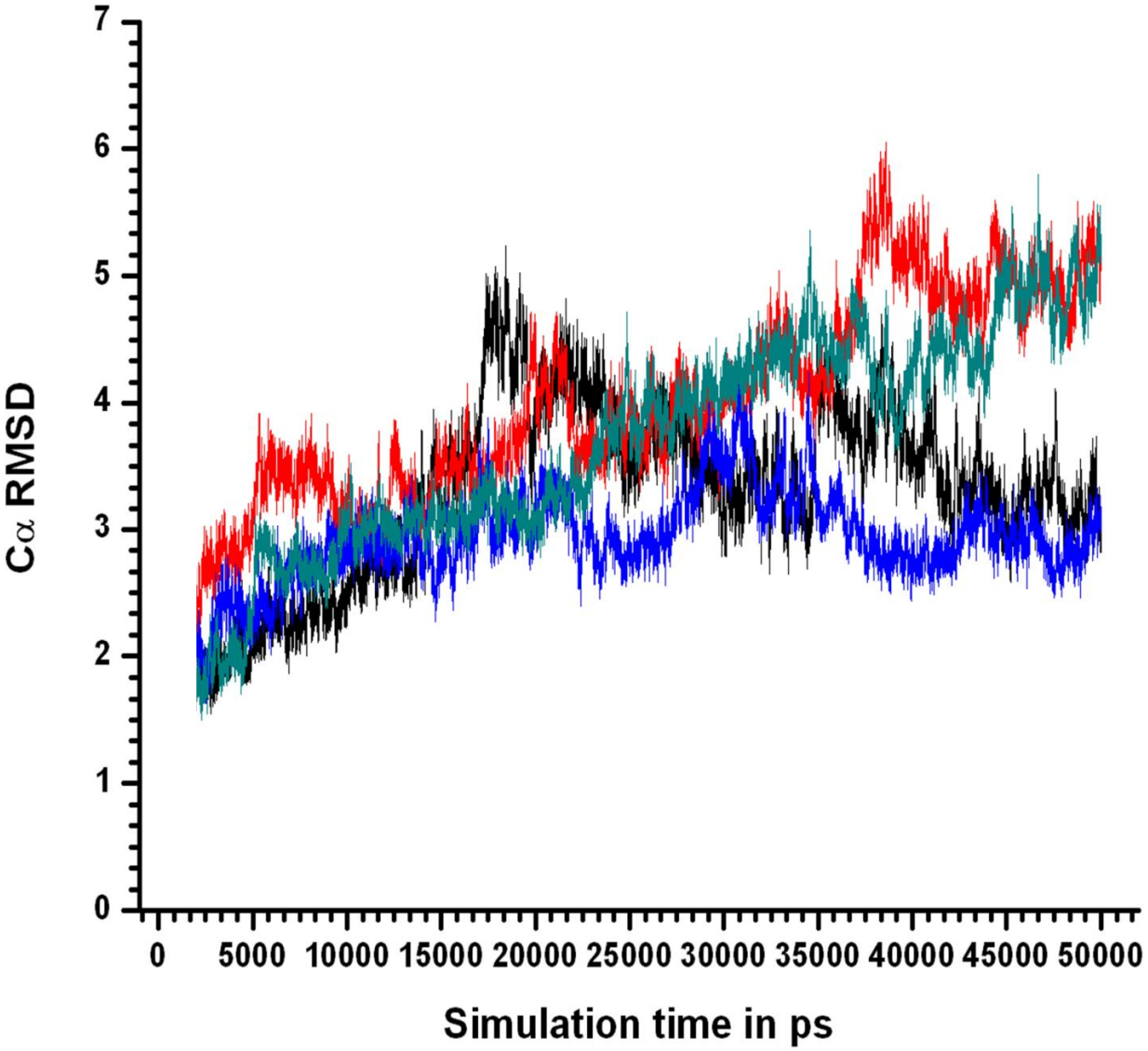


Figure 5: C^{α} RMSDs of snapshots for simulations at 450K a) 450K_SIM1 (—) b) 450K_SIM2 (—) c) 450K_SIM3 (—) and d) 450K_SIM4 (—).

At 500K, 4 trajectories tended to cluster together (500K_SIM1-500K_SIM4) with C^{α} RMSD in excess of 6 Å while the fifth 500K_SIM5 **(Figure 6)** diverged sharply after 25ns (C^{α} RMSD increasing to about 14 Å).

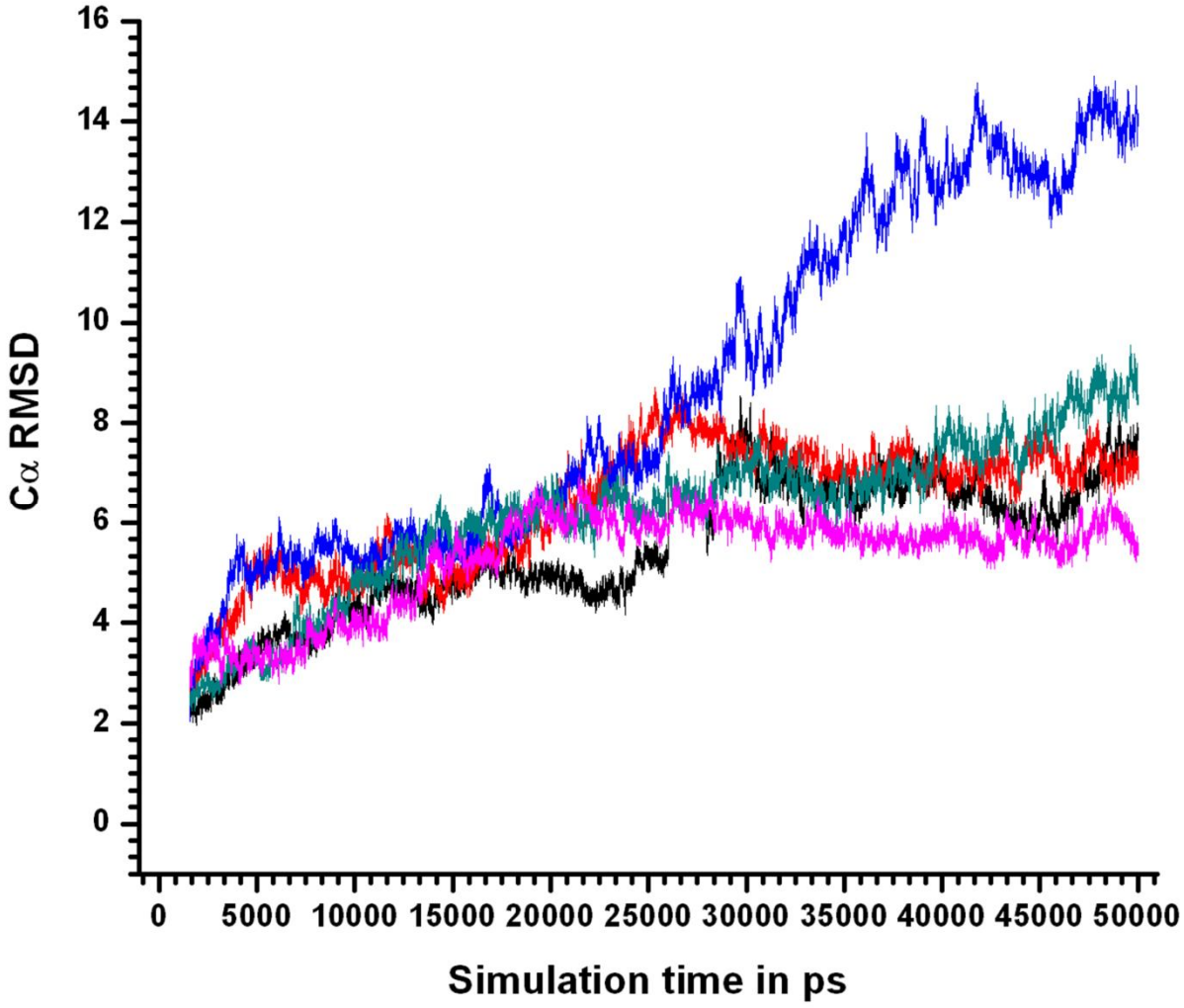


Figure 6: C^{α} RMSD of snapshots for simulations at 500K a) (—) 500K_SIM1 b) (—) 500K_SIM2 c) (—) 500K_SIM3 d) (—) 500K_SIM4 e) (—) 500K_SIM5.

However, all simulations at 500 K demonstrated the complete unraveling of the protein structure. By and large with the exception of the MD run at 500K, there was fairly good convergence between the trajectories at lower temperatures**.**

3.3.2 Disnet

It has been pointed out that C^{α} RMSD could be a misleading metric as peripheral distortions in the structure (for example in the loops) could lead to high RMSD values, even though little deviation is observed in the main body of the protein. Such is the case for cyclophilin, where the initial rise in C^{α} RMSD at 400K is primarily due to the movement of helix H2, rather than distortions in the main ß barrel of the protein. Network based metrics such as Disnet **(Materials and Methods),** wholly circumvent this problem. Disnet is the ratio of the uncommon links between two networks to the total number of possible links. For snapshots at an interval of 10 ps, surface contact networks **(SCN: Materials and Methods)** were constructed only considering residues constituting the hydrophobic core of the molecule. As has been mentioned previously, SCN's based on surface complementarity (S_m) and overlap (Ov) between side chain surfaces of interacting residues, contains implicit information with regard to the relative geometry of inter-digitating side chains and is an exceptionally sensitive indicator of subtle changes in side chain packing within proteins. The 'distances' (Disnet values) between SCN's derived from the snapshots and the native crystal structure were then studied for the entire course of the simulation. For every MD simulation run (2-50ns blocks)**,** Disnet was calculated between the SCN adjacency matrices of the snapshots (sampled every 10ps) and the SCN derived from the native crystal structure (2HAQ) and then averaged (<Disnet>) over the entire block. <Disnet> recorded a characteristic increase with rise in temperature (310K_SIM1, 400K_SIM1 etc, **Table**

2). At 310K_SIM1 <Disnet> was 0.44 (±0.07) which is to say, that on an average 56% of the links in the core were identical to those found in the native crystal structure **(Table 2).** Relative to the native simulation (at 310K), marginal difference was observed at 400K_SIM1 (0.49±0.08) and at 400K_SIM2 (0.47 ± 0.08) (**Figure 7**).

Table 2: <Disnet> values for different simulation temperatures for (310K_SIM1, 400K_SIM1, 450K_SIM1 and 500K_SIM1) calculated with native crystal structure (2HAQ) as baseline averaged over the entire simulation block of 50ns with standard deviations in parentheses.

Temperature (K)	<Disnet>
310K_SIM1	0.44(0.07)
400 K_SIM1	0.49(0.08)
450 K_SIM1	0.68(0.12)
500 K_SIM1	0.80(0.11)

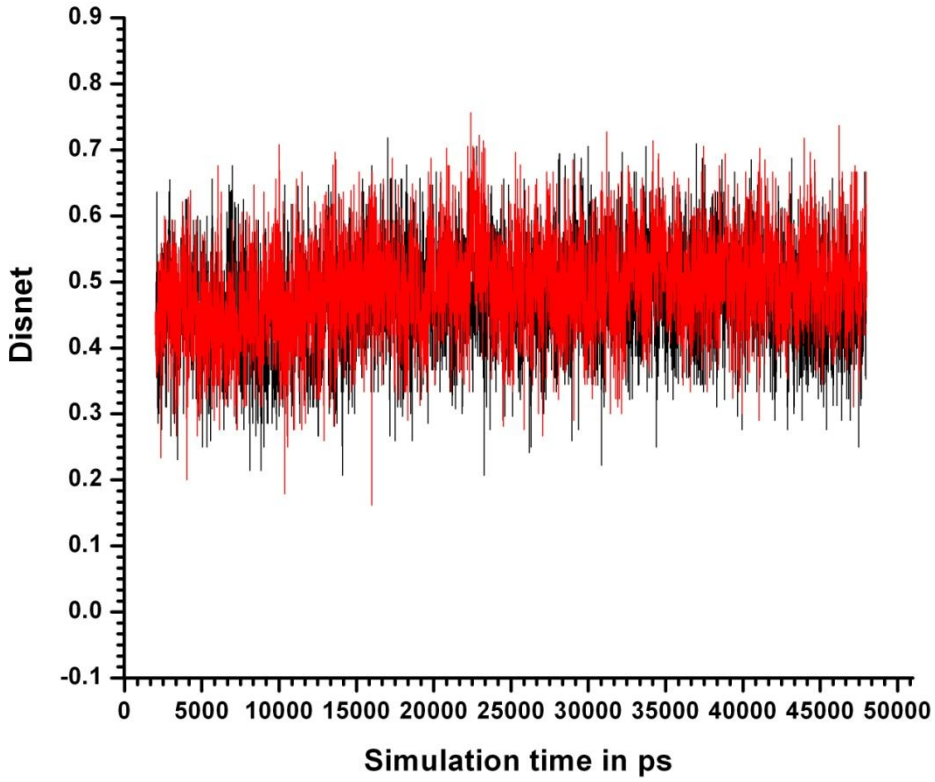


Figure 7: Disnet values for simulation set of 400K_SIM2 shown in solid black (—) lines and 400K_SIM1 in (—) lines

However, at 450K_SIM1 <Disnet> increased rather abruptly to 0.68 (±0.12) and further rose to 0.80 (±0.11) at 500K_SIM1 **(Table 3).**

Table 3: <Disnet> values for different simulation temperatures for (310_KSIM1, 400K_SIM1, 450K_SIM1 and 500K_SIM1) calculated with native crystal structure (2HAQ) as baseline averaged over the entire simulation block of 50ns with standard deviation in parentheses

Temperature(K)	<Disnet>
310K_SIM1	0.44(0.07)
400 K_SIM1	0.49(0.08)
450 K_SIM1	0.68(0.12)
500 K_SIM1	0.80(0.11)

At 450K_SIM1, Disnet values were stable up to ~23ns (**Figure 8**) even though somewhat elevated relative to 310 or 400K. However, subsequent to 23ns a sharp rise in Disnet was observed till about 30 ns (peaking at 28 ns); after which it again resumed a more stable progression though at a significantly higher value (about 0.5). The behavior of Disnet from 20 to 30ns exhibited a similar tendency in the other simulation trajectories at 450K **(Figure 9).**

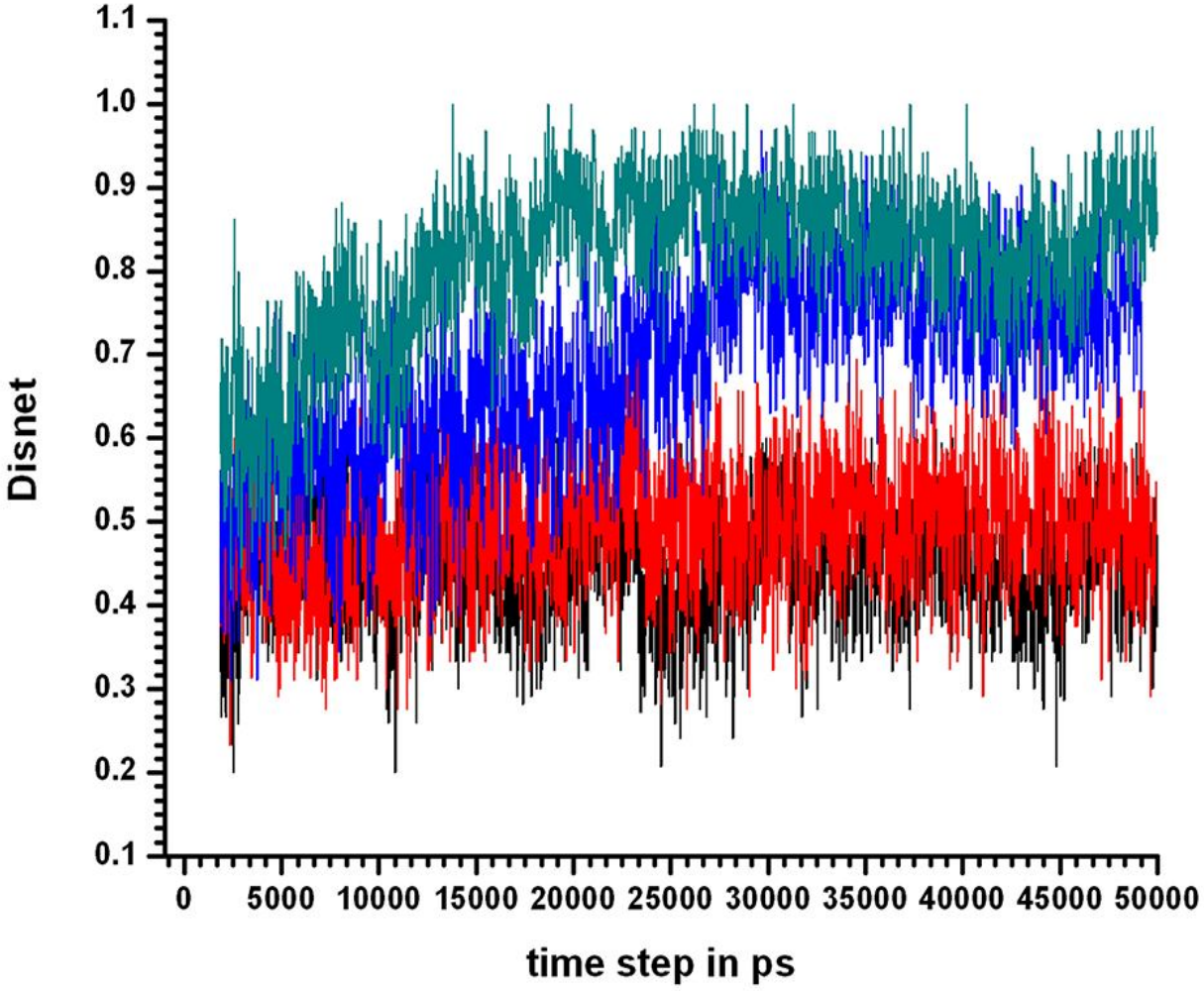


Figure 8: Disnet values between the Surface Contact Networks (SCNs) of the snapshots and the SCN derived from the crystal structure (2HAQ) plotted as a function of simulation time for the simulations 310K_SIM1 (—), 400K_SIM1 (—), 450K_SIM1 (—), 500K_SIM1 (—). Residues constituting the hydrophobic core of LdCyp alone were considered in the construction of the SCNs.

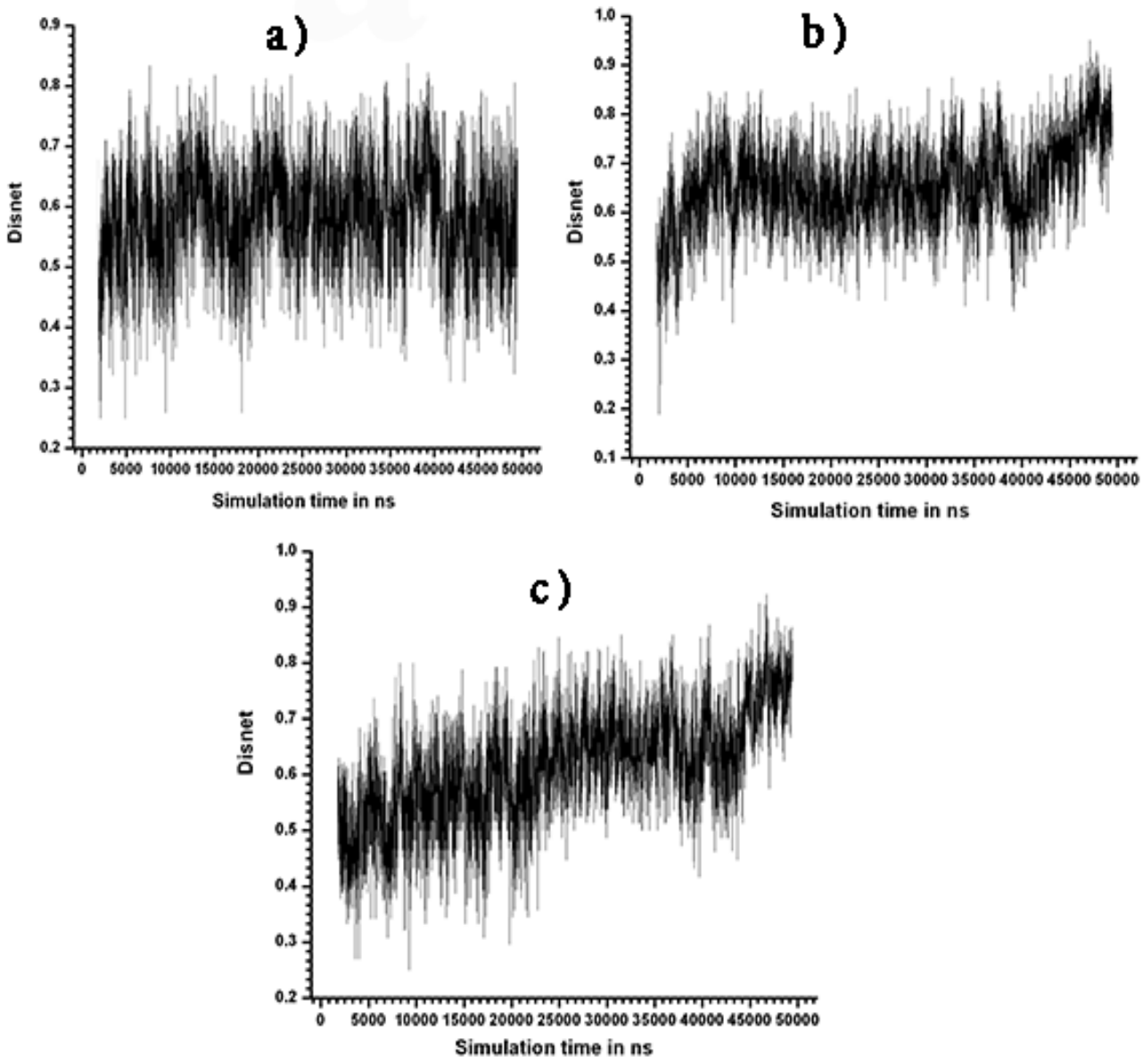


Fig 9: Disnet values for the hydrophobic core of LdCyp at a) 450K_SIM1 b) 450K_SIM2 and c) 450K_SIM3 with the native crystal structure of cyclophilin (2HAQ) as a baseline

3.3.3 Dlf and Persf

To estimate the degree of stability or volatility in the network two more measures ('persf' and 'dlf') were defined **(Materials & Methods).** These measures are based on the partitioning of each MD simulation run (from 2 – 50ns), into non – overlapping, contiguous 2ns intervals (referred to as epochs). Persf estimates the 'persistence' **(Materials and Methods)** of native contacts in snapshots constituting an epoch ; whereas dlf is simply the pairwise average of Disnet values between SCN adjacency matrices (spanning 2ns epochs), making it somewhat analogous to the previously reported measure 'alf' **[21]**. Both measures persf and dlf are expected to rise in the event of increased conformational flexibility in the structure, symptomatic of enhanced network instability**.**

For the native simulation at 310K_SIM1 (which in a sense serves as a baseline) the values for <persf> and <dlf> (averaged over all the epochs) were 0.15 (.03) and 0.39 (.01) respectively, which implies that for <persf>, on an average about 85 % of the contacts prevalent in the core of LdCyP 'persisted' throughout the course of the simulation at 310K_SIM1. At 500 K_SIM1, the same measure dropped to about 35%. Both <persf> and <dlf> exhibited regular and graduated increase as a function of temperature, increasing by about 0.1 – 0.2 for each temperature increment (**Table 4**).

Table 4: Dlf and persf values averaged over the entire simulation block (between 2 – 50ns) for the hydrophobic core of LdCyp at different simulation temperatures for (310K_SIM1,400K_SIM1,450K_SIM1 and 500K_SIM1). The standard deviations are given in parentheses.

Temperature (K)	<dlf>	<Persf>
310K_SIM1	0.39(0.01)	0.15(0.04)
400K_SIM1	0.49(0.02)	0.22(0.04)
450K_SIM1	0.65(0.05)	0.37(0.10)
500K_SIM1	0.76(0.07)	0.64(0.16)

Significant differences were observed in the time course of the two measures. Specifically, around 25 ns at 450 K_SIM1, a sharp surge in persf from about 0.20 to 0.55 was observed (**Figure 10**), preceded by a relatively unstable region beginning from 13 ns. A similar rise was observed in the other simulations at 450K (_SIM2-4), though not beginning from the same time point (**Figure 11)**

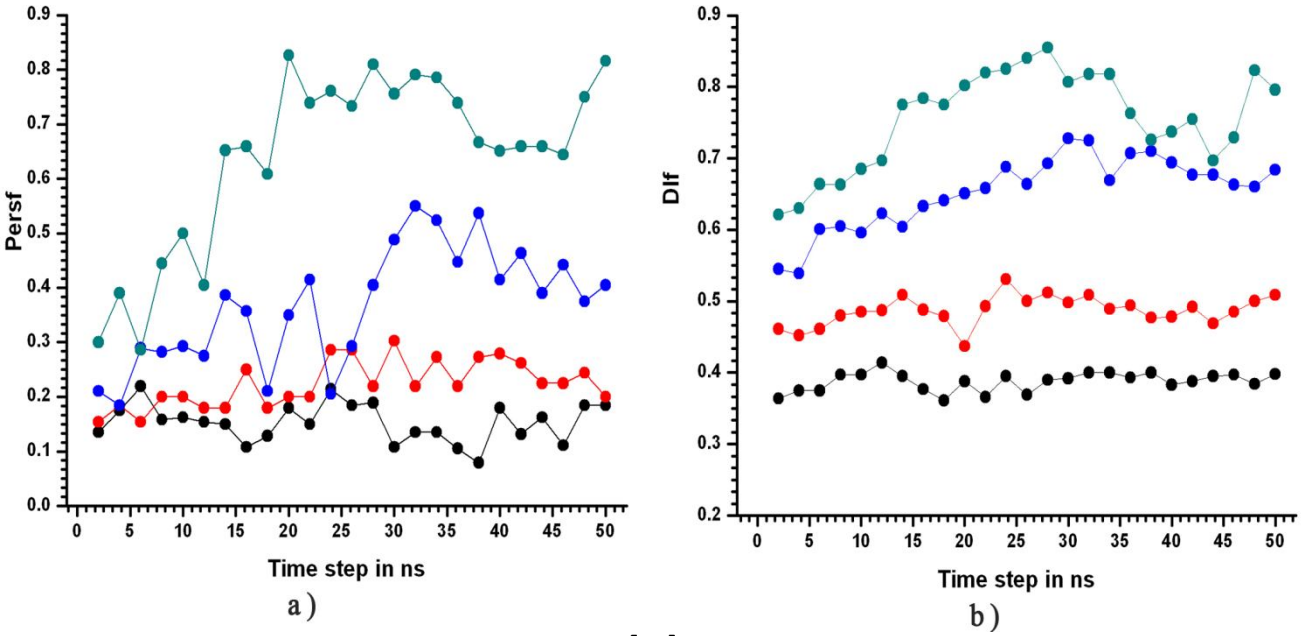


Figure10: a) Persf values for 2ns epochs indicated by filled circles considering the hydrophobic core of LdCyp as a function of simulation time with simulation sets of (310K_SIM1,400K_SIM1,450K_SIM1and500K_SIM1). The simulations at different temperatures are indicated by the colors : (—) 310K (—)400K (—)450K (—) 500K.

b) Dlf values for 2ns epoch indicated by filled circles considering the hydrophobic core of LdCyp ; plotted as a function of simulation time with simulation sets of (310K_SIM1,400K_SIM1,450K_SIM1 and 500K_SIM1). The simulations at different temperatures are indicated by the colors : (—) 310K (—)400K (—)450K (—) 500K.

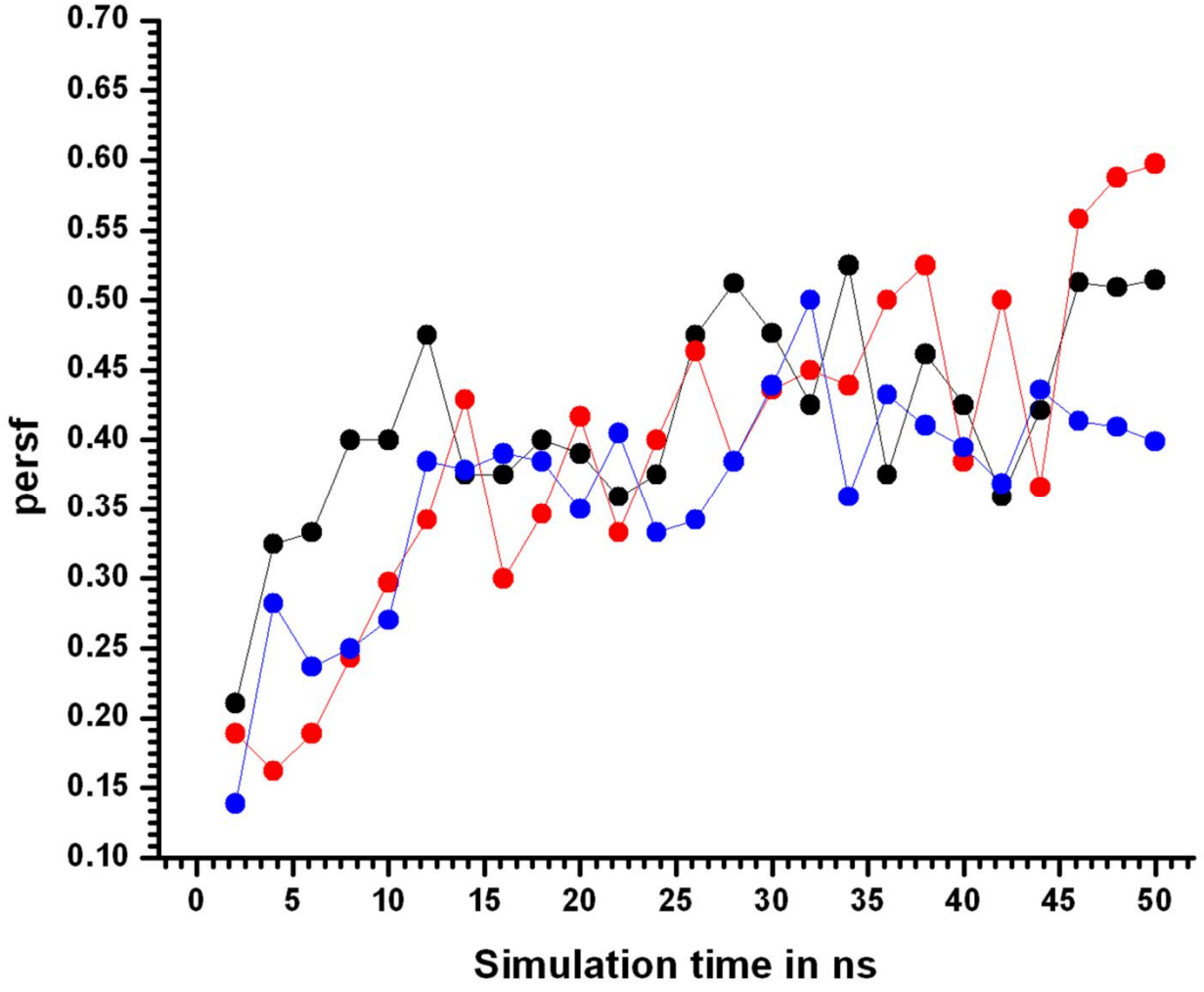


Figure11 Persf values for 450K simulation temperatures plotted as a function of simulation time a) (—) 450K_SIM2 b) (—) 450K_SIM3 and c) (—) 450K_SIM4.

In the same temporal region "dlf" exhibited a more gradual increase (from 0.60 to 0.75) **(Figure 10)** probably due to smoothening of the function by averaging over all pairs of snapshots in an epoch (450K_SIM1). However, persf demonstrated lesser fluctuations in the other simulation trajectories (450K_SIM2-4) compared to dlf (**Figure 12, Table: 5).** At 500 K_SIM1, both persf and dlf values had increased to about 0.90 indicating complete unraveling of native packing in the core. The most notable feature in both measures was their highly sensitive signaling of the gradual dynamic relaxation of packing constraints in the hydrophobic core of LdCyP, even at 400 K, indicated both by average and real time values. It will be recalled that such sensitivity was not prominently exhibited even by Disnet alone or probably any other previously studied metric for that matter.

Table 5: Dlf and persf values averaged over the entire simulation block (between 2 – 50ns) for the hydrophobic core of LdCyp at different simulation temperatures for (450K _SIM2, 450K_SIM3 and 450K_SIM4). The standard deviations are given in parentheses.

Temperature (K)	<dlf>	<Persf>
450K_SIM2	0.66(0.09)	0.39(0.08)
450K_SIM3	0.60(0.04)	0.41(0.10)
450K_SIM4	0.59(0.02)	0.41(0.11)

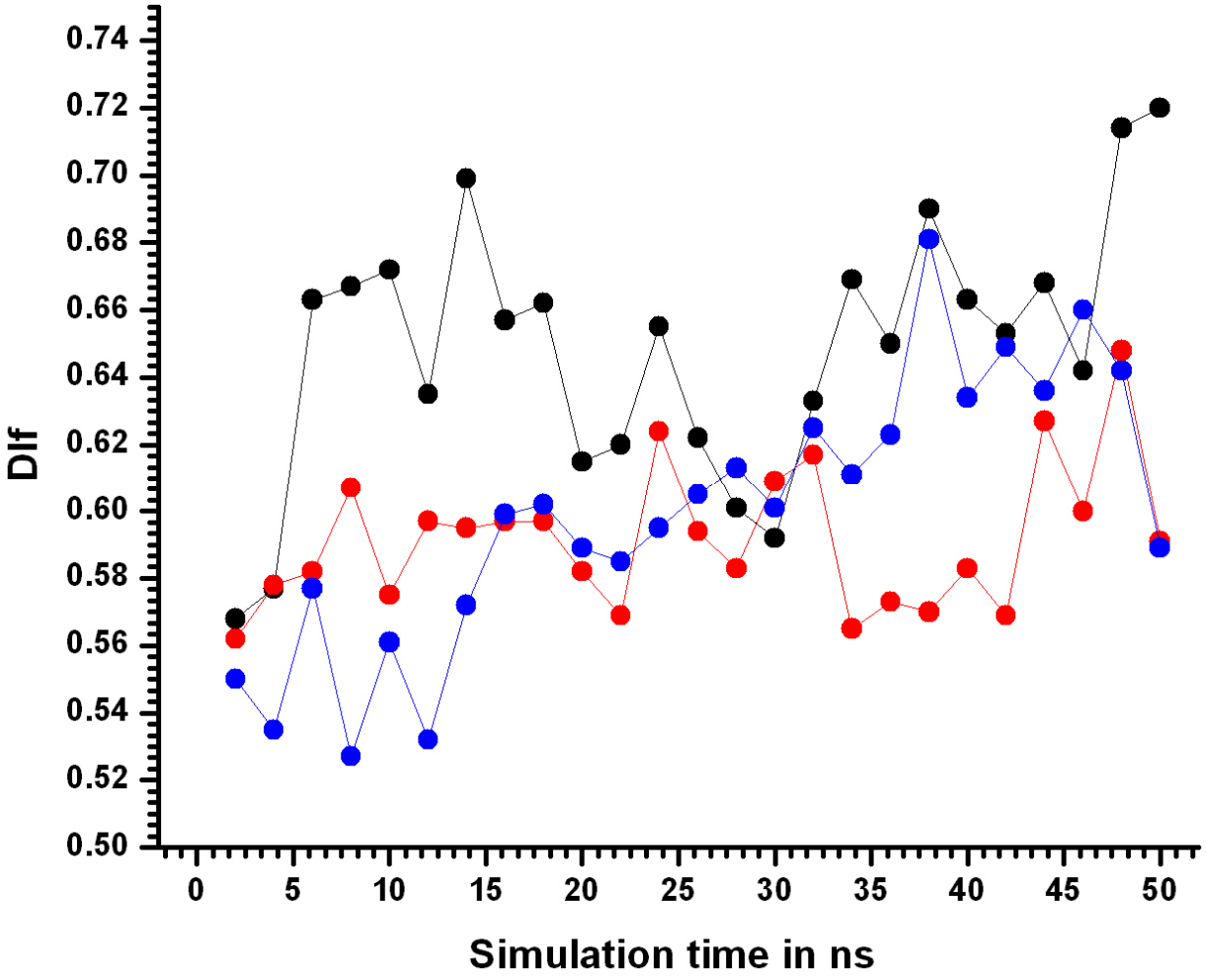


Figure 12 Dlf values for simulations at 450K plotted versus simulation epochs, a) (—) 450K_SIM2 b) (—) 450K_SIM3 and c) (—) 450K_SIM4.

3.3.4. Solvent Accessibility and Secondary Structural Elements

To probe the solvent accessibility of core residues in the course of unfolding simulations and to study their correlation with the SCN based metrics, the 'burial' **(bur: Materials and Methods)** of the 24 amino acids constituting the hydrophobic core of LdCyP were computed for every snapshot. The 'burial' (*bur*) of a residue (X) in a polypeptide chain essentially denotes the fraction of its total solvent accessible area (estimated in a Gly – X – Gly peptide fragment with probe radius of 1.4 Å), accessible to solvent in its current conformational position in the polypeptide chain. Thus, a bur value of zero for any residue denotes complete burial within the interior of the protein molecule. The net solvent accessibility of the core was estimated by averaging *bur* over all the 24 core residues (bur_{core}), and bur_{core} from all the snapshots spanning a 2ns epoch were averaged in order to derive $bur_{epoch.}$ For the native simulation at 310 K and even at 400 K values of bur_{epoch} were consistently close to zero, denoting the existence of an integral core, by and large preventing the penetration of solvent into the interior of the molecule. Concomitant to the surge in persf around 25 – 30 ns at 450K_SIM1, an abrupt increase in bur_{epoch} was also observed in the same region, in two increments, first rising to about 0.04 and then to 0.10 at 35 ns **(Figure 13).** Even at 500 K, the maximum value observed for bur_{epoch} was in the range of 0.15.

Analysis of the fractional secondary structural content of LdCyp **(Materials & Methods)** which is the fraction of residues constituting helices/strands in a snapshot (as found in the native crystal structure) and averaged over a 2ns epoch (SSC_{epoch}), confirmed stable secondary structural elements in the native 310 K_SIM1 simulation block. At 400 K_SIM1, helix H2 exhibited a tendency to unwind upon dissociation from the hydrophobic core (for more details see section on Contacts given below), reflected in a 10 % drop in $SSC_{epoch.}$ At 20ns

(450K_SIM1) there was a sharp decline in SSC_{epoch} **(Figure 13)** falling to a minimum value of 0.55 (at 32 ns) which indicated pronounced disruption for both helices and the barrel. Even though SSC_{epoch} subsequently recovered somewhat to about 0.75 at 40ns, post 20ns all secondary structural features appeared to be irreversibly destabilized, culminating in relatively volatile secondary structural features at 500K_SIM1.

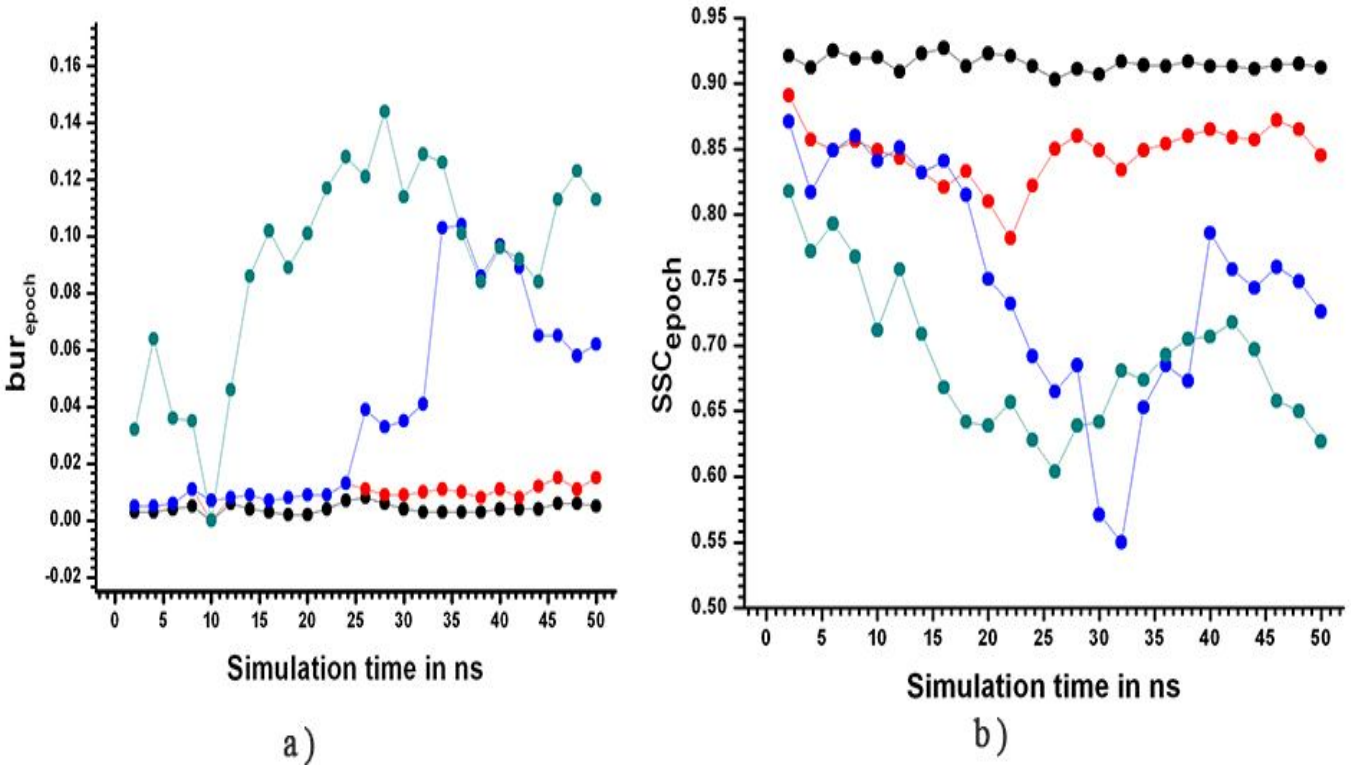


Figure 13: a) Bur_{epoch} values for every 2ns epoch indicated by filled circles plotted versus simulation time at different simulation temperatures (—) 310K_SIM1, (—) 400K_SIM1, (—)450K_SIM1, (—)500K_SIM1.

b) Fractional secondary content or SSC_{epoch} values for every 2 ns epochs indicated by filled circles plotted as a function of simulation time at different simulation temperatures , the color coding following the same convention given above.

3.3.5 Evolution of Contacts

Next, the time evolution of individual surface contacts sustaining the hydrophobic core of the molecule was studied. As has been mentioned previously, the native surface contacts **(Materials and Methods)** of the core were parsed into three subsets, namely S1: contacts between residues located on ß strands, S2: between helix H1 and the remaining residues of the core and S3: between H2 and the remaining core residues

All contacts were analyzed in terms of 'persistence' **(Materials and Methods)** which in effect is the number of times (or snapshots), the surface contact was observed in an epoch, divided by the total number of snapshots in the same duration (200). That is a persistence of 1.00 implies that the specific surface contact was observed in all the snapshots in the epoch. Only three surface contacts were observed in the set S3, out of which one (164I-179V) had a rather low persistence of 0.50, even at 310K_SIM1 (**Table 6**). A drop of about 40 - 50 % in persistence for contacts in S3 at 400 K_SIM1 provided structural rationale to the early dislocation of helix H2 from the barrel or the unique core of the molecule. In contrast, 9 contacts involving helix H1 (set S2) and 11 in set S1 more or less maintained persistence levels at 400K_SIM1 similar to those at 310 K_SIM1, implying conservation of the native geometry with regard to the beta strands and helix H1

Table 6: Average persistence of contacts between residues within the core, divided into subsets of S1 (strand –strand or strand-loop contact), S2 (Strand-Helix1) and S3 (Strand-Helix2) at simulation temperatures of 310,400,450 and 500K with simulations sets of (310K_SIM1,400K_SIM1, 450K_SIM1 and 500K_SIM1) and the standard deviation given in parentheses. The abbreviations in bracket represent: L-Loop, S-strand, H1-Helix1 and H2-Helix2.

	SET S1		SET S2		SET S3	
Temp(K)	Core-Core	Perst(SD)	Helix H1-Core	Perst(SD)	Helix H2-Core	Perst(SD)
310K_SIM1	120LEU(S)-151PHE(S)	0.98(0.01)	59 PHE(H1)-151PHE(S)	0.98(0.08)	85 PHE(S)-164PHE(H2)	0.98(0.01)
	29VAL(S)-47LEU(S)	0.96(0.03)	62 LEU(H1)-71 TYR(L)	0.98(0.01)	85 ILE(S)-161VAL(H2)	0.78(0.05)
	122 MET(S)-151PHE(S)	0.94(0.03)	29VAL(S) -63CYS(H1)	0.96(0.02)	164 ILE(H2)179VAL(S)	0.50(0.07)
	31PHE (S)-181ILE(S)	0.92(0.04)	55THR(H1)-151PHE(S)	0.96(0.03)		
	76 PHE(L)-85 ILE(S)	0.84(0.06)	59 PHE (H1)-71TYR(L)	0.96(0.02)		
	31PHE(S)-45ILE(S)	0.71(0.09)	55THR(H1)-122MET(S)	0.93(0.03)		
	76 PHE(L)-181 ILE(S)	0.68(0.18)	47LEU(S)- 59PHE(H1)	0.67(0.14)		
	71 TYR(L) -134 PHE(S)	0.64(0.10)	59PHE(H1)-134PHE(S)	0.59(0.09)		
	31PHE (S)-134PHE(S)	0.46(0.15)	31PHE(S)-59PHE(H1)	0.26(0.20)		
	31PHE (S)-71TYR(S)	0.30(0.31)				
	33VAL(S -179VAL(S)	0.18(0.16)				
400K_SIM1	120 LEU(S)-151 PHE(S)	0.94(0.03)	62 LEU(H1)-71 TYR(L)	0.99(0.01)	85 ILE(S)-164ILE(H2)	0.49(0.19)
	31 PHE(S)-181 ILE(S)	0.88(0.03)	55 THR(H1)-151PHE(S)	0.98(0.02)	164 ILE(H2)179VAL(S)	0.47(0.09)
	29 VAL(S)- 47 LEU(S)	0.82(0.08)	29 VAL(S)-63 CYS(H1)	0.94(0.03)	85 ILE(S)-161VAL(H2)	0.39(0.12)
	76 PHE(L)- 85 ILE(S)	0.78(0.08)	59 PHE(H1)-151PHE(S)	0.93(0.05)		
	76 PHE(L)-181 ILE(S)	0.77(0.08)	59 PHE(H1)-71 TYR(L)	0.87(0.07)		
	71 TYR(L)-134 PHE(S)	0.77(0.01)	59 PHE(H1)-134PHE(S)	0.78(0.07)		
	122 MET(S)-151PHE(S)	0.70(0.10)	47 LEU(S)-59 PHE(H1)	0.72(0.08)		
	31 PHE(S)-71 TYR(S)	0.65(0.18)	55 THR(H1)-122MET(S)	0.56(0.23)		
	31 PHE(S)- 45 ILE(S)	0.55(0.06)	31 PHE(S)-59PHE(H1)	0.49(0.11)		
	31 PHE(S)-134 PHE(S)	0.54(0.10)				
	33 VAL(S) -179 VAL(S)	0.47(0.10)				
450K_SIM1	31PHE (S)- 181ILE(S)	0.87(0.06)	47LEU(S)-59PHE(H1)	0.55(0.22)	164 ILE(H2)179VAL(S)	0.46(0.09)
	71 TYR(L) - 134PHE(S)	0.68(0.20)	62 LEU(H1)-71 TYR(L)	0.54(0.28)	85ILE(S) -164ILE(H2)	0.45(0.10)
	122 MET(S)-151PHE(S)	0.59(0.14)	59 PHE (H1)-71TYR(L)	0.51(0.27)	85 ILE(S) -161VAL(H2)	0.37(0.09)
	76 PHE(L) - 85 ILE(S)	0.57(0.11)	29VAL(S) - 63CYS(H1)	0.49(0.34)		
	31PHE(S) -45ILE(S)	0.56(0.08)	59 PHE(H1)-151PHE(S)	0.48(0.40)		
	120 LEU(S) -151PHE(S)	0.53(0.26)	55THR(H1)-151PHE(S)	0.35(0.37)		
	76 PHE(L)- 181 ILE(S)	0.51(0.07)	59PHE(H1)-134PHE(S)	0.26(0.27)		
	31PHE (S)- 71TYR(S)	0.47(0.17)	55THR(H1)-122MET(S)	0.15(0.24)		
	29VAL(S)- 47LEU(S)	0.41(0.25)	31PHE(S)-59PHE(H1)	0.10(0.13)		
	31PHE (S)- 134PHE(S)	0.20(0.14)				
	33VAL(S) - 179VAL(S)	0.17(0.12)				
500K_SIM1	31PHE (S)-181ILE(S)	0.60(0.14)	55THR(H1) -151PHE(S)	0.55(0.25)	164 ILE(H2)179VAL(S)	0.07(0.10)
	76 PHE(L)-85 ILE(S)	0.48(0.13)	47LEU(S)-59PHE(H1)	0.50(0.17)	85 ILE(S)-161VAL(H2)	0.06(0.10)
	29VAL(S)- 47LEU(S)	0.41(0.20)	29VAL(S)-63CYS(H1)	0.39(0.31)	85ILE(S)-164ILE(H2)	0.02(0.05)
	33VAL(S)-179VAL(S)	0.40(0.09)	59PHE(H1)-134PHE(S)	0.31(0.24)		
	120 LEU(S) -151PHE(S)	0.35(0.23)	62 LEU(H1)-71TYR(L)	0.31(0.26)		
	76 PHE(L)-181ILE(S)	0.29(0.16)	59 PHE(H1)-151PHE(S)	0.26(0.32)		
	31PHE(S)-45ILE(S)	0.27(0.18)	55THR(H1)-122MET(S)	0.16(0.18)		
	122 MET(S)-151PHE(S)	0.25(0.19)	59 PHE (H1)-71TYR(L)	0.08(0.19)		
	71 TYR(L)-134PHE(S)	0.22(0.27)	31PHE(S)-59PHE(H1)	0.07(0.09)		
	31PHE (S)-134PHE(S)	0.17(0.18)				
	31PHE (S)-71TYR(S)	0.14(0.16)				

However, the contacts in both sets (S1 and S2) exhibited non – uniform levels of persistence at 400K_SIM1, the strongest being 31PHE-181ILE, 120 LEU – 151PHE, 29VAL – 47LEU,

122MET – 151PHE and 59PHE – 151PHE, 62LEU-71TYR, 29VAL-63CYS,55THR – 151PHE, 59PHE-71TYR in sets S1,S2 respectively. Non-uniform levels of persistence and drop in persistence for S3 were also observed in 400K_SIM2 (**Table 7**). At 450 K , an overall decline of about 40 – 50 % in the persistence of the majority of surface contacts was observed in sets S1, S2 indicating profound disruption in native core packing (or side chain association geometries) and also dislocation of the second helix (H1). In S1 only contacts 31PHE – 181 ILE, 71TYR- 134PHE were able to maintain persistence levels comparable to the native simulation (310K). Similar trends were observed in the other trajectories at 450K (_SIM2-4) with a decline (of about 25 – 30%) in the average persistence of the stable contacts (120LEU-151PHE, 31PHE-181ILE, 122MET-151PHE in S1 and 59PHE-151PHE,29VAL-63CYS,62LEU-71TYR in S2 (**Table 8**).

Table 7: Persistence of contacts between residues within the core, divided into subsets of S1 (strand–strand or strand-loop contact), S2 (Strand-Helix1) and S3 (Strand-Helix2) at simulation temperature of 400K_SIM2, the standard deviation given in parentheses. The abbreviations in bracket represent: L-Loop, S-strand, H1-Helix1 and H2-Helix2.

Temp(K)	SET S1		SET S2		SET S3	
	Core-Core	Persistence(SD)	Helix H1-Core	Persistence(SD)	Helix H2-Core	Persistence(SD)
400K_SIM2	120LEU(S)-151PHE(S) 31PHE (S)-181ILE(S) 122 MET(S)-151PHE(S) 29VAL(S)-47LEU(S) 31PHE (S)-71TYR(S) 76 PHE(L)-181 ILE(S) 76 PHE(L)-85 ILE(S) 71 TYR(L) -134 PHE(S) 31PHE(S)-45ILE(S) 31PHE (S)-134PHE(S) 33VAL(S -179VAL(S)	0.92(0.22) 0.83(0.11) 0.82(0.20) 0.81(0.08) 0.47(0.19) 0.54(0.19) 0.79(0.19) 0.73(0.19) 0.47(0.13) 0.45(0.14) 0.14(0.23)	59 PHE(H1)-151PHE(S) 29VAL(S) -63CYS(H1) 62 LEU(H1)-71 TYR(L) 59PHE(H1)-134PHE(S) 55THR(H1)-151PHE(S) 47LEU(S)- 59PHE(H1) 59 PHE (H1)-71TYR(L) 55THR(H1)-122MET(S) 31PHE(S)-59PHE(H1)	0.77(0.19) 0.91(0.03) 0.95(0.27) 0.63(0.18) 0.91(0.23) 0.48(0.20) 0.88(0.22) 0.53(0.24) 0.40(0.16)	164 ILE(H2)-179VAL(S) 85 PHE(S)-164PHE(H2) 85 ILE(S)-161VAL(H2)	0.73(0.24) 0.66(0.25) 0.50(0.19)

Likewise the three contacts (164 ILE-179VAL, 85 PHE-164PHE, 85 ILE-161 VAL) in S3 exhibited average persistence in the range of 0.13-0.27(**Table 8**). Thus, by and large the

behavior of the contacts, in terms of persistence, was replicable in the other trajectories at 450K. At 500 K helix H2 completely dissociated from the main body of the protein and despite further decline in persistence in sets S1,S2 indicating both profound disruption of the core and packing of helix H1, a few selected contacts still exhibited residual values of about 0.50 (31F- 181I, 76F-85I,55T-151F, 47L-59F).

Table 8: Persistence of contacts between residues within the core, divided into subsets of S1 (strand –strand or strand-loop contact), S2 (Strand-Helix1) and S3 (Strand-Helix2) at simulation temperatures of 310,400,450 and 500K with simulations sets of (310KSIM1,400KSIM1, 450KSIM1 and 500KSIM1) and the standard deviation given in parentheses. The abbreviations in bracket represent: L-Loop, S-strand, H1-Helix1 and H2-Helix2. Perst (Persistence).

Temp(K)	SET S1		SET S2		SET S3	
	Core-Core	Perst (SD)	Helix H1-Core	Perst (SD)	Helix H2-Core	Perst (SD)
450K_SIM2	120LEU(S)-151PHE(S) 31PHE (S)-181ILE(S) 122 MET(S)-151PHE(S) 29VAL(S)-47LEU(S) 31PHE (S)-71TYR(S) 76 PHE(L)-181 ILE(S) 76 PHE(L)-85 ILE(S) 71 TYR(L) -134 PHE(S) 31PHE(S)-45ILE(S) 31PHE (S)-134PHE(S) 33VAL(S -179VAL(S)	0.79(0.21) 0.69(0.14) 0.61(0.22) 0.57(0.18) 0.49(0.25) 0.46(0.20) 0.45(0.20) 0.41(0.27) 0.37(0.15) 0.26(0.21) 0.17(0.20)	59 PHE(H1)-151PHE(S) 29VAL(S) -63CYS(H1) 62 LEU(H1)-71 TYR(L) 59PHE(H1)-134PHE(S) 55THR(H1)-151PHE(S) 47LEU(S)- 59PHE(H1) 59 PHE (H1)-71TYR(L) 55THR(H1)-122MET(S) 31PHE(S)-59PHE(H1)	0.73(0.19) 0.68(0.23) 0.58(0.27) 0.54(0.18) 0.46(0.27) 0.37(0.18) 0.28(0.27) 0.16(0.27) 0.09(0.23)	164 ILE(H2)-179VAL(S) 85 PHE(S)-164PHE(H2) 85 ILE(S)-161VAL(H2)	0.27(0.24) 0.26(0.25) 0.22(0.25)
450K_SIM3	120 LEU(S)-151 PHE(S) 31 PHE(S)-181 ILE(S) 71 TYR(L)-134 PHE(S) 29 VAL(S)- 47 LEU(S) 122 MET(S)-151 PHE(S) 76 PHE(L)-181 ILE(S) 31 PHE(S)-71 TYR(S) 31 PHE(S)-134 PHE(S) 31 PHE(S)- 45 ILE(S) 76 PHE(L)- 85 ILE(S) 33 VAL(S) -179 VAL(S)	0.76(0.20) 0.74(0.19) 0.73(0.26) 0.69(0.07) 0.66(0.22) 0.52(0.19) 0.45(0.15) 0.37(0.18) 0.35(0.14) 0.34(0.25) 0.17(0.21)	62 LEU(H1)-71 TYR(L) 59 PHE(H1)-151 PHE(S) 29 VAL(S)-63 CYS(H1) 59 PHE(H1)-71 TYR(L) 55 THR(H1)-151 PHE(S) 59 PHE(H1)-134 PHE(S) 47 LEU(S)- 59 PHE(H1) 55 THR(H1)-122 MET(S) 31 PHE(S)-59 PHE(H1)	0.92(0.24) 0.80(0.21) 0.76(0.09) 0.63(0.23) 0.62(0.24) 0.61(0.16) 0.38(0.19) 0.30(0.22) 0.28(0.17)	85 ILE(S)-164ILE(H2) 85 ILE(S)-161VAL(H2) 164 ILE(H2)-179VAL(S)	0.18(0.34) 0.16(0.29) 0.14(0.32)
450K_SIM4	120 LEU(S) -151PHE(S) 31PHE (S)- 181ILE(S) 122 MET(S)-151PHE(S) 71 TYR(L) - 134 PHE(S) 29VAL(S)- 47LEU(S) 76 PHE(L)- 181 ILE(S) 31PHE (S)- 134PHE(S) 31PHE (S)- 71TYR(S) 76 PHE(L) - 85 ILE(S) 31PHE(S) -45ILE(S) 33VAL(S) - 179VAL(S)	0.81(0.24) 0.79(0.17) 0.79(0.21) 0.73(0.20) 0.63(0.09) 0.53(0.18) 0.42(0.16) 0.42(0.16) 0.40(0.22) 0.36(0.13) 0.17(0.20)	62 LEU(H1)-71 TYR(L) 29VAL(S) - 63CYS(H1) 55THR(H1)-151PHE(S) 59 PHE (H1)-71TYR(L) 59 PHE(H1)-151PHE(S) 59PHE(H1)-134PHE(S) 47LEU(S)-59PHE(H1) 31PHE(S)-59PHE(H1) 55THR(H1)-122MET(S)	0.93(0.28) 0.77(0.13) 0.77(0.21) 0.75(0.20) 0.70(0.19) 0.64(0.17) 0.50(0.17) 0.30(0.16) 0.25(0.22)	164 ILE(H2)-179VAL(S) 85 ILE(S) - 161VAL(H2) 85ILE(S) -164ILE(H2)	0.21(0.29) 0.15(0.28) 0.13(0.30)

That is unlike H2, helix H1 maintained some form of non – native association with a disrupted core. It can also been seen that in the vicinity of the transition state specific high persistent strategic contacts (29Val-47Leu, 76Phe-85Ile, 120Leu-151Phe and 122Met-151Phe) in the set of S1 and (29Val-63Cys, 55Thr-151Phe, 59Phe-71Tyr, 59Phe-151Phe and 62Leu-71Tyr) in S2 declined rather sharply (**Table 9a, 9b**)

Table 9a: Epoch wise persistence values of strategic high persistent contacts in Set S1 for 450K_SIM1

Epochs (ns)	29Val-47Leu	76Phe-85Ile	120Leu-151Phe	122Met-151Phe
2	0.76	0.83	0.95	0.88
4	0.73	0.69	0.97	0.85
6	0.75	0.59	0.95	0.73
8	0.70	0.46	0.89	0.64
10	0.75	0.41	0.92	0.49
12	0.60	0.52	0.76	0.46
14	0.68	0.55	0.59	0.61
16	0.64	0.43	0.56	0.50
18	0.40	0.55	0.24	0.59
20	0.51	0.44	0.68	0.51
22	0.16	0.62	0.17	0.70
24	0.31	0.69	0.39	0.69
26	0.35	0.69	0.31	0.34
28	0.20	0.65	0.27	0.39
30	0.05	0.52	0.61	0.41
32	0.00	0.57	0.43	0.57
34	0.01	0.40	0.50	0.82
36	0.10	0.39	0.76	0.42
38	0.09	0.56	0.53	0.47
40	0.44	0.67	0.40	0.62
42	0.25	0.66	0.22	0.67
44	0.36	0.57	0.32	0.67
46	0.38	0.57	0.28	0.55
48	0.44	0.65	0.31	0.58
50	0.56	0.57	0.28	0.47

Table 9b: Epoch wise persistence values of strategic high persistent contacts in Set S2 for 450K_SIM1.

Epochs (ns)	29Val-63Cys	55Thr-151Phe	59Phe-71Tyr	59Phe-151Phe	62Leu-71Tyr
2	0.82	0.92	0.87	0.89	0.97
4	0.76	0.94	0.85	0.96	0.99
6	0.69	0.85	0.63	0.97	0.96
8	0.74	0.84	0.69	0.94	0.91
10	0.90	0.88	0.75	0.86	0.95
12	0.82	0.74	0.52	0.65	0.92
14	0.87	0.86	0.62	0.86	0.88
16	0.90	0.74	0.50	0.88	0.94
18	0.74	0.41	0.49	0.86	0.93
20	0.77	0.31	0.60	0.84	0.90
22	0.67	0.41	0.61	0.90	0.63
24	0.49	0.38	0.66	0.58	0.71
26	0.61	0.16	0.58	0.69	0.37
28	0.74	0.23	0.29	0.67	0.65
30	0.59	0.01	0.00	0.02	0.00
32	0.29	0.00	0.05	0.04	0.00
34	0.44	0.00	0.05	0.09	0.00
36	0.19	0.00	0.00	0.00	0.00
38	0.16	0.00	0.03	0.03	0.00
40	0.00	0.00	0.51	0.11	0.10
42	0.00	0.00	0.73	0.01	0.37
44	0.00	0.00	0.64	0.01	0.47
46	0.00	0.00	0.71	0.01	0.44
48	0.00	0.00	0.73	0.01	0.51
50	0.00	0.00	0.64	0.03	0.52

possibly suggesting that the onset of the transition state is signaled by the collapse of high persistent contacts that are integral in maintaining the stability of both the core and the association of Helix H1 with the core. Thus, as per the simulations the dissolution of internal packing or the disjuncture of native side chain association within the protein appeared to proceed in a series of sequential steps.

3.3.6 Classical Mutidimensional Scaling

Classical multidimensional scaling (MDS) was utilized to transform (Disnet based) 'distances' into coordinates **(Materials & Methods),** in order to ascertain whether the distribution of points could enable the identification of the transition state region. Snapshots were sampled every 100 ps for every simulation block (at temperatures 310, 400 K etc.) and the matrix of inter-snapshot 'distances' (Disnet) converted to coordinates independently for each temperature by MDS**.**

The distribution **(Figure 14)** at 310 and 400K_SIM1 were very similar, dominated by a unique dense cluster of points with a few outliers. The only difference between the distributions was the marginal increase in the number of outliers at 400K_SIM1 (with respect to 310). The pattern for the 450K_SIM1 simulation block was however substantially different, with the total number of points being partitioned into two distinct distributions **(Figure 15)**; one retaining the characteristics of the previously observed dense cluster whereas the other fanning out into a widely dispersed array. If a rapid expansion in the number of conformational states of the protein can be expected immediately subsequent to the transition state then the departure from the densely packed region of the plot **(**between 20 to 22 ns**: Figure 15)** could possibly be identified as exit from the transition state. A dense cluster of points represents an ensemble of structures with very similar internal architecture. On the other hand, the wide dispersal of points denotes the dissolution of the compact core substituted by labile, non-native interactions in the interior of the molecule. By 500 K the entire distribution was dominated by a widely dispersed array of points indicative of a completely dissolved core.

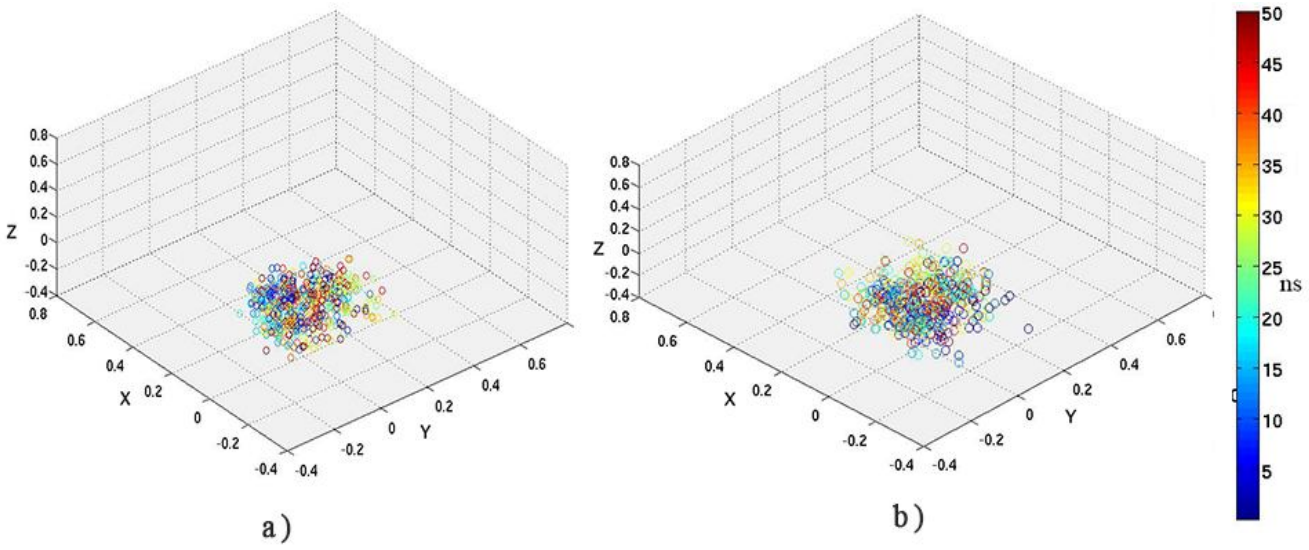


Figure 14: Coordinates derived by MDS, based on the metric Disnet at simulation temperatures of a) 310K_SIM1 and b) 400K_SIM1.The color bar for the plotted points denoting snapshots at indicated time intervals during the course of the simulation.

.

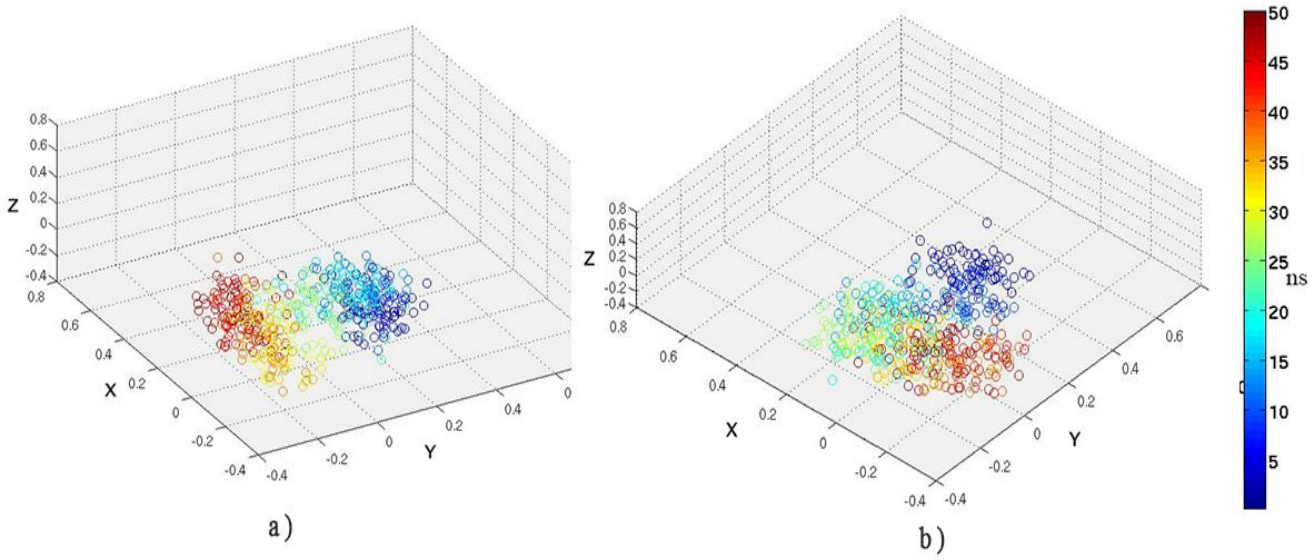


Figure 15: Coordinates derived by MDS, based on the metric Disnet at simulation temperatures of a) 450K_SIM1 and b) 500K_SIM1.The color bar for the plotted points denoting snapshots at indicated time intervals during the course of the simulation

3.4 Conclusions

The dual use of surface complementarity measures defined on a surface representation of a protein structure and network based metrics provide several insights into the unfolding of the cyclophilin molecule which can be summarized as follows.

1. The dissociation of helix H2 from the core or the main body of the protein appears to be the first step in the unfolding of cyclophilin.
2. Even at 400K when the hydrophobic core , the β barrel and the association of the other helix (H1) is reasonably intact, though greater conformational flexibility/fluctuations is exhibited by the surface contact network (SCN) spanning the core , with respect to the native structure (simulated at 310 K). This increased dynamic fluctuation is captured by the metrics 'persf', 'dlf' and the examination of the 'persistence' in case of individual contacts. At this stage the core of cyclophilin is wholly inaccessible to the solvent, native like and compact. Thus, the initiation of unfolding could be concomitant to a subtle enhancement in network fluctuations.
3. The unraveling of the SCN is a consequence of sequential disintegration of specific links in the network apparent from following the pattern in their persistence values as a function of time and temperature. Entry into the transition state is signaled by the abrupt destabilization of strategic high persistence contacts. A significant fraction of these contacts could centre about 'hubs' such as 151 and 59 Phe.
4. The main transition state of the unfolding appears between 25 – 30 ns in the 450 K simulation block, indicated by a sudden rise in all the network parameters (Disnet, persf, dlf) subsequent to which there is a collapse of the surface contact network and rapid expansion into non – native conformational states. Passed the transition state the

secondary structures tend to get disordered and the core is relatively exposed to solvent. The measure for solvent accessibility (bur) does not appear to be a very sensitive measure to clearly demarcate between a possible DMG and WMG. No unambiguous identification of a possible WMG prior to the transition state was possible by solvent accessibility calculations. Prior to the transition state (25 – 30 ns: 450 K) all the secondary structural elements are conserved.

5. Conversion of inter – snapshot 'distances' based on network based metrics appears to demarcate the entry into and exit from the transition state ensemble.

Thus, the first steps in the thermal unfolding of LdCyP (which could involve a DMG) exhibits increased dynamical fluctuations in the surface contact network. This subtle increase in the dynamical flexibility of the network could only be captured by network based metrics in terms of surface complementarity. Further entry into the transition state is heralded by the drop in persistence of specific high persistence contacts. In view of the above observations network based metrics involving surface complementarity could provide additional insights into the unfolding process denied to the traditional metrics.

References

1. R.S. Prajapati, S. Indu, R. Varadarajan (2007) *Biochemistry*, 46, 10339-10352.
2. S. Bhattacharyya, R. Vardarajan (2013) *Curr. Opi. Struct. Biol.* 23, (11-21). http://dx.doi/10.1016/j.sbi.2012.10.010.
3. S. Neumaier, T. Kiefhaber (2014) *J. Mol. Biol.* 426, 2520-2528.
4. M. Levitt, M. Gerstein, E. Huang, S. Subbiah J. Tsai (1997) *Annu. Rev. Biochem.* 66, 549-579.

5. E. Gabellieri, G.B. Strambini (2006) *Biophys J.* 90 (9): 3239–3245.

6. P. Malhotra , J.B. Udgaonkar (2014) *Biochemistry,* 53 (22):3608-20.

7. S.S. Sarkar, J.B. Udgaonkar, G. Krishnamoorthy (2013) *Biophys J,* 105(2392-2402).

8. B.R. Rami, J.B. Udgaonkar (2002) *Biochemistry* 41, 1710-1716.

9. R.L. Baldwin, C. Frieden, G.D. Rose (2013) *Proteins: Struct. Func. Bion*. 78: 2725–2737.

10. T. Kiefhaber, R.L. Baldwin (1995) *Proc Natl Acad Sci USA*; 92:2657–2661.

11. T. Kiefhaber, A.M. Labhardt, R.L. Baldwin. (1995) *Nature* 375:513–515.

12. S.D. Hoeltzli, C. Frieden (1995) *Proc Natl Acad Sci USA,* 92: 9318–9322.

13. S.K. Jha, J.B. Udgaonkar (2009) *Proc Natl Acad Sci USA;* 106:12289–12294.

14. A. Reiner, P. Henklein, T. Kiefhaber (2010) *Proc Natl Acad Sci USA*; 107:4955–4960.

15. R.L. Baldwin, G.D. Rose (2013) *Curr. Opi. Struct. Biol* 23 (1): 4-10.

16. E.I. Shakhnovich, A.V. Finkelstein (1989) *Biopolymers* 28:1667–1680.

17. J. Juneja, J.B. Udgaonkar (2002) *Biochemistry* 41: 2641–2654.

18. V. Daggett, A.Li, L.S. Itzhaki, D.E. Otzen, A.R. Fersht (1996) *J. Mol. Biol*. 257, 430-446.

19. K. E. Fulton, E.R.G. Main, V. Daggett, S.E. Jackson (1999) *J. Mol. Biol*. 291, 445-461.

20. A. Li, V. Daggett (1996) *J. Mol. Biol.* 257,412-429

21. N.L. Salimi, H. Bosco, D.A. Agard (2010) *Plos Computational Biology* 2(6) doi: 10.1371/journal.pcbi.1000689

22. R. Banerjee, M. Sen, D. Bhattacharyya, P. Saha (2003) *J. Mol. Biol*. 333, 211–226.

23. S. Basu, D. Bhattacharyya, R. Banerjee (2011) *BMC Bioinformatics* (2011) 12:195, doi: 10.1186/1471-2105-12-195

24. L.H. Greene, V.A. Higman (2003) *J. Mol. Biol.* 334, 781-791.

25. M. Vendruscolo, N.V. Dokholyan, E. Paci, M. Karplus (2001) *Phys Rev E*. 65(6).

26. A.R. Atilgan, P. Akan, C. Baysal (2004) *Biophys J.* 86, 85-91.

27. S. Basu, D. Bhattacharyya, R. Banerjee (2012) *Biophys J*. 102 (11): 2605-2614.

28. S. Basu, D. Bhattacharyya, R. Banerjee (2014) *Indian Journal of Biochemistry and Biophysics*, 51, 188-200.

29. H. Ke, (1992) *J. Mol. Biol,* 228, 539–550.

30. V. Venugopal, B. Sen, A. K. Datta, R. Banerjee (2007) *Acta. Cryst. F*. 63, 60–64.

31. S. Roy, S. Basu, A. K. Datta , D. Bhattacharyya, R. Banerjee, D. Dasgupta (2014) *Int. J. Biol. Macromol.* 69, 353-360.

32. A. Jakalian, D.B. Jack, C.I. Bayly (2002) *J Comput Chem* 23, 1623-1641

33. J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten. (2005) *J. Comput. Chem.* 26 (16) 1781–1802.

34. J.P. Ryckaert, G. Ciccotti, H.J.C., Berendsen, (1977) *J. Comput. Phys*. 23, 327–341.

35. W. Humphrey, A. Dalke, K. Schulten, (1996) *J.Mol.Graph*.14,33–38 http://www.ks.uiuc.edu/Research/vmd/

36. W.D. Cornell, P. Cieplak , C. I. Bayly , I. R. Gould , K. M. Merz , D. M. Ferguson , D. C. Spellmeyer , T. Fox ,J. W. Caldwell , P. A. Kollman (1995) *J. Am. Chem. Soc.* 117, 5179-5197.

37. D. Frishman, P. Argos (1995) *Proteins: Struct. Funct. Genet*. 23, 566–579.

38. I. Borg, P. J. F Groenen (2005) Modern Multidimensional Scaling Springer Series in Statistics, 2nd ed.

39. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus (1983). *J Comp Chem* 4 (2): 187–217

40. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

Chapter 4

Stability Studies of Cyclophilin Mutants

4.1 Introduction

Over the last few decades the thermal stability of proteins has been studied utilizing a plethora of experimental and computational methods. Close packing of side chains, strength of hydrogen bonding, extent of polar solvent accessible surface areas, deletion, shortening of loops and the number of stable amino acid residues in the polypeptide chain [**1-12**] has been identified to be some of the major causes contributing to the thermal stability of proteins. Extensive information in this regard has been obtained from studying the effect of mutations on the three dimensional structure and stability of proteins. The third chapter described the methodology to construct protein surface contact networks and also to identify 'persistent' interactions or contacts within the protein, specifically for cyclophilin. This led to the identification of strategic amino acid residues which served as 'hubs' within the interaction network of cyclophilin, effectively mediating a large number of interactions. In this chapter, mutation of such hubs and their effect on the thermal stability has been studied utilizing differential scanning calorimetry (DSC). A total of seven mutants V33A, T55A, L62A, I85A, L120A, F151V and I164A (**Figure1**) have been thermally characterized and of these the crystal structures of two mutants (V33A, L120A) have been solved at fairly high resolution. The chapter concludes with an attempt to assess the utility of surface complementarity measures as a descriptor of perturbations in core packing as a consequence of mutations. Given that only two crystal structures are currently available further work will have to be done to confirm the utility of

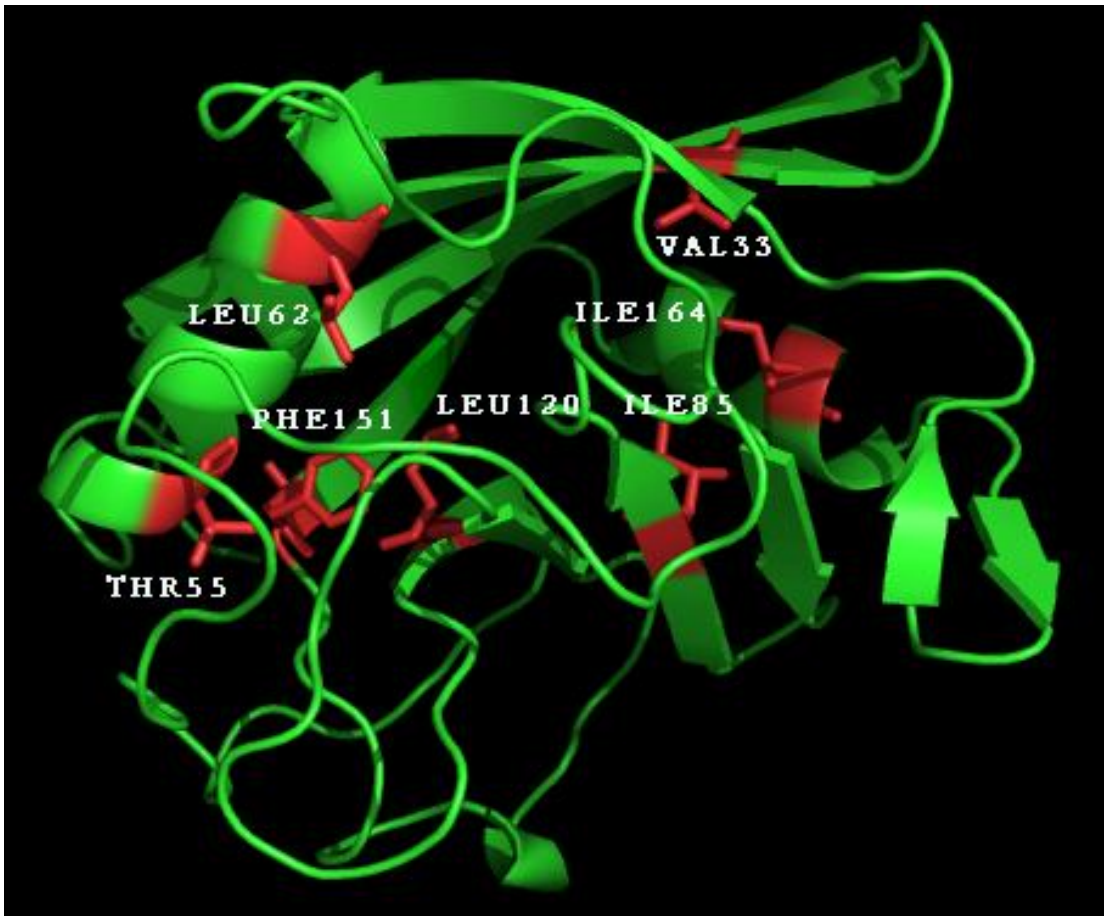


Figure 1: Spatial positions of the residues which have been mutated are colored in red, located within the hydrophobic core of the molecule. The figure has been generated from the crystal structure of LdCyp (Pdb: 2HAQ) using Pymol (Schrodinger LLC).

surface complementarity measures to assess packing defects incurred in the protein hydrophobic core upon mutation of native residues.

4.2 Materials and Methods

4.2.1. Surface complementarity, scores and basis of Mutations:

The definition of surface complementarity (S_m) and the construction of associated surface contact networks has been discussed in detail in Chapter 3. Briefly, surface complementarity estimates the 'goodness of fit' between two molecular surfaces in close association. In order to

estimate Sm between two surfaces **A**, **B** the nearest neighbouring surface point on **B** is identified for every dot surface point on **A** (within a distance cutoff of 3.5 Å) and S_m between the two surfaces is computed by the application of Equation 1 (Chapter 3). Thus, for any buried residue (referred to as 'target') the surface complementarity of its side chain (atoms) with respect to the rest of the atoms constituting the polypeptide chain can be estimated. Further, side chain surface of the target can be effectively partitioned such that each surface patch (Chapter 3) is in 'surface contact' with a specific residue in its vicinity (or in other words the set of dot surface points constituting a surface patch have all found their nearest surface points, belonging to the same neighbouring residue). The mutants were designed such that the mutated residues were not active site residues, located within the unique hydrophobic core of the molecule and could be considered as 'hubs' in the interaction network, involving a large number of 'surface contacts' with residues in their immediate neighborhood. It was one of the objectives of the present study to ascertain, whether surface complementarity measures could effectively estimate the perturbation of native packing upon mutation. Two crystal structures of mutants (V33A and L120A) have been solved, while the rest of the mutants whose crystal structure was not available, was manually mutated to the requisite residues by Cootv0.7.1 [**13**]. Initial energy minimizations with steepest descent of 200 steps and ABNR (adopted basis newton raphson) of 20000 steps was performed followed by solvation and final energy minimizations after charge neutralization of the system by CHARMM [**14**] until the energy of the system converged.

To start with the surface contact neighborhood of every residue which had been targeted for mutation was determined from the crystal structure of the native protein (2HAQ). This involved identification of the set of those residues in its vicinity, with which it was in surface contact (as given in **Table 1**). The S_m values of this set of residues was computed both for the

native and mutant crystal structures. In an earlier report from the laboratory [**15**] the average S_m ($\mu_{database} \pm \sigma_{database}$) had been determined individually for all the 20 amino acids, buried within the interior of proteins from a database of highly resolved protein crystal structures. The S_m values for the neighbouring residues along with the original residue and its mutated amino acid were converted to Z – scores utilizing the formula:

$$Z_{sm} = (Sm_{crystal/model} - \mu_{database}) / \sigma_{database}$$

where, $Sm_{crystal}$ is the actual S_m value obtained either from the native or mutant crystal structure and $\mu_{database}$, $\sigma_{database}$ refer to the average S_m for the specific residue from the database , as mentioned above. Subsequent to this normalization process the corresponding Z scores were subtracted between the native and mutant crystal structures (including the original residue at the mutation site and its mutated counterpart) and finally summed to yield Δ_{Sm}. Δ_{Sm} was considered to be an indicator of the difference in packing between the native and mutant crystal structures. Only side chain atoms were used in the computation of S_m values.

The surface contact network of native cyclophilin (LdCyp: pdb id: 2HAQ) showed the presence of a set of residues mediating a large number of interactions within the hydrophobic core of the molecule (**Figure 1**). These residues exhibited a high degree of interaction with their neighbours qualifying them as 'hubs' (**Table 1**) and also showed a propensity to form "cliques" that is a closed set of interaction involving three residues. All these residues were buried deep within the hydrophobic core of the protein, away from the active site of the molecule. Apart from Phe151, all other residues mentioned above were designed to be mutated to Ala, while Phe 151 was mutated to Val. Point mutations of Val 33 to Ala 33 and Leu 120 to Ala 120 was performed in house through site-directed mutagenesis while mutants T55A, L62A, I85A, F151V and I164A were procured from Genscript Inc. (USA).

Table 1: Residues chosen for mutation in cyclophilin, their immediate neighbours in surface contact and the number of such neighbours, as determined from the native crystal structure (2HAQ).

Residues to be mutated	Residues in surface contact	No. of residues in surface contact
Val33	Phe31,Ile35,Leu40,Leu43,Phe76,Val160,Ile164,Val179,Ile181	9
Thr55	Ala52,Thr56,Asn58,Phe59,Phe106,Glu109,Met122,Val149,Phe151	9
Leu62	Asn58,Phe69,Tyr71,Phe90,Ile101,Ser132	6
Ile85	Phe76,Val79,Ile136,Thr138,Val161,Ile164,Glu165	7
Leu120	Phe31,Ile48,Phe59,Phe134,Ile136,Phe151,Val154	7
Phe151	Leu47, Ala52,Thr55,Thr56,Phe59,Leu120,Met122,Phe134,Val149	9
Ile164	Val33,Ile35,Leu40,Ile43,Phe76,Val79,Ile85,Val160,Thr167,Val179	10

4.2.2 Site-directed Mutagenesis

Polymerase Chain Reaction (PCR) technique **[16]** was used to amplify the genes for the mutants V33A and L120A with pre-designed forward and reverse primers (**Table 2**).

Table 2: Primer and template sequences for point mutants V33A and L120A with the region of nucleotide sequences (sense and anti-sense template) aimed for mutations highlighted in yellow and the desired change in nucleotide sequence for mutation highlighted in green in the forward and reverse primers.

Mutations	Primer Sequences and Template Sequences
V33A	Sense template : 5'-AAGGTGTATTTTGATGTGATGATCGATAGCGA-3' Anti-sense template : 5'-TCGCTATCGATCATCACATCAAAATACACCTT-3' Forward primer : 5'-AAGGTGTATTTTGATGCGATGATCGATAGCGA-3' Reverse primer : 5'-TCGCTATCGATCATCGCATCAAAATACACCTT-3' Nucleotide change : GTG to GCG
L120A	Sense template : 5'-CACTTTGTGGGTGCGCTGTCCATGGCAAACGC-3' Anti-sense template : 5'-GCGTTTGCCATGGACAGCGCACCCACAAAGTG-3' Forward primer : 5'-CACTTTGTGGGTGCGGCGTCCATGGCAAACGC-3' Reverse primer : 5'-GCGTTTGCCATGGACGCCGCACCCACAAAGTG-3' Nucleotide change : CTG to GCG

Plasmid DNA of wild-type cyclophilin (LdCyp) was used as a template in the PCR reaction. The primer annealing temperature T_m was set 5K lower than the theoretical melting temperature of the oligo-nucleotides. Initially 30 cycles of gradient PCR with various annealing temperatures were employed for amplification and to estimate the T_m of the oligo-nucleotides. A reaction mixture for a typical PCR reaction is given below in (**Table 3**)

Table 3: A typical PCR reaction mixture

PCR Components	Concentrations
Template-plasmid DNA(cyclophilin wild type)	~10-20ng
Forward primer	0.6pmol
Reverse primer	0.6pmol
d-NTP mix-100X	1X
10X PCR buffer	1X
Pfu Turbo (enzyme)#	1U
Remaining volume made up to 25µL with autoclaved H_20	

Pfu Turbo was used as a polymerase owing to its superior proof-reading capacity.

The amplified PCR products were then subjected to digestion with the enzyme Dpn-I (Fermentas, USA) to cleave the original parental DNA molecules and restriction enzymes BamH1 (Fermentas, USA), Pvu-1 (New England Biolabs, USA) were utilized to digest the final PCR products. Inserts for V33A and L120A were cloned into the pQE-32 vector (Qiagen, USA), identical to the vector used in case of wild-type cyclophilin. The vector had been previously digested with BamH1 (Fermentas, USA) and HindIII (New England Biolabs, USA) to create sticky ends, to facilitate the ligation of the insert. Finally, the inserts for V33A and L120A were ligated into the target vector with T4 DNA Ligase (Promega Ltd. USA).

The ligation product was directly transformed into electro-competent cells (XL1-blue, Stratagene) according to the manufacturer's protocol by Bio-Rad gene pulser and plated on to agarose plates supplemented with 100µg/ml ampicillin. The plasmids were then purified from the

colonies by the alkaline lysis method **[17]** utilizing the plasmid purification kit (Axygen) and the molecular weight of the inserts were estimated by running agarose gels, subsequent to digestion of the plasmid with appropriate restriction enzymes. Final confirmation of the targeted mutation was confirmed by sequencing the recombinant gene product. Over expressed mutant proteins were obtained from the *E.Coli* strain BL 21(DE3) after transformation of BL-21(DE3) electro-competent cells with pQE-32 vectors containing mutant inserts of V33A and L120A.

Rest of the recombinant plasmids encoding the targeted mutations were procured from Genscript Inc. USA. The mutants were constructed with the same wild-type plasmid of cyclophilin in pQE-32 vector. The sequences of the forward and the reverse primers are given in **Table 4.**

Table 4: Primer and template sequences for point mutants procured from Genscript Inc. with the region of nucleotide sequences (sense and anti-sense template) aimed for mutations highlighted in yellow and the desired change in nucleotide sequence for mutation highlighted in green in the forward and reverse primers.

Mutations	**Primer Sequences and Template Sequences**
T55A	Sense Template: 5'–AGGATGCGCCGCTCACAACAGAGAACTTCCG-3' Antisense Template: 5'-CGGAAGTTCTCTCTTGTGAGCTTCGCATCCT-3' Forward Primer: 5'–AGGATGCGCCGCTCGCAACAGAGAACTTCCG-3' Reverse Primer: 5'-CGGAAGTTCTCTCTTGCGAGCTTCGCATCCT-3' Nucleotide change: ATC to GCC

L62A	Sense template:5'-GAGAACTTCCGGCAGCTCTGCACGGGTGAGCA-3' Antisense template: 5'-TGCTCACCCGTGCAGAGCTGCCGGAAGTTCTC-3' Forward Primer: 5'-GAGAACTTCCGGCAGGCCTGCACGGGTGAGCA-3' Reverse Primer: 5'-TGCTCACCCGTGCAGGCCTGCCGGAAGTTCTC-3' Nucleotide change: CTC to GCC
I85A	Sense template: 5'-ATCCAAAACTTCATGATCCAGGGCGGCGACTT-3' Antisense template: 5'-AAGTCGCCGCCCTGGATCATGAAGTTTTGGAT-3' Forward Primer: 5'-ATCCAAAACTTCATGGCCCAGGGCGGCGACTT-3' Reverse Primer: 5'-AAGTCGCCGCCCTGGGCCATGAAGTTTTGGAT-3' Nucleotide change: ATC to GCC
F151V	Sense template : 5'-GGCCGCCATGTGGTTTTCGGCAAGGTGCTTGA-3' Antisense template:5'-TCAAGCACCTTGCCGAAAACCACATGGCGGCC-3' Forward primer : 5'-GGCCGCCATGTGGTTGTCGGCAAGGTGCTTGA-3' Reverse primer : 5'-TCAAGCACCTTGCCGACAACCACATGGCGGCC-3' Nucleotide change: TTC to GTC
I164A	Sense template : 5'-GACGTGGTGCTCCGCATCGAGAAGACAAAGAC-3' Antisense template :5'-GTCTTTGTCTTCTCGATGCGGAGCACCACGTC-3' Forward Primer : 5'-GACGTGGTGCTCCGCGCCGAGAAGACAAAGAC-3' Reverse Primer : 5'-GTCTTTGTCTTCTCGGCGCGGAGCACCACGTC-3' Nucleotide change: ATC to GCC

All the recombinant plasmids procured from Genscript Inc were inserted into XL-1 blue electro-competent cells and stored till further use. For over-expression *E.Coli* BL-21(DE3) electrocompetent cells, were transformed by the plasmids containing the mutant gene. The sequences of all the recombinant plasmids (mutants) were confirmed by sequencing.

4.2.3 Protein Purification

All the cyclophilin mutants were purified using the identical purification protocol described in **Materials and Methods**, Chapter 3. SDS-PAGE electrophoresis was used to confirm the purity and molecular weight of the purified protein sample.

4.2.4 DSC experiments

10μM of LdCyp and cyclophilin mutants (V33A, L120A, I85A, F151V, L62A and T55A) with the same concentration were scanned on a VP-DSC Microcalorimeter (Microcal LLC, Northampton, MA, USA) from 10°C to 70°C with a scan rate of 30°C/h and at approximately 28 psi pressure. All samples (buffer and protein) were extensively degassed prior to loading. Initially, the buffer (25mM potassium phosphate, pH 7.5) was repeatedly scanned to ensure a stable baseline. The buffer baseline was subtracted from the protein thermogram using Microcal Origin version7.0 software provided with the instrument. The data was then normalized by the protein concentration and a non-linear curve fitting algorithm was employed to obtain the thermodynamic parameters of the transition. The reversibility of unfolding was checked at different scan rates, buffer conditions and temperature ranges. All DSC experiments were repeated at least three times.

4.2.5 Fluorescence Measurements

Intrinsic tryptophan fluorescence of the native protein LdCyp and the cyclophilin mutants

(V33A, L120A, I85A, F151V, L62A and T55A) were measured at 25°C on a Perkin Elmer LS55 spectrofluorimeter (Perkin Elmer Ltd., UK). LdCyp and the mutants at 4 μM concentration in 25mM potassium phosphate buffer (pH 7.5) was excited at 295nm and the emission spectra was recorded in the range 310nm to 450nm with both excitation and emission slit widths kept at 5nm. The spectra were recorded at a scan speed of 200nm/min and averaged over 4 scans. Appropriate controls for the background emission were subtracted in each case.

4.2.6. Crystallization

Crystallization of cyclophilin mutants L120A and V33A was carried out by the hanging drop vapor diffusion method. Diffraction quality crystals were grown in 10 μl drops containing 6-7 mg ml^{-1}protein, 7.5% PEG 3350 inverted over a reservoir containing 1ml of 40% PEG 3350 at 293 K. Crystals suitable for data collection appeared in about two weeks (**Figure 2**).

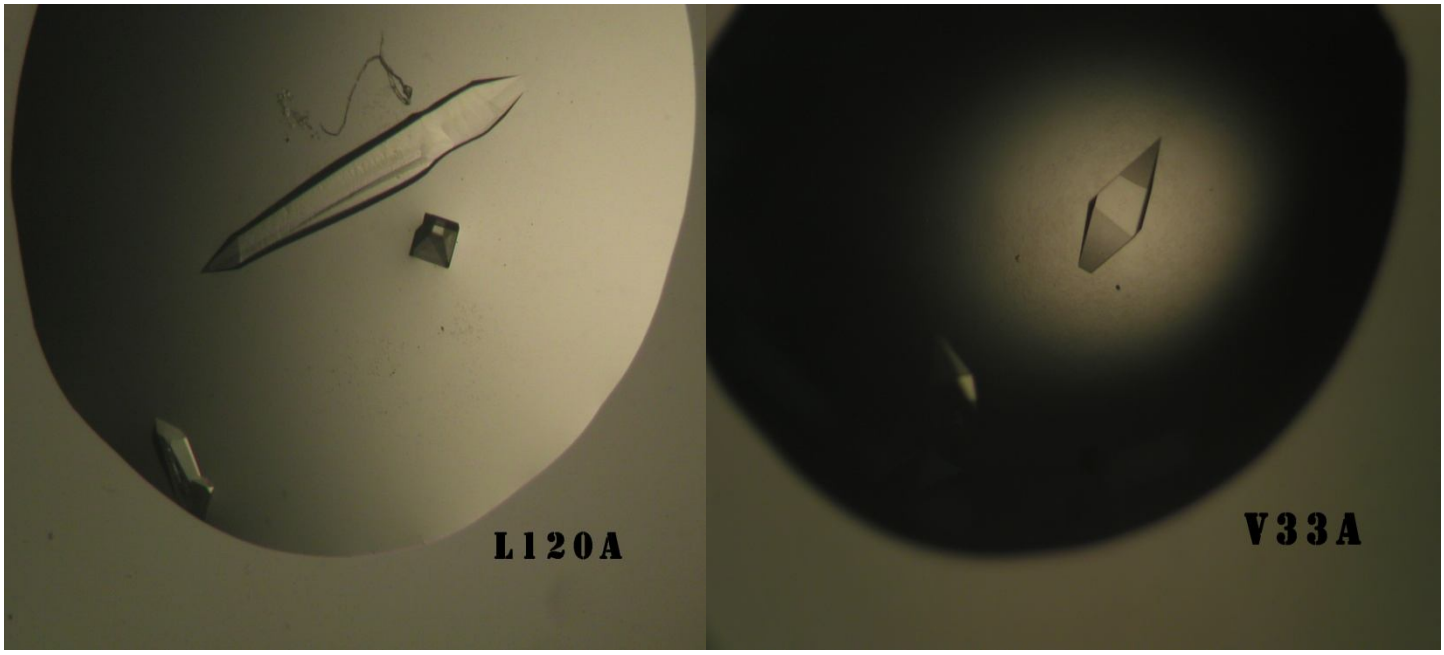


Figure 2: Crystals of cyclophilin mutants L120A and V33A grown with PEG 3350 as a precipitant utilizing hanging drop vapor diffusion method. The crystals grew in temperature of 293K in about two weeks time.

4.2.7 Data collection and processing

Crystals of both L120A and V33A mutants were mounted in locally produced glass capillary tubes utilizing the reservoir solution which served as the mother liquor for the crystals. Data collection for L120A at 25°C was done on a MAR image plate scanner mounted on an in-house Rigaku Rotaflex Cu Kα rotating anode X- ray source, with the crystal to detector distance set at 180.0 mm. The oscillation per frame was set to 1°. The reflection spots were then indexed to define a lattice. The reflection data was then integrated and scaled by *DENZO* and *SCALEPACK* [**18**], and the Laue symmetry of the crystal was determined to be 4/*m*.

Diffraction data for the V33A mutant crystals was collected utilizing a RAXIS-IV^{++} image plate system mounted on a Rigaku RU-H3R Cu Kα rotating anode X-ray source. The crystal to detector distance was set to 190.0mm and incremented by 1° per frame at 293K. The reflections of V33A were indexed, integrated and scaled to generate the final reflection file by the software Crystal Clear v1.4.0, provided with the instrument. L120A and V33A diffracted to a resolution of 2.21 and 2.30Å respectively. The summary of data collection statistics are given in **Table 5**. The structure solutions for both the mutants were obtained by molecular replacement with *AMoRe* [**19**] from the CCP4 program suite, using the native crystal structure of cyclophilin from *leishmania donovani* (LdCyP) (Pdb id: 2HAQ) as the search model.

Table 5: Data collection summary of cyclophilin mutants of V33A and L120A, the values in the parentheses indicate values in the last resolution shell. The values in the parentheses indicate values in the last resolution shell.

V33A

Space group	$P4_3$
Unit cell parameters	a=b=48.53, c=141.7 Å, $\alpha = \beta = \gamma = 90^0$
Resolution Range	50.0 - 2.30 Å (2.38-2.30 Å)
Mosaicity	0.170
Completeness%	97.2 (81.7)
Rmerge % *	4.90(17.8)
No. of Unique Reflections	13585 (1369)
I/σ(I) ratio	13.8 (4.8)
Multiplicity	4.00(2.85)

L120A

Space group	$P4_3$
Unit cell parameters	a=b=48.62, c=141.04 Å, $\alpha = \beta = \gamma = 90^0$
Resolution Range	30.0 - 2.21 Å (2.30-2.21 Å)
Mosaicity	0.201
Completeness%	99.3 (93.7)
Rmerge % *	4.45(12.6)
No. of Unique Reflections	16104(1497)
I/σ(I) ratio	16.8 (8.0)
Multiplicity	5.91(5.37)

*$R_{merge} = \Sigma \Sigma | I(k) - \langle I \rangle | / \Sigma I(k)$, where $I(k)$ and $\langle I \rangle$ represent the diffraction intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements.

:

.

4.3 Results and Discussions

4.3.1 Differential Scanning Calorimetry

The DSC thermogram for the native protein and the mutants exhibited a wide variation in the Cp(T) values of transition. As mentioned in Chapter 3, reversibility of the transition was observed in the native protein only till its T_m1 of 54.6°C. Reversibility in case of the mutant proteins were also exhibited up to the major T_m of the transition. Different scan rates were employed to take care of kinetic effects if any, however the shape of the Cp(T) curve as well as the value(s) of T_m's did not alter significantly. Non-linear curve fitting provided with the instrument was employed to fit the DSC thermograms of the mutants, all of which fitted to a non-2 state model with one or two T_m's as the case may be. T55A appeared to be the most stable mutant thermodynamically with T_m's of 53.82 and 51.09 °C with apparent ΔH of 37.2 kcal $mole^{-1}$ and 23.1 kcal $mole^{-1}$ respectively (**Table 6**). The thermograms of F151V, I85A and L62A were fitted with a single T_m, in contrast to the thermograms of V33A, T55A and L120A which were fitted with two T_m's (**Figure 3**). Thermally, I85A proved to be the least stable mutant with a single T_m of 43.11°C and apparent ΔH of 7.84 kcal $mole^{-1}$. Despite repeated efforts no statistically significant peak was observed for the mutant I164A in the DSC measurements. L120A seemed to be the next stable mutant after T55A, followed by L62A, V33A, F151V and lastly by I85A on the basis of their T_m's and corresponding apparent ΔH of transition.

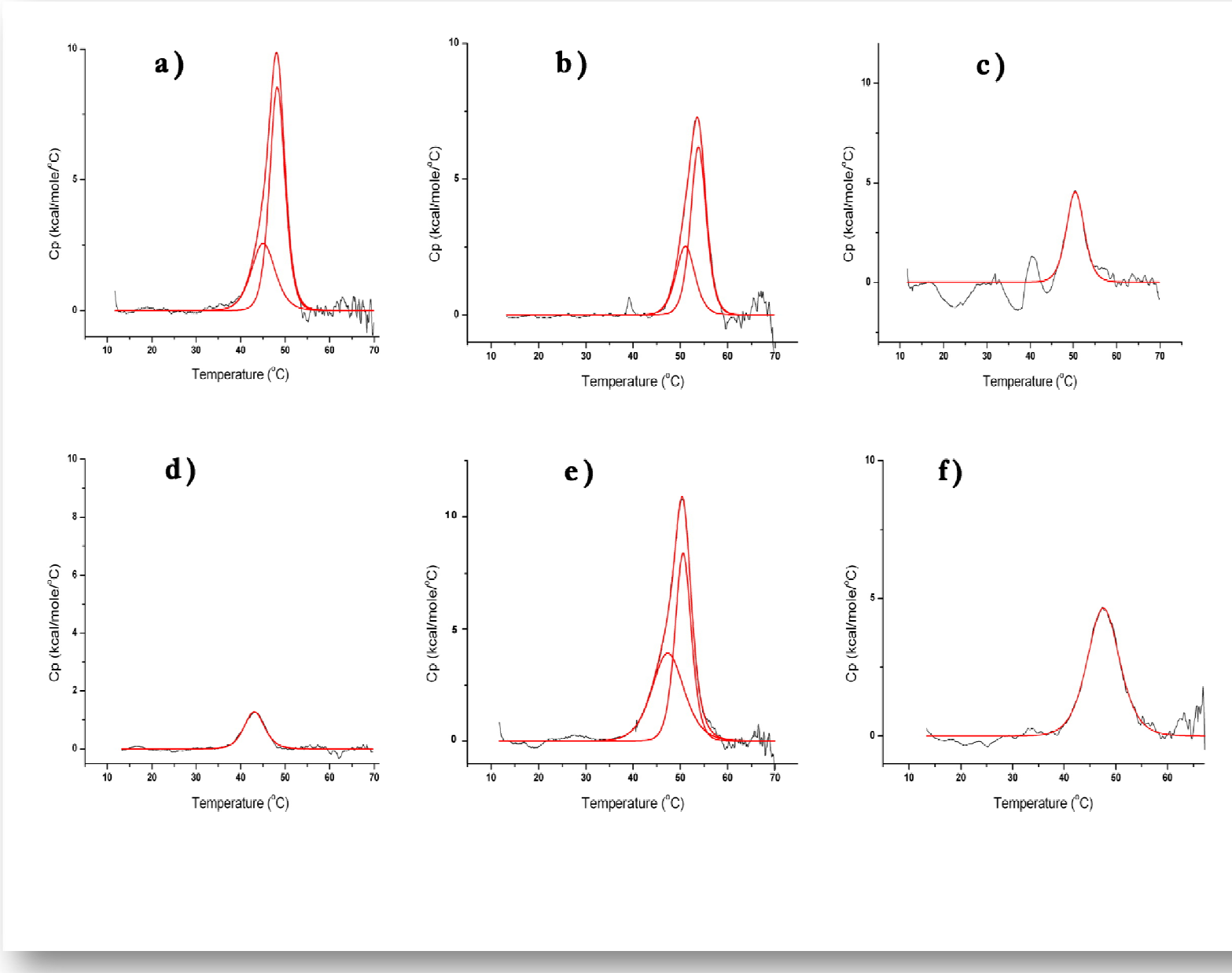


Figure 3: DSC thermograms of the cyclophilin mutants a) V33A b) T55A c) L62A d) I85A e) L120A and f) F151V

Table 6: Relative thermodynamic parameters of cyclophilin mutants V33A, T55A, L62A, I85A, L120A and F151V

Mutants	Unfolding Process	T_m's (°C)	ΔH(Kcal mole^{-1})
V33A	Non-2-state, 2 T_m's	48.25, 45.12	43.14, 16.23
T55A	Non-2-state, 2 T_m's	53.82, 51.09	38.21, 14.11
L62A	Non-2-state, 1 T_m	50.49	23.78
I85A	Non-2-state, 1 T_m	43.11	7.84
L120A	Non-2-state, 2 T_m's	50.60, 47.42	41.29, 15.17
F151V	Non-2-state, 1 T_m	47.63	27.30

4.3.2. Unfolding Fluorescence Curves of the Mutants

Fluorescence unfolding curves of all the cyclophilin mutants were sigmoidal and fitted best to a 3-state equation previously described in Chapter2 (equation 3).

$$S_{obs} = \frac{S_N + S_I e^{\frac{-\Delta G_{NI}}{RT}} + S_U e^{\frac{-\Delta G_{NU}}{RT}}}{1 + e^{\frac{-\Delta G_{NI}}{RT}} + e^{\frac{-\Delta G_{NU}}{RT}}} \qquad (1)$$

The fluorescence signals in case of each of the mutants were normalized and represented in terms of the fraction of the unfolded protein at specific GdmCl concentrations, assuming

$$F_f + F_u = 1 \qquad (2)$$

Where, F_f is the fraction of the protein that is folded and F_u the fraction in the unfolded conformation. The signal at any GdmCl concentration can thus be represented as :

$$Sobs = S_f F_f + S_u F_u \qquad (3)$$

where, S_f is the signal due to the folded state and S_u due to the unfolded state. Combining equations (2) and (3) leads to

$$F_u = \frac{(S_f - S_{obs})}{(S_f - S_u)} \qquad (4)$$

For all the mutants the pre and the post transition baselines were first fixed and the fraction unfolded at 360nm of wavelength was plotted at each GdmCl concentrations to generate the fluorescence unfolding curves (**Figure 4**).

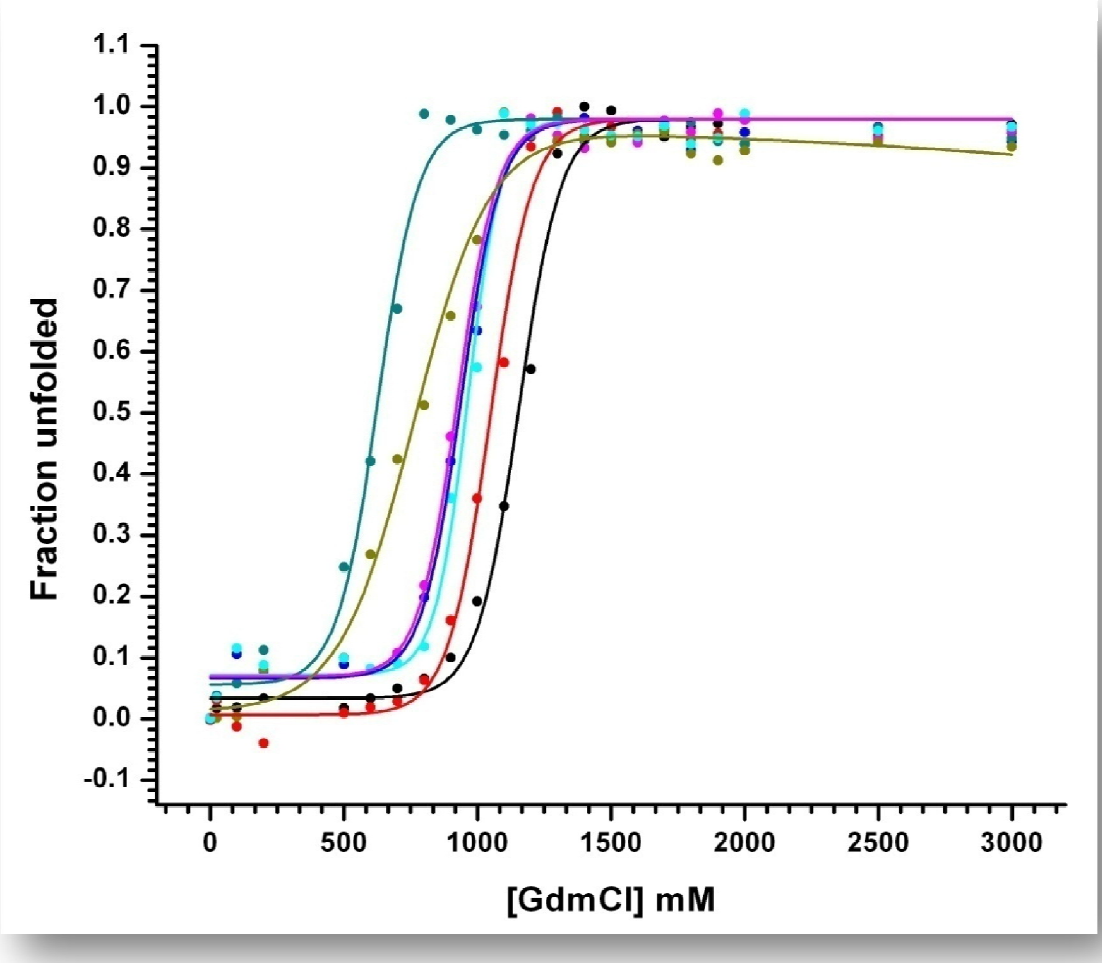


Figure 4: Fluorescence unfolding curves for LdCyp (—) and mutants a) (—) V33A b) (—) T55A c) (—) L62A d) (—) I85A e) (—) L120A and f) (—) F151V in terms of fraction unfolded at various GdmCl concentrations

Signals indicative of unfolding were observed as early as 0.4 M GdmCl concentration in case of mutant F151V, with the fluorescence signal saturating at 0.8 M of GdmCl. On the other hand, the unfolding transition of T55A was similar to native LdCyp with the initiation of unfolding transition at a GdmCl concentration in excess of 1.0M and saturating at a concentration of 1.5M. Thermodynamic parameters of unfolding of both ΔG_{NI} (H_2O) and ΔG_{NU} (H_2O) along with the corresponding mid-point of transitions [C_m ($N \leftrightarrow I$) and C_m($N \leftrightarrow U$)] suggests T55A to be the most stable mutant with thermodynamic parameters close to the native protein whereas F151V was the least stable mutant. ΔG_{NI} (H_2O) and ΔG_{NU} (H_2O) of T55A are 10.12 kcal/mole, -7.39 kcal/mole respectively (**Table 7**) which is approximately 2 kcal/mol, 1 kcal/mole less than the corresponding values of the native protein. The mid-points of transition for T55A also lies close to LdCyp differing by 0.1 M for C_m ($N \leftrightarrow I$) and is comparable to C_m ($N \leftrightarrow U$) (**Table 7**). The mid-points of transition for the mutants V33A, L62A and L120A are also very similar and almost overlap with each other. Concomitant to the mid-points of transition the associated thermodynamic parameters of ΔG_{NI} (H_2O) and ΔG_{NU} (H_2O) for these mutants are also similar corresponding to conformational stability of similar magnitude. The unfolding transition for I85A (**Figure 4**) appears to be somewhat gradual relative to the steep unfolding transition of the other mutants and native LdCyp.

Table 7: Equilibrium thermodynamic parameters for native LdCyp and mutants V33A, L120A, T55A, L62A, I85A and F151V

Native LdCyp and Mutants	Thermodynamic parameters					
	$\Delta G_{NI}(H_2O)$ kcal mol^{-1}	m_{NI} Kcal mol^{-1} M^{-1}	$\Delta G_{NU}(H_2O)$ kcal mol^{-1}	m_{NU} Kcal mol^{-1} M^{-1}	$C_m(N\leftrightarrow I)$ M	$C_m(N\leftrightarrow U)$ M
LdCyp	12.13 ± 0.38	-8.17± 0.92	6.72 ± 0.54	-5.89±0.65	1.47±0.06	1.16 ±0.04
V33A	9.36± 0.27	-7.74 ±0.81	5.13±0.71	-4.89±0.39	1.21± 0.09	1.05± 0.08
T55A	10.12±0.31	-7.39 ±0.57	5.87±0.51	-5.28±0.43	1.37±0.10	1.11 ± 0.07
L62A	9.01±0.19	-7.38 ±0.58	5.02±0.48	-5.02±0.33	1.22± 0.09	1.00±0.08
I85A	8.38±0.16	-7.68±0.47	4.33±0.41	-5.03±0.37	1.09±0.08	0.86±0.11
L120A	9.03±0.18	-7.34 ±0.63	5.02±0.48	-4.97±0.31	1.23± 0.07	1.01±0.06
F151V	6.97±0.23	-7.11±0.38	3.47±0.29	-4.51±0.46	0.98±0.11	0.77±0.23

4.3.3 Crystal structure of V33A and L120A

The crystal structures based on the data sets (**Table 5**) for both the mutants were solved by the molecular replacement method using AMoRe [**19**] from the CCP4 program suite, with the native crystal structure of cyclophilin from leishmania donovani (LdCyp) (pdb:2HAQ) as a search model. The rotation and translation function peaks with the lowest R-factor and the highest correlation coefficient were observed in space group $P4_3$ for both V33A and L120A with two molecules in the asymmetric unit. The R-factors for V33A and L120A upon including both

solutions from AMoRe were 31.3. A total of 692 (~5%) for V33A and 784 reflections (~5%) for L120A were randomly selected from their respective entire resolution ranges of 50-2.3 Å and 30.0-2.21 Å for the calculation of R_{free} [**20**]. After a few cycles of rigid body refinement using CNS version 1.1 [**21**], a 2Fo-Fc map (**Figure5**) was calculated and manual model building performed using the program O [**22**] and Coot v0.7.2.1 [**13**]. A few iterations of model building and refinement lowered the Rfactors for both mutants (V33A: R_{free}: 28.1, R_{work}: 24.2 and L120A: R_{free}: 27.1, R_{work}: 23.6).

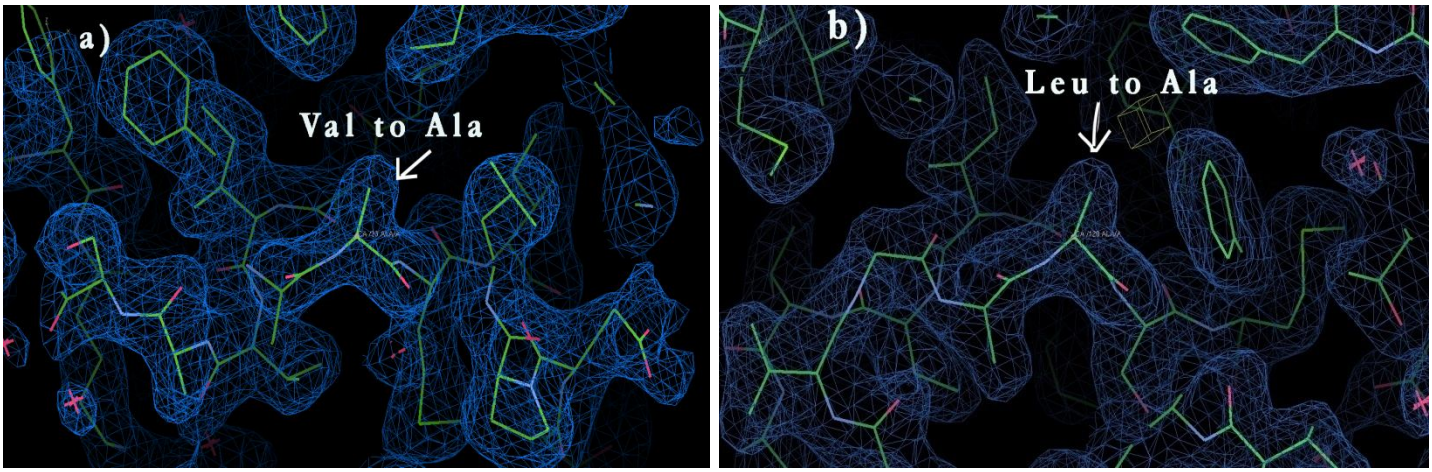


Figure 5: A 2Fo-Fc electron density map for mutants V33A and L120A with cutoff of 1σ with side chains for alanine in place of valine in a) V33A and leucine in b) L120A marked with arrows. Figures generated using Coot v0.7.2.1 [13].

At this stage water molecules were included in the refinement process and solvent molecules with unrealistic B factors (>60Å^2) or irregular electron density were excluded in the course of the refinement. Individual B factors were refined in the final stage with an initial value of 20 Å^2. At the conclusion of the refinement process R_{free}, R_{work} for the crystal structures of V33A and L120A were 22.15, 19.13 and 22.56, 17.56 respectively (**Table 8**). The number of water molecules for the structures were 109 (V33A) and 119 (L120A). PROCHECK [**23**] was used

for the validation of the final models. A total of 2562 and 2566 protein atoms were included in the final structures of V33A, L120A and the coordinates were deposited in the Protein Data Bank with pdb codes 4S1J and 4S1E respectively. For both the structures no residues were placed in the disallowed region of the Ramachandran plot, with the r.m.s deviations in bond lengths and bond angles being 0.008 Å, 1.5° and 0.007 Å, 1.5°. The overall B-factor from the Wilson plot and averaged from the coordinates were 36.40 $Å^2$; 33.83 $Å^2$ (V33A) and 32.50 $Å^2$; 30.85 $Å^2$ (L120A) and thus in fairly good agreement. The refinement and structure validation statistics are summarized in **Table 8** for both structures.

Table 8: Summary of refinement statistics for V33A and L120A

Statistics	**V33A**	**L120A**
PDB code	4S1J	4S1E
Sigma cut-off (*F*)	1.00	1.00
No. of reflections (working)	13028	15280
No. of reflections (test)	692	784
No. of protein atoms	2562	2566
No. of water molecules	109	119
R_{cryst}/R_{free} (%)	19.13/22.15	17.56/22.16
R.m.s. deviation from ideal values:		
Bond length (Å)	0.008	0.007
Bond angle (°)	1.50	1.50
Mean *B* Value ($Å^2$)	33.83	30.85
Ramachandran plot		
Most favored (%)	83.8	83.8
Additional allowed (%)	14.7	14.3
Generously allowed (%)	1.5	1.8

Native crystal structure of LdCyp (pdb: 2HAQ) was used to calculate side-chain torsion angles of residues in surface contact (**Materials and Methods**, Chapter3) with Val 33 and Leu 120 (**Table 9**) which have been mutated to Ala.

Table 9: Side chain torsion angles of residues in surface contact with Val 33 and Leu 120 calculated from native crystal structure of LdCyp (2HAQ). In case of V33A and L120Avalues of χ_1, χ_2 in the top row corresponds to values in chain A and bottom row represents that for chain B.

Residues	LdCyp		V33A		L120A	
	χ_1	χ_2	χ_1	χ_2	χ_1	χ_2
31 Phe	-70.161	-37.810	- 72.410 - 71.281	-41.123 -38.699	-71.596 -69.403	-41.207 -40.282
35 Ile	-59.723	-171.941	-52.361 -61.498	-172.910 -175.422		
40 Leu	-90.577	56.740	-93.138 -87.522	41.533 40.669		
43 Ile	-61.704	175.787	-63.103 -64.075	169.558 168.937		
45 Ile	-64.634	157.254			-66.857 -67.037	168.238 168.947
59 Phe	-164.843	81.273			-164.598 -164.684	74.857 75.397
76 Phe	-75.059	-80.113	-76.470 -79.203	-74.628 -71.536		
134 Phe	59.110	-84.532			62.825 63.136	-80.358 -81.112
136 Ile	-61.341	179.063			-66.027 -66.496	170.036 168.897
151 Phe	60.603	-82.778			58.421 58.042	-85.084 -82.976
154 Val	176.697				174.992 173.119	
160 Val	-179.084		-178.534 -177.822			
164 Ile	-66.059	171.241	-67.968 -65.619	175.737 179.980		
179 Val	179.002		173.392 176.633			
181 Ile	-67.057	171.820	-68.220 -66.305	-178.761 174.866		

In order to compare the change in the torsion angles of these residues the final coordinates of V33A and L120A were utilized for the determination of torsion angles for both chains (A and

B) in the asymmetric unit. A total of 9 and 7 residues were found to be in surface contact with V33A and L120A (**Table 9**).

A close inspection of the torsion angles form Table 9 suggests that in general there is a minor change of ~ 2° to 3° in the individual side chain torsions of the residues involved in surface contact with Val33 and Leu 120 as a consequence of the mutations. The most notable changes in torsion upon mutation from Val33 to Ala were observed for residues 35 Ile, 40Leu and 43 Ile by about 7-15°. In the second case of mutation of Leu 120 to Ala, the torsion angle of the residues that suffers a considerable change are that of 45Ile, 59Phe and 134Phe with a deviation in the range of 5-13°.

4.3.4 Surface Complementarity Scores

The difference in side chain surface complementarity (ΔS_m : See **Materials and Methods**) was computed from the native and mutant crystal structures (or modeled mutant structures). With the exception of I85A, all mutations had negative ΔS_m values which indicated a deleterious disruption in the native packing. The most negative ΔS_m value was observed for F151V (**Table 10**) whilst values for T55A and I164A indicated a comparatively lesser disruptive effect in the native packing as consequence of mutations, reflected in the relatively less negative ΔS_m values. Barring the anomalous value of ΔS_m obtained for I85A the correlation coefficient between ΔS_m for the mutants and the RMSD's , (between the native and mutant/modeled structures) calculated for all side chain atoms belonging to residues which were in surface contact with the mutated residue (as given **Table 1**) was around 0.68.

.

Table 10: Side chain ΔS_m values of cyclophilin mutants along with the RMSD between the coordinates of the native crystal structure (2HAQ) and mutant crystal or modeled structures. Only side chain atoms of the residues in the immediate neighbourhood that is in surface contact of the mutated residue in the native crystal structure (as given in Table 1) were used in the RMSD calculation. Calculations from the experimentally determined crystal structures are marked in bold.

Mutations	ΔS_m (sidechain)	RMSD (Å)
V33A	-9.39	0.76
T55A	-3.62	0.61
L62A	-12.04	0.73
I85A	+4.13	0.97
L120A	-8.46	0.74
F151V	-13.59	0.87
I164A	-3.96	0.69

4.4 Conclusion

All mutants were thermally destabilized with respect to the native protein. Amongst the mutants the most stable mutant was T55A for both DSC and GdmCl meditaed unfolding monitored by fluorescence. The least stable mutant as characterized by DSC was I85A followed by F151V whereas their positions were interchanged in terms of conformational stability (fluorescence). ΔS_m for all the mutants appeared to indicate deleterious disruptions in side chain packing with the sole exception of I85A. However, this result will have to be experimentally verified as I85A was not an experimentally determined crystal structure but was computationally modeled. The RMSD's of mutant side chains atoms (with respect to native) in the neighbourhood of the mutation (**Table 10**) appeared to correlate reasonably well with ΔS_m values.

References:

1. P. Haney, J. Konisky, K.K. Koretke, Z. Luthey-Schulten, P.G. Wolynes (1997) *Proteins* 28, 117-130.

2. R.J.M Russell, J.M.C Fergusson, D.W. Haugh, M.J. Danson, G.L. Taylor (1997) *Biochemistry* 36, 9983-9994.

3. T.Salminen, A. Teplyakov, J. Kankare, B.S. Cooperman, R. Lahti, A. Goldman (1995) *Prot. Sci.* 5, 1014-1025.

4. H. Zuber (1988*) Biophys Chem*. 29, 170-179.

5. R.J.M Russell, U. Gerike, M.J. Danson, D.W. Hough, G.L. Taylor (1998) *Structure*, 6, 351-361

6. K. Watanabe, Y. Hata, H. Kizaki, Y. Katsube, Y. Suzuki (1997) *J. Mol. Biol*., 269, 142-153.

7. Y. Bogin, M. Peretz, Y. Hacham, Y. Korkhin, F. Frolow, A.J. Kalb, Y. Burstein (1998) *Prot. Sci*. 7, 1156-1163.

8. G. Vogt, P. Argos (1997) *Fold Des*. 2, S40-S46.

9. G. Vogt, S. Woell, P. Argos (1997) *J.Mol.Biol*. , 269, 631-643.

10. K.S.P. Yip, (1995) Structure, 3, 1147-1158.

11. K.S.P. Yip, K.L. Britton, T.J. Stillman, J. Lebbink, W.M. De Vos, F.T. Robb, C. Vetriani, D. Maeder, D.W. Rice (1998) *Eur. J. Biochem., 255, 5759-5765.*

12. A.H. Elcock (1998) *J. Mol. Biol*. 284, 489-502.

13. P. Emsley, B. Lohkamp, W.G. Scott,K. Cowtan (2010) *Acta Cryst. D*66, 486-501.

14. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus (1983). *J Comp Chem* 4 (2): 187–217

15. R. Banerjee, M. Sen, D. Bhattacharyya, P. Saha (2003) *J.Mol.Biol* 333, 211–226

16. K. Kleppe, E. Ohtsuka, R. Kleppe, I. Molineux, H.G. Khorana (1971) *J. Mol. Biol.* 56 (2): 341–361.

17. H.C. Birnboim, J. Doly (1979) *Nucleic Acids Res.* 7 (6): 1513–1523.

18. Z. Otwinowski, W. Minor, (1997) *Methods Enzymol*. 276, 307-374.

19. J.Navaza (1994). *Acta .Cryst. A* 50, 157-163.

20. A.T. Brünger (1992) *Nature (London)*, 355, 472-475.

21. A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G. L. Warren (1998). *Acta. Cryst*. *D* 54, 905-921

22. T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard (1991). *Acta Cryst*. A 47, 110-119.

23. R. A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton (1993). *J. Appl*. *Cryst*. 26, 283-291.

International Journal of Biological Macromolecules 69 (2014) 353–360

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

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

Equilibrium unfolding of cyclophilin from *Leishmania donovani*: Characterization of intermediate states

Sourav Roy[a,1], Sankar Basu[a,1], Alok K. Datta[b], Dhananjay Bhattacharyya[a,1], Rahul Banerjee[a,*], Dipak Dasgupta[a,*]

[a] Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, West Bengal, India
[b] Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Kolkata 700032, West Bengal, India

ARTICLE INFO

Article history:
Received 7 March 2014
Received in revised form 20 May 2014
Accepted 24 May 2014
Available online 2 June 2014

Keywords:
Cyclophilin
Intermediate
Protein unfolding

ABSTRACT

Cyclophilin from *Leishmania donovani* (LdCyp) is a ubiquitous peptidyl-prolyl cis-trans isomerase involved in a host of important cellular activities, such as signaling, heat shock response, chaperone activity, mitochondrial pore maintenance and regulation of HIV-1 infectivity. It also acts as the prime cellular target for the auto-immune drug cyclosporine A (CsA). LdCyp is composed of a beta barrel encompassing the unique hydrophobic core of the molecule and is flanked by two helices (H1, H2) on either end of the barrel. The protein contains a lone partially exposed tryptophan. In the present work the equilibrium unfolding of LdCyp has been studied by fluorescence, circular dichroism and the non-coincidence of their respective C_m's, indicates a non-two state transition. This fact was further corroborated by binding studies of the protein with bis-ANS and the lack of an isochromatic point in far UV CD. The thermal stability of the possible intermediates was characterized by differential scanning calorimetry. Further, MD simulations performed at 310, 400 and 450 K exhibited the tendency of both helices to partially unwind and adopt non-native geometries with respect to the core, quite early in the unfolding process, in contrast to the relatively stable beta barrel.

1. Introduction

Leishmaniasis, a broad spectrum of diseases caused by *Leishmania* spp. is widely prevalent in third world countries, among the poorer sections of the populace. The appearance of strains resistant to drugs (pentavalent antimonials) [1], traditionally used as the first line of defense against the pathogen, stress the need to continue the search for alternative drug targets, as second line drugs are generally expensive and have reportedly severe side effects [2]. Cyclophilin from *Leishmania donovani* (LdCyp) belongs to the ubiquitous class of peptidyl-prolyl cis-trans isomerases (PPIases), also known to be the intracellular receptor of the immunosuppressive drug cyclosporine A (CsA), a cyclic undecapeptide constituted of non-natural amino acids. CsA derivatives formed by specific modification of its side chains have been shown to lack immunosuppressive activity though retaining its anti-trypanosomatid and anti-parasitic character [3,4]. In addition, cyclophilins have also been implicated in signal transduction, cell division, cell surface recognition, chaperone activity and heat shock response [5]. Other studies show Cyp's to be involved in the maintenance of mitochondrial transition pores [6] and regulation of HIV-1 infectivity by functional association with HIV-1 virions in humans [7].

Several crystal and NMR structures of CyPs in both ligated/unligated forms are currently available [8,9]. Cyclophilin (LdCyp) has been used as the model system in the present study. The three dimensional structure of LdCyp (2HAQ: 1.97 Å) [10] consists of an 8 stranded ß-barrel with two α helices located at either end. The location of the helices with respect to the barrel effectively blocks the solvent accessibility of the only hydrophobic core of the molecule, located in the interior of the barrel. The single domain molecule consists of a single cysteine residue and its enzymatic activity is hindered by the binding of CsA as there is considerable overlap between the native active site of the enzyme and the binding site for CsA, both being located on the face of the barrel. LdCyp consists of a single tryptophan (Trp 143 also involved in its active site) which is situated on a flexible 3_{10} helix (141–145) and is partially exposed to the solvent.

Given the intrinsic interest of Cyp's in general and specifically in *Leishmania*, LdCyp has been selected for unfolding studies, both thermally and with denaturant. There are reports of thermal

* Corresponding authors. Tel.: +91 33 2337 5345; fax: +91 33 2337 4637.
E-mail addresses: rahul.banerjee@saha.ac.in (R. Banerjee), dipak.dasgupta@saha.ac.in (D. Dasgupta).
[1] Tel.: +91 33 2337 5345; fax: +91 33 2337 4637.

http://dx.doi.org/10.1016/j.ijbiomac.2014.05.063
0141-8130/

denaturation studies on cyclophilin from mycobacterium tuberculosis both in the presence and absence of CsA. CsA binding did not appear to impart any extra stabilization to the protein, though facilitated the formation of secondary structural elements at lower temperatures (25 °C). The PPiA (cyclophilin A from *Mycobacterium tuberculosis*) tended to aggregate beyond 70 °C and unfolding induced via denaturant demonstrated that the loss in tertiary structure preceded the loss in secondary structure both in the presence and absence of CsA [11].

In the present study we probed the unfolding of LdCyp by guanidium chloride (GdmCl) by means of fluorescence, circular dichroism (CD), dynamic light scattering (DLS) spectroscopy and differential scanning calorimetry (DSC). The results from the studies indicate that most probably the unfolding of LdCyp proceeds via at least one equilibrium intermediate. We have made an attempt to structurally characterize this possible intermediate(s) using multiple spectroscopic tools and molecular dynamics simulations.

2. Materials and methods

2.1. Reagents and chemicals

Guanidium chloride (GdmCl), bis-anilino-8-napthalenesulphonate (bis-ANS) were obtained from Sigma–Aldrich (USA). Potassium dihydrogen orthophosphate (KH_2PO_4) and anhydrous dibasic potassium phosphate (K_2HPO_4) were purchased from Sisco Research Laboratories (SRL) and Nickel-nitrilotriacetic acid-agarose, superflow (Ni-NTA) from Qiagen (USA). All other chemicals used were of analytical grade.

2.2. Protein purification

As reported previously, residues 22–187 of LdCyp has been cloned into PQE32 (Qiagen) expression vector (with an additional 6x-His tag at the N-terminal), over expressed in M15 *Escherichia coli* cells and purified to homogeneity using a Ni-NTA column. The purified protein sample was dialyzed extensively against a buffer containing 0.02 M Tris (pH 8.5) and 0.02% NaN_3 and stored at 4 °C. Prior to all spectroscopic and calorimetric experiments the protein was extensively re-dialyzed in 25 mM potassium phosphate buffer (pH 7.5).

2.3. Fluorescence measurements

Intrinsic tryptophan fluorescence of the protein was measured at 25 °C on a Perkin Elmer LS55 spectrofluorimeter (Perkin Elmer Ltd., UK). LdCyp at 4 μM in 25 mM potassium phosphate buffer (pH 7.5) was excited at 295 nm and the emission spectra recorded in the range 310 nm to 450 nm with both excitation and emission slit widths of 5 nm. The spectra recorded at a scan speed of 200 nm/min were averaged over 4 scans. Appropriate controls for the background emission were subtracted in each case.

Fluorescence quenching of the lone tryptophan residue by neutral quenchers serves to indicate the degree of solvent exposure of the tryptophan residues in a protein. Quenching experiments with the neutral quencher acrylamide was performed for LdCyp (4 μM) pre-incubated with the denaturant. The acrylamide concentration in the protein sample was varied from (10 to 50 mM) and fluorescence spectra measured ranging from 310 to 450 nm, with excitation at 295 nm. Stern–Volmer quenching constant was calculated as the slope of the plot of F_0/F values at 360 nm against the input concentration of quencher, acrylamide in the Stern–Volmer equation:

$$\frac{F_0}{F} = 1 + K_{SV}[Q] \quad (1)$$

where K_{SV} is the Stern–Volmer constant and [Q] is the concentration of the quencher. All quenching experiments were repeated thrice.

Binding of hydrophobic fluorescent dye, bis-ANS (bis-anilino-8-napthalenesulphonate), with LdCyp was carried out to probe the intermediate conformational states of LdCyp (4 μM, 25 mM potassium phosphate buffer, pH 7.5) in varying concentrations of the denaturant (GdmCl, 0.0–3.0 M). The protein along with the denaturant was equilibrated with bis-ANS for 20 min before recording the spectra. Ensuring complete saturation of the dye binding sites in the native protein the concentration of bis-ANS in the samples was kept fixed at 10 μM. Bis-ANS emission spectra was recorded from 410 to 675 nm, with excitation at 395 nm and both emission and excitation slit widths were kept at 5 nm. Appropriate buffer corrections were done.

2.4. Circular dichroism measurements

The CD spectra of LdCyp were recorded on a JASCO J715 (Jasco Corporation, Tokyo, Japan) in a rectangular quartz cell of path length 1 mm. Far UV measurements of LdCyp were recorded in the wavelength range of 200–250 nm For all CD experiments in 200–250 nm region, spectra were obtained by averaging over 4 scans, utilizing a slit width of 1 nm. For protein unfolding with chemical denaturants far UV spectra were recorded from 210 to 250 nm in case of GdmCl concentrations exceeding 1.0 M. Near UV spectra were recorded in the region 250–300 nm to probe the denaturant induced alteration of tertiary structure of the protein. Buffer contributions alone or with denaturants were subtracted from all protein spectra.

The observed values (after appropriate buffer corrections) were converted to mean residue ellipticity (MRE) in deg cm^2 $dmol^{-1}$ defined as

$$\text{MRE} = \frac{M\theta_\lambda}{10dcr} \quad (2)$$

where M is the molecular weight of the protein (in Da), θ_λ is the CD in millidegree, d is the path length in cm, c is the protein concentration in mg/ml and r the number of amino acid residues in the protein polypeptide chain.

2.5. Differential scanning calorimetry

10 μM of LdCyp was scanned on a VP-DSC Microcalorimeter (Microcal LLC, Northampton, MA, USA) from 10 °C to 70 °C with a scan rate of 30 °C/h and at approximately 28 psi pressure. All samples were extensively degassed prior to loading. Initially, the buffer (25 mM potassium phosphate, pH 7.5) or buffer combined with denaturant (GdmCl, 0.2–1.4 M) were repeatedly scanned to ensure a stable baseline. The buffer baseline was subtracted from the protein thermogram using Microcal Origin version 7.0 provided with the instrument, subsequent to normalization of the data by the protein concentration a non-linear curve fitting algorithm was employed to obtain the thermodynamic parameters of the transition. The reversibility of unfolding was checked at different scan rates, buffer conditions and temperature ranges. All experiments were repeated at least three times.

2.6. Guanidium chloride (GdmCl) mediated unfolding

Equilibrium unfolding of LdCyp was induced by 16 h of incubation with varying concentrations of GdmCl (0.025–3.0 M) at 25 °C. For refolding experiments, the unfolded protein with 3.0 M GdmCl was extensively dialyzed for 3 h against separate buffers containing 2.0, 1.5, 1.0, 0.5, 0.2 M GdmCl and was subsequently equilibrated again for 2 h. Fluorescence and CD spectra under different

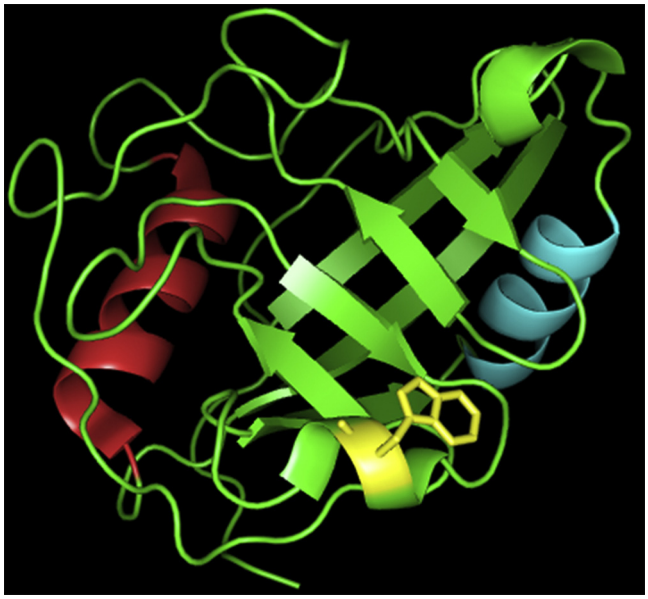


Fig. 1. Native crystal structure of cyclophilin from *Leishmania donovani* (LdCyp) in green (cartoon representation) with the partially solvent exposed lone tryptophan residue Trp 143 (in yellow, ball and stick mode). Helix H1 is colored red and helix H2 colored cyan. The figure is generated using Pymol.

conditions were then recorded as described above. The increase in GdmCl concentration in the buffer caused a negligible change in the pH of the solution.

2.7. Molecular dynamics simulations

Unfolding molecular dynamics simulations were performed on cyclophilin (coordinates obtained from: 2HAQ) at temperatures 310, 400 and 450 K. Each simulation run was for a duration of 50 ns. The simulation at 310 K could be considered to be 'native'. Initially, the protein was solvated in a cuboidal box of dimensions 78.683, 68.897 and 78.137 Å by the addition of 11,088 waters following the TIP3P model and charge neutralization of the system was accomplished by the addition of Na^+ ion utilizing the xleap module of AMBER [12]. The structure was then energy minimized (AMBER 2002 force field) for 200 steps of steepest descent followed by 19,800 steps of ABNR incorporated in the SANDER module. All simulations were performed using NAMD 2.0 [13]. For each 50 ns run the targeted temperature was attained in increments of 10 K/ps for a NPT ensemble with the Langevin piston set to the targeted temperature (310, 400, 450 K, etc.) and the pressure fixed at 1.032 bar. The bond lengths were constrained at a tolerance level of 0.005 Å by the SHAKE [14] algorithm. Visual molecular dynamics program (VMD 1.9.1) [15] was used to view the trajectories obtained from NAMD. The time step (δt) for the velocity–verlet algorithm was kept fixed at 1 fs. Snapshots were sampled at intervals of 10 ps. Secondary structural content for the snapshots were calculated using STRIDE [16].

3. Results and discussion

3.1. Unfolding of LdCyp with GdmCl

As mentioned previously, LdCyp contains a single tryptophan residue (Fig. 1). The maximum intensity of intrinsic fluorescence, λ_{max}, was observed at 348 nm for the native protein suggestive of the partial solvent exposure of the lone tryptophan. The emission spectrum changed with the variation of denaturant concentration till 1.4 M GdmCl. Further increase in the concentration of the

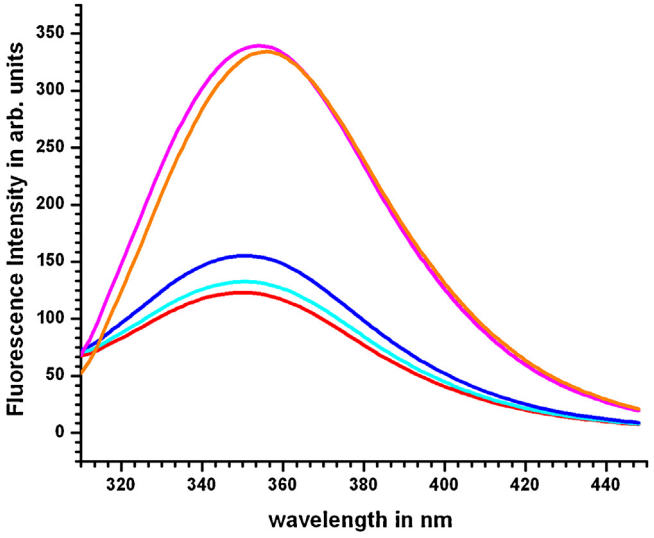


Fig. 2. Fluorescence unfolding spectra of native cyclophilin at various GdmCl concentrations ranging from (0 to 3) M. (a) ( ) 4 μM LdCyp, (b) ( ) 4 μM LdCyp + 0.8 M GdmCl, (c) ( ) 4 μM LdCyp + 1.0 M GdmCl, (d) ( ) 4 μM LdCyp + 1.4 M GdmCl, (e) ( ) 4 μM LdCyp + 3.0 M GdmCl.

denaturant led to a slight decrease in the intensity (Fig. 2). We plotted the intensity at 360 nm as a function of GdmCl concentration (Fig. 3). With increase in GdmCl concentration unfolding was initiated around 1.0 M and completed by 2.0 M, as clearly indicated by the saturation in fluorescence intensity.

The unfolding curve of LdCyp fitted best to a three state equation originating from the following equilibrium: $N \rightleftharpoons I \rightleftharpoons U$

$$S_{obs} = \frac{S_N + S_I e(-\Delta G_{NI}/RT) + S_U e(-\Delta G_{NU}/RT)}{1 + e(-\Delta G_{NI}/RT) + e(-\Delta G_{NU}/RT)} \quad (3)$$

where S_{obs} is the signal at any denaturant concentration, S_N is the signal due to the native state, S_I the signal due to the intermediate state and S_U the signal due to the unfolded state. ΔG_{NI} and ΔG_{NU} are the corresponding free energies of the transitions from $N \rightleftharpoons I$ and $N \rightleftharpoons U$ states [17] where ΔG_{NI} and ΔG_{NU} are assumed to have

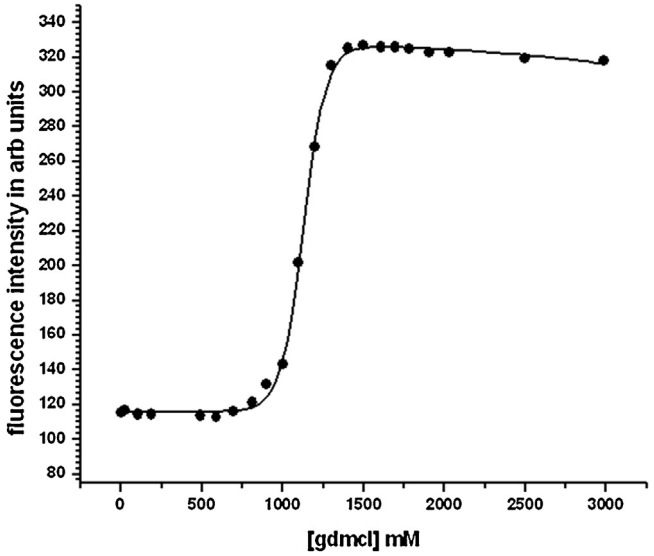


Fig. 3. Unfolding curve of LdCyp induced by GdmCl as monitored by intrinsic tryptophan fluorescence with emission maxima points at 360 nm, fitted to a 3-state equation. The black circles (•) indicate the fluorescence intensity in arbitrary units at the corresponding denaturant concentrations (in mM) and ( ) is the fitted line to the data points.

Table 1
Equilibrium thermodynamic parameters of unfolding for LdCyp mediated by GdmCl.

Spectroscopic tools employed	Thermodynamic parameters					
	$\Delta G_{NI}(H_2O)$ (kcal mol^{-1})	m_{NI} (kcal mol^{-1} M^{-1})	$\Delta G_{NU}(H_2O)$ (kcal mol^{-1})	m_{NU} (kcal mol^{-1} M^{-1})	$C_m(N \leftrightarrow I)$ (M)	$C_m(N \leftrightarrow U)$ (M)
Intrinsic tryptophan fluorescence	12.13 (±0.38)	−8.17 (±0.92)	6.72 (±0.54)	−5.89 (±0.65)	1.47 ± 0.06	1.16 ± 0.04
Far UV circular dichroism	2.12 (±0.13)	−2.18 (±0.29)	5.24 (±0.45)	−3.47 (±0.43)	0.99 ± 0.02	1.51 ± 0.04

The thermodynamic parameters in this table were obtained from the 3-state equation, Eq. (3) by employing a non-linear least square algorithm. $\Delta G_{NI}(H_2O)$ and $\Delta G_{NU}(H_2O)$ refers to the free energy of transition from the native to the intermediate state and native to the unfolded state, respectively, at zero denaturant concentration with ΔG_{NI} and ΔG_{NU} going to zero. m_{NI} and m_{NU} are the corresponding slopes of the transition from the native to the intermediate and native to the unfolded states, $C_m(N \leftrightarrow I)$ and $C_m(N \leftrightarrow U)$ are the midpoints of transition (denaturant concentration) obtained from Eqs. (4) and (5), where ΔG_{NI} and ΔG_{NU} goes to zero.

linear dependence with the denaturant concentration [D], resulting in the following equations,

$$\Delta G_{NI} = \Delta G_{NI}(H_2O) + m_{NI}[D] \quad (4)$$

$$\Delta G_{NU} = \Delta G_{NU}(H_2O) + m_{NU}[D] \quad (5)$$

where m_{NI} and m_{NU} are the corresponding slopes of transitions from the native to the intermediate and native to the unfolded states respectively. C_m for any unfolding curve is defined as the midpoint of the transition from the native to the unfolded state ($C_m(N \leftrightarrow U)$: for a 2-state transition) or between the native to intermediate and native to the unfolded state, respectively (that is $C_m(N \leftrightarrow I)$, $C_m(N \leftrightarrow U)$ for a three state transition) obtained from Eqs. (4) and (5) when the corresponding free energies of transition from the native to the intermediate state ΔG_{NI} and from the native to the unfolded state ΔG_{NU} becomes zero. The above fit yielded $\Delta G_{NI}(H_2O)$ of 12.13 kcal mol^{-1} with m_{NI} of −8.166 kcal mol^{-1} M^{-1} and $\Delta G_{NU}(H_2O)$ of 6.72 kcal mol^{-1} and m_{NU} of 5.89 kcal mol^{-1} M^{-1} (Table 1) which suggests that the unfolding of N to I exposes more solvent accessible surface area than the unfolding of N to U which is also clear from other experiments like (acrylamide quenching and bis-ANS fluorescence studies).

The far UV CD of the native protein exhibited a pronounced minimum at 225 nm characteristic of a protein with prevalently β-sheets. With the addition of GdmCl, MRE value at 222 nm progressively reduced from 0.8 to 2.0 M, saturating thereafter. This was also accompanied by a gradual and concomitant reduction in the overall secondary structural content till 1.2 M GdmCl, after which there was an abrupt loss of all secondary structural features in the protein (Fig. 4). Thus the CD signature of secondary structural feature was observed till 1.2 M GdmCl. Similar to the feature obtained from fluorescence titration profile (Fig. 3), the unfolding curve from CD value at 222 nm was best fitted to the above three state equation yielding ΔG_{NI} (H_2O) and C_m1 of 2.12 (±0.13) kcal mol^{-1} and 0.97 (±0.11) M and ΔG_{NU} (H_2O) of 5.24 (±0.16) kcal mol^{-1} and C_m2 of 1.51 (±0.2) M, respectively (Fig. 5). In addition, CD unfolding curves at MRE_{225} and MRE_{215} (Supplementary Information, Fig. S1 and Table S1a, S1b, S1c) gave similar C_m values though some deviation in the curves were observed between MRE_{222}, MRE_{225} on one hand and MRE_{215} on the other.

The lack of superposition of the unfolding curves obtained by two independent probes CD and fluorescence and the consequent non-equivalence of their respective C_m values (Table 1) appears to be supportive of a non-two state transition and thereby indicates the presence of a possible intermediate in the unfolding of LdCyp: N ⇌ I ⇌ U (see Supplementary Information, Fig. S2). There are numerous instances in the literature where non-coincidence of C_m values in equilibrium unfolding experiments [17–20] via two independent probes has been cited as evidence for the presence of an equilibrium intermediate. The adoption of a three state model improved the fitting in the protein unfolding curve monitored by fluorescence, when compared to a 2 state model (see

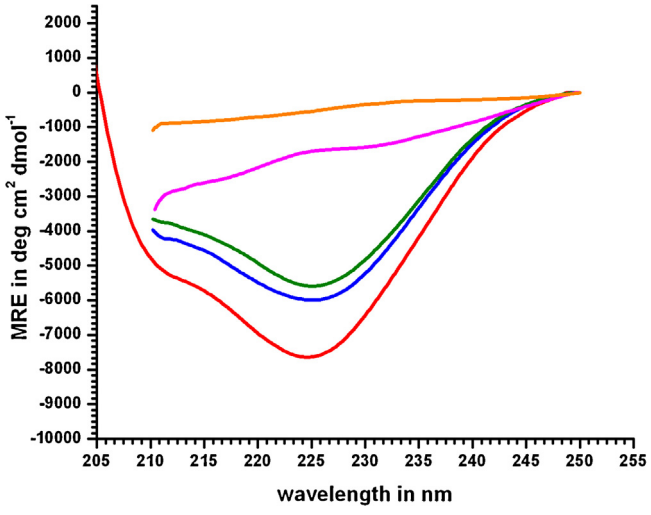


Fig. 4. Far-UV CD spectra of LdCyp (native) along with different denaturant concentrations. (a) ( ) 4 μM LdCyp (b) ( ) 4 μM LdCyp + 1 M GdmCl, (c) ( ) 4 μM LdCyp + 1.2 M GdmCl, (d) ( ) 4 μM LdCyp + 1.4 M GdmCl, (e) ( ) 4 μM LdCyp + 3.0 M GdmCl.

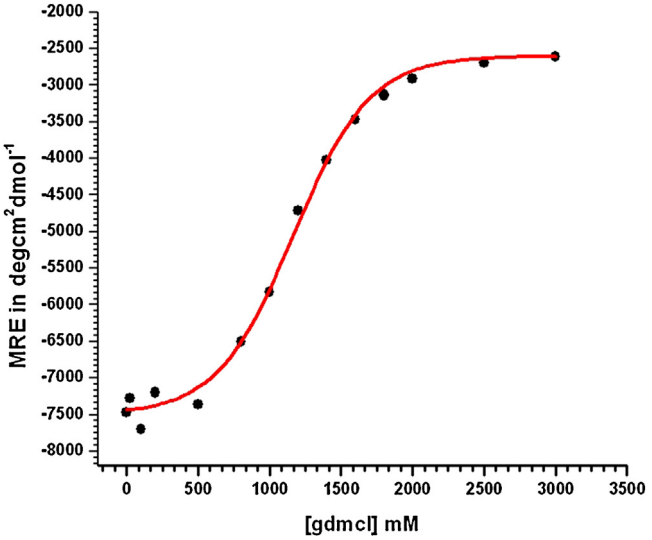


Fig. 5. MRE_{222} [in black circles (•)] values obtained from far-UV CD spectra plotted as a function of the denaturant and fitted to a 3-state equation.

Supplementary Information, Fig. S3). However, for unfolding curves monitored by CD (MRE_{222}) the improvement was only marginal (3-state compared to 2-state) (see Supplementary Information, Fig. S4). Notably, a 2-state transition could also be considered adequate in protein unfolding probed by CD. In any case for both fluorescence, CD the improved fitting (3-state versus 2-state) was only mildly reflected in their respective R^2 values. However, three state fittings were retained for both fluorescence and CD for consistency with bis-ANS fluorescence intensity (recorded at 488 nm) curves as function of GdmCl concentration and also because of the lack of an isochromatic point in the far UV CD data (ranging from 210 to 250 nm), generally indicative of a non-two state transition. Near UV CD (250–300 nm) spectra of native LdCyp show a broad shoulder in the region of 260–285 nm. CD spectra of LdCyp incubated with low concentration of GdmCl (0.2 M) indicate that the native tertiary interactions are relatively preserved. In contrast LdCyp incubated with 1.0 M and 1.2 M of GdmCl shows considerable perturbation in the native tertiary contacts as manifested by the reduced intensity of the broad shoulders (Fig. 6). Absence of tertiary interaction at high GdmCl concentration is evident from the featureless spectra with a significant reduction in band intensity. Summing up, the results from CD studies suggest that LdCyp thus incubated with 1.0 and 1.2 M of GdmCl shows considerable disruption in the original tertiary interactions whilst conserving their secondary structure.

To further characterize the intermediate hydrophobic dye binding studies with bis-ANS were performed.

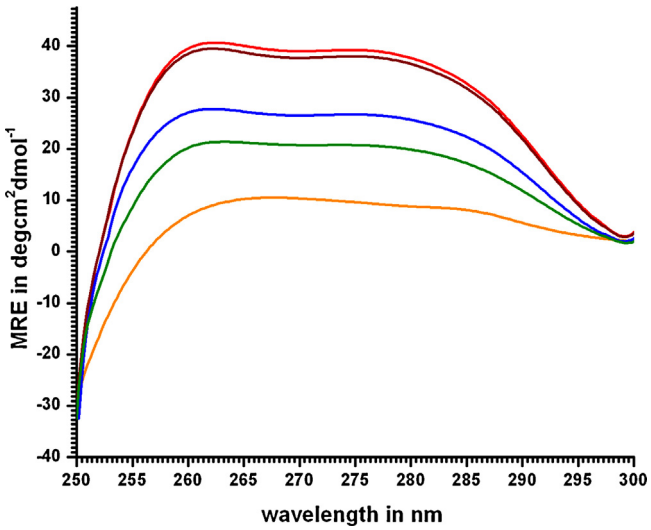


Fig. 6. Near UV CD spectra of LdCyp for (a) native protein LdCyp ( ) 40 μM, (b) LdCyp + 0.2 M GdmCl ( ) maintaining a characteristic broad shoulder which considerably flattens out for (c) LdCyp + 1.0 M GdmCl ( ), (d) LdCyp + 1.2 M GdmCl ( ), (e) LdCyp + 3.0 M GdmCl ( ), thereby depicting the loss of tertiary interactions with increasing denaturant concentrations.

3.2. Bis-ANS fluorescence experiment

Bis-ANS fluorescence studies have been widely used to identify intermediate states in protein unfolding [21,22]. Natively folded or completely unfolded proteins are unable to bind to bis-ANS, as their hydrophobic cores are effectively shielded in the former, whereas in the latter dissolution of the core negates any dye binding even though the hydrophobic residues are solvent exposed. Thus the partial exposure of erstwhile buried hydrophobic residues characteristic of "molten globules" are ideal for bis-ANS interaction. For LdCyp, bis-ANS fluorescence intensity (recorded at 488 nm) initially rose with GdmCl concentration upto 1.2 M and declined thereafter with a minima at 3.0 M (Figs. 7 and 8).

Both of these observations are indicative of a molten globule like intermediate in the equilibrium unfolding of LdCyp.

3.3. Acrylamide quenching

Stern–Volmer coefficients estimated for different denaturant concentrations (see Section 2.3) yielded a K_{SV} value of 7.64 M^{-1} for the native protein indicating partial solvent exposure of its lone tryptophan. K_{SV} increased to 13.18 M^{-1} at 1.4 M concentration of

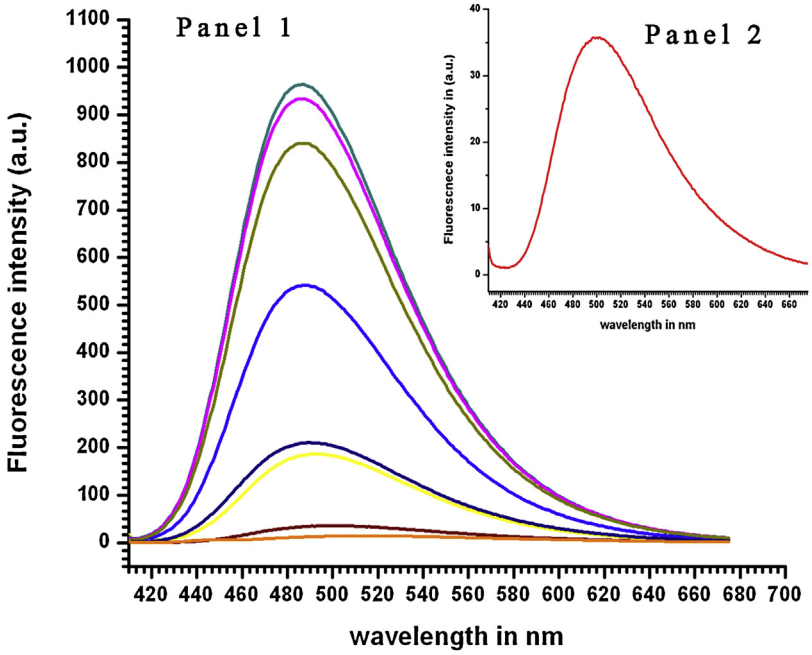


Fig. 7. Panel 1: Bis–ANS fluorescence (10 μM) spectra for LdCyp (4 μM) incubated with various GdmCl concentrations, panel 1: (a) ( ) 4 μM LdCyp + 0.2 M GdmCl (b) ( ) 4 μM LdCyp + 0.7 M GdmCl (c) ( ) 4 μM LdCyp + 1.0 M GdmCl (d) ( ) 4 μM LdCyp + 1.2 M GdmCl (e) ( ) 4 μM LdCyp + 1.4 M GdmCl, (f) ( ) 4 μM LdCyp + 1.5 M GdmCl (g) ( ) 4 μM LdCyp + 1.9 M GdmCl (h) ( ) 4 μM LdCyp + 3.0 M GdmCl. Panel 2: The native spectra ( ) 4 μM LdCyp is shown in Panel 2.

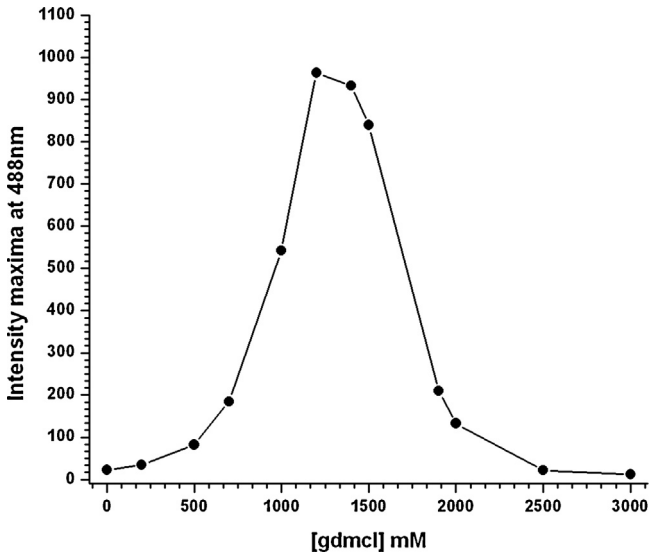


Fig. 8. Plot of Bis–ANS fluorescence (10 μM) intensity maximum in black circles (•) at maximum blue shifted wavelength of 488 nm plotted as a function of GdmCl concentration, with maximum peak intensity obtained for protein sample (4 μM) incubated with 1.2 M GdmCl.

GdmCl after which the change is not significant (see Table 2) at higher denaturant concentrations suggesting maximum exposure of the said tryptophan to occur by 1.4 M. The dynamic character of the quenching was also confirmed by repeating the experiment at a higher temperature at 45 °C, which gave Stern–Volmer coefficients different from those at 25 °C (Supplementary Information, Fig. S5 and Table S2).

Table 2
Dependence of K_{SV} values on denaturant concentrations.

Denaturant concentration in M	K_{SV} values in M^{-1}
0.0	7.64 ± 0.23
0.2	8.84 ± 0.87
1.2	11.65 ± 0.94
1.4	13.18 ± 1.05
3.0	12.87 ± 1.11

The K_{SV} values in this table were calculated using the Stern–Volmer equation, Eq. (1) by linear least square fitting.

3.4. Differential scanning calorimetry

In order to thermally characterize the intermediate states, differential scanning calorimetry was performed with the native protein and also in the presence of different denaturant concentrations. For the native protein reversibility was observed only till 54.6 °C, despite being tested on a wide range of buffering conditions and scan rates. The DSC thermogram of native LdCyp could be best fitted to a non-two state process with two T_m's at 49.6 °C and 54.6 °C. 'Different scan rates (at 20 and 60 °C/h) did not significantly alter the position of the peak in Cp(T), the shape of the curve in the thermogram or the value of the two T_m's (Supplementary information. Figs. S6, S7 and Table S3), confirming that the shoulder observed in the theremograms were unlikely due to kinetic effects.' The apparent ΔH of the transitions were 100 kcal mol^{-1} and 117 kcal mol^{-1}, respectively. DSC measurements under identical conditions with the addition of 0.2 M GdmCl gave a similar thermogram, but with reduction in both T_m's by about 5 °C (44.9 and 50.6 °C) and a concomitant decrease in their ΔH (75.2 and 104.7 kcal mol^{-1}), respectively. Although

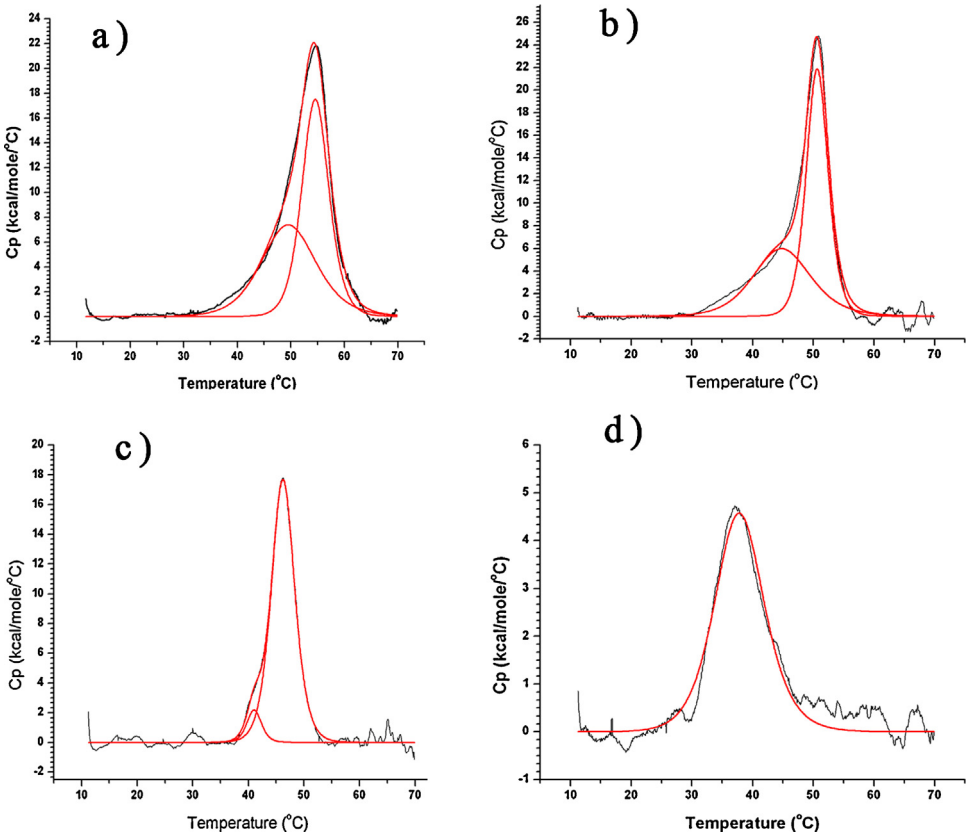


Fig. 9. Thermal characterization of the intermediate states of LdCyp through DSC (a) 10 μM LdCyp, (b) 10 μM LdCyp + 0.2 M GdmCl, (c) 10 μM LdCyp + 0.6 M GdmCl, (d) 10 μM LdCyp + 1.0 M GdmCl, two step (two T_m values) to a single step (one T_m value) transition. The black solid line (▬) corresponds to the experimental data points while the solid red line (▬) are the fitted lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 3
The fraction of native contacts conserved between the residues constituting the helices (H1, H2) and the remaining residues in the core, averaged over 50 ns for each MD simulation run at 310, 400 and 450 K, respectively. The standard deviation is given in parentheses. The fractional conservation of secondary structural content with respect to the native crystal structure is also given.

Temperature (K)	Fraction of native contacts conserved		Fraction of secondary structural content		
	Helix H1 – core	Helix H2 – core	Helix H1	Helix H2	Beta strands
310	0.73(0.16)	0.65(0.23)	0.92(0.02)	0.89(0.01)	0.97(0.02)
400	0.65(0.18)	0.47(0.29)	0.92(0.03)	0.60(0.27)	0.86(0.05)
450	0.48(0.21)	0.37(0.30)	0.70(0.28)	0.51(0.29)	0.81(0.12)

the non-two state character (with two T_m's) of the thermogram was still maintained at 0.6 M GdmCl, yet there was a significant reduction in the apparent enthalpy ΔH of the first transition (T_m = 41.0 °C, ΔH = 7.7 kcal mol^{-1}) relative to the second (T_m = 46.3, ΔH = 98 kcal mol^{-1}), with further decrease in the two T_m's. At 1.0 M GdmCl a single transition was observed at T_m = 37.9 °C and ΔH = 49.5 kcal mol^{-1} (Fig. 9). At still higher concentrations (1.4 M and above) no statistically significant peak was observed in the DSC measurement most probably indicating the initiation of the molten globule like intermediate, mentioned previously.

3.5. Molecular dynamics simulations

LdCyp consists of a single extended hydrophobic core, whose constituent residues are primarily contributed by the beta strands composing the barrel (see Fig. 1), the two helices (H1, H2) and a few from surrounding loops. Initially, three sets of residues were constructed; S1: residues constituting helix H1; S2: residues constituting helix H2 and S3: residues of the hydrophobic core excluding those from helices H1, H2 contributed by beta strands and loops (Supplementary Information, Table S4). To study the disposition of the helices with respect to the barrel (upon raising temperature in an MD simulation; see Section 2.7), the contacts (see Supplementary Table S5) between S1–S3 and S2–S3 were first calculated from the crystal structure (2HAQ). 4.0 Å was considered to be the distance cutoff for two atoms to be in contact. The fractional conservation of these native contacts (between the helices and the remaining residues of the core) were then estimated for all snapshots and averaged over 50 ns for each simulation run at temperatures 310, 400 and 450 K (Table 3). In addition (average) fractional conservation of secondary structural content (with respect to the native crystal structure) was also calculated for each simulation (50 ns).

On an average, about 0.65–0.75 of the native contacts between helices and core, were conserved in the simulation at 310 K. However, helix H2 appeared to be relatively loosely bound to the core compared to H1, with a sharper decline in the conserved fraction upon elevation of temperature to 400 K (H1 – 0.65(0.18); H2 – 0.47(0.29)). At 450 K, 0.48, 0.37 of the native contacts were conserved for H1, H2 respectively. Again, H2 also exhibited an increased tendency to unwind as only 40% of its constituent residues (as found in the crystal structure) retained their helical conformation even by 400 K. In general, the strands constituting the barrel appear to have greater stability (0.81 secondary structural content conserved, Table 3), in contrast to both the helices. Thus, the simulations appeared to indicate the tendency of both helices to gradually unwind and adopt non-native geometries with respect to the core, quite early in the unfolding process.

4. Conclusion

The work reported in this paper characterizes the equilibrium unfolding of cyclophilin mediated by the denaturant GdmCl. Initial evidence for the existence of intermediate states appears with regard to the non-superposition of the unfolding curves monitored by fluorescence and CD. Analysis of the near and far UV CD spectra indicates the pronounced disruption of the native tertiary contacts at 1.2 M GdmCl while conserving significant fraction of the native secondary structural content, characteristic of a molten globule [23–35]. This was further confirmed by hydrophobic dye-binding studies utilizing bis-ANS where a prominent peak was observed at 1.2 M GdmCl. Thermal characterization of the protein incubated with 0.2, 0.6 and 1.0 M of GdmCl by DSC exhibited a gradual shift from a two (involving two T_m's) to a single step transition (with a unique T_m).

Examination of the native crystal structure and data from MD simulations leads us to hypothesize that the possible equilibrium intermediate could involve disjuncture of the helices from their native geometry with respect to the barrel/core during initial stages of the unfolding process.

Authors' contributions

Dipak Dasgupta and Rahul Banerjee conceived the experiments, Sourav Roy performed the experiments. Dhananjay Bhattacharyya gave guidance and Sankar Basu participated in Molecular Dynamics Simulations. Rahul Banerjee, Dipak Dasgupta and Sourav Roy wrote the manuscript.

Conflict of interest

None declared.

Acknowledgements

The authors thank Prof. Krishnananda Chattopadhayay (Indian Institute of Chemical Biology) and Dr. Amrita Dasgupta (National Centre for Biological Sciences) for providing valuable suggestions and comments. This work was supported by the intramural grants from Department of Atomic Energy, Government of India (Projects of SINP:- MSACR [XII-R&D-SIN-5.04-0102] and BARD -[XII-R&D-SIN-5.04-0103]).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijbiomac.2014.05.063.

References

[1] Ashutosh, S. Sundar, N. Goyal, J. Infect. Dis. 56 (2007) 143–153, http://dx.doi.org/10.1099/jmm.0.46841-0.
[2] D. Sereno, P. Holzmuller, J.L. Lemesre, Acta Trop. 74 (2000) 25–31.
[3] J. Bua, L.E. Fichera, A.G. Fuchs, M. Potenza, M. Dubin, R.O. Wenger, G. Morretti, C.M. Scabone, A.M. Ruiz, Parasitology 135 (2008) 217–228.
[4] M.E. Perkins, T.W. Wu, S.M. Le Blancq, Antimicrob. Agents Chemother. 42 (1998) 843–848.
[5] P. Wang, J. Heitman, Genome Biol. 6 (2005) 226.
[6] H. Du, L. Guo, F. Fang, D. Chen, A.A. Sosunov, G.M. McKhann, Y. Yan, C. Wang, H. Zhang, J.D. Molkentin, F.J. Gunn-Moore, J.P. Vonsattel, O. Arancio, J.X. Chen, S.D. Yan, Nat. Med. 14 (2008) 1097–1105.
[7] Z. Keckesova, L.M. Ylinen, G.J. Towers, J. Virol. 80 (2006) 4683–4690.

[8] J. Kallen, V. Mikol, P. Taylor, M.D. Walkinshaw, J. Mol. Biol. 283 (1998) 435–449.
[9] H. Ke, J. Mol. Biol. 228 (1992) 539–550.
[10] V. Venugopal, B. Sen, A.K. Datta, R. Banerjee, Acta Crystallogr. F 63 (2007) 60–64.
[11] D. Mitra, S. Mukherjee, A.K. Das, FEBS Lett. 580 (2006) 6846–6860.
[12] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. De Bolt, D. Ferguson, G. Siebel, P. Kollman, Comput. Phys. Commun. 91 (1995) 1–41.
[13] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, J. Comput. Chem. 26 (16) (2005) 1781–1802.
[14] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, J. Comput. Phys. 23 (1977)327–341.
[15] W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. 14 (1996) 33–38 http://www.ks.uiuc.edu/Research/vmd/
[16] D. Frishman, P. Argos, Proteins: Struct. Funct. Genet. 23 (1995) 566–579.
[17] A.H. Wani, J.B. Udgaonkar, J. Mol. Biol. 387 (2009) 348–362.
[18] A.K. Mandal, S. Samaddar, R. Banerjee, S. Lahiri, A. Bhattacharyya, S. Roy, J. Biol. Chem. 278 (2003) 36077–36084.
[19] R.F. Latypov, H. Cheng, N.A. Roder, J. Zhang, H. Roder, J. Mol. Biol. 357 (2006) 1009–1025.
[20] S. Lindhoud, A.H. Westphal, J.W. Borst, C.R.M. van Mierlo, PLoS ONE 7 (10) (2012), http://dx.doi.org/10.1371/journal.pone.0045746.
[21] L. Zhu, Y.X. Fan, J.M. Zhou, Biochim. Biophys. Acta 1544 (2001) 320–332.
[22] G.V. Semisotnov, N.A. Rodionova, O.I. Razgulyaev, V.N. Uversky, A.F. Gripas, R.I. Gilmanshin, Biopolymers 31 (1991) 119–128.
[23] A.L. Fink, L.J. Calciano, Y. Goto, T. Kurotsu, D.R. Palleros, Biochemistry 33 (1994) 12504–12511.
[24] B.K. Das, T. Bhattacharyya, S. Roy, Biochemistry 34 (1995) 5242–5247.
[25] R.L. Baldwin, C. Frieden, G.D. Rose, Proteins 78 (2010) 2725–2737.
[26] D. Barrick, R.L. Baldwin, Protein Sci. 2 (1993) 869–876.
[27] A. Iram, T. Alam, J.M. Khan, T.A. Khan, R.H. Khan, A. Naeem, PLoS ONE 8 (8) (2013), http://dx.doi.org/10.1371/journal.pone.0072075.
[28] K. Kuwajima, Proteins 6 (1989) 87–103.
[29] K. Kuwajima, FASEB J. 10 (1996) 102–109.
[30] M. Ferrer, G. Barany, C. Woodward, Nat. Struct. Biol. 2 (1995) 211–217.
[31] H. Christensen, R.H. Pain, Eur. Biophys. J. 19 (1991) 221–229.
[32] S.H. Park, J. Biochem. Mol. Biol. 37 (2004) 676–683.
[33] S. Sheshadri, G.M. Lingaraju, R. Vardarajan, Protein Sci. 8 (1999) 1689–1695.
[34] O.B. Ptitsyn, V.N. Uversky, FEBS Lett. 341 (1994) 15–18.
[35] S.K. Jha, J.B. Udgaonkar, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 12289–12294.