Essay describing the thesis work

It has been the central dogma since long that the function of a protein is related to its significant and unique state, which is a well-defined three-dimensional structure called native structure. Though this paradigm about folded structures and functions remains valid, recent structural and genomic data have clearly shown that not all proteins have unique folded structures under normal physiological conditions. The disorder can be local as well as global. A special term 'natively unfolded' was coined to describe the properties of this class of proteins. Recently a Protein Quartet Model has been proposed for generalization of the structure-function paradigm. According to this, a biological function arises as a result of interplay between four specific conformational forms, namely, ordered forms, molten globules, pre-molten globules, and random coils.

Detailed characterization of the unfolded state and consequent identification of the folding initiation sites in a given protein provide valuable insight into its folding mechanism. Well-formed or transient residual structures in the unfolded state can be possible candidates for folding initiation sites. Unfolded or partly unfolded states of globular proteins can be created by use of denaturants, which disrupt the non-covalent interactions and propel them to lose their biological activity. This state is referred to as the 'denatured state'. However, often, denaturation is not accompanied by complete unfolding of the protein, and the denatured state is an ensemble of conformations between native and completely unfolded states. For all these reasons it is necessary to characterize all states of protein at high resolution.

NMR methodology of investigating unfolded or partly folded proteins

In addition to its normal application for determination of 3D structure at atomic resolution of folded proteins in solution, Nuclear Magnetic Resonance (NMR) spectroscopy is unparalleled in its ability to provide detailed structural and dynamical information of unfolded and flexible proteins. The importance of use NMR methodology for study of biological macromolecules in solution has recently emphasized again by awarding 2002 Nobel Prize in chemistry for this technique. The proton chemical shift dispersion and line width give first-hand information on the state of proteins; poor proton chemical shift dispersion and sharp lines are indicative of disorder.

New strategy for resonance assignment in labeled proteins

Sequence specific resonance assignment is the first step in detailed NMR characterization of proteins. In context of fast growing genomic and proteomic data, there is a high demand on the speed of resonance assignment, whereas, the presently available methods based either on NOESY or on some triple resonance experiments are

rather slow. They also have limited success with unfolded proteins because of lack of NOEs, and poor dispersion of amide and carbon chemical shifts. I developed an efficient and high throughput approach for rapid resonance assignment suitable for both folded and unfolded proteins, making use of the triple resonance experiments HNN and HN(C)N developed recently (see fig 1).

It has three underlying principles: First, the experiments exploit the ^{15}N chemical shift dispersions which are generally very good for both folded and unfolded proteins (see fig 2), along two of the three dimensions, second, they directly display sequential amide and ^{15}N correlations along the polypeptide chain, and, third, the sign patterns of the diagonal and the sequential peaks originating from any residue are dependent on the nature of the adjacent residues, especially the glycines and the prolines. These lead to so called 'triplet fixed points' which serve as starting points and/or check points during the course of sequential walks, and explicit side chains assignment becomes less crucial for unambiguous backbone assignment. These features significantly enhance the speed of data analysis, reduce the amount of experimentation required and thus result in a substantially faster and unambiguous assignment. Following the amide and ^{15}N assignments, the other proton and carbon assignments can be obtained in a straightforward manner, from the well-established 3D triple resonance experiments.

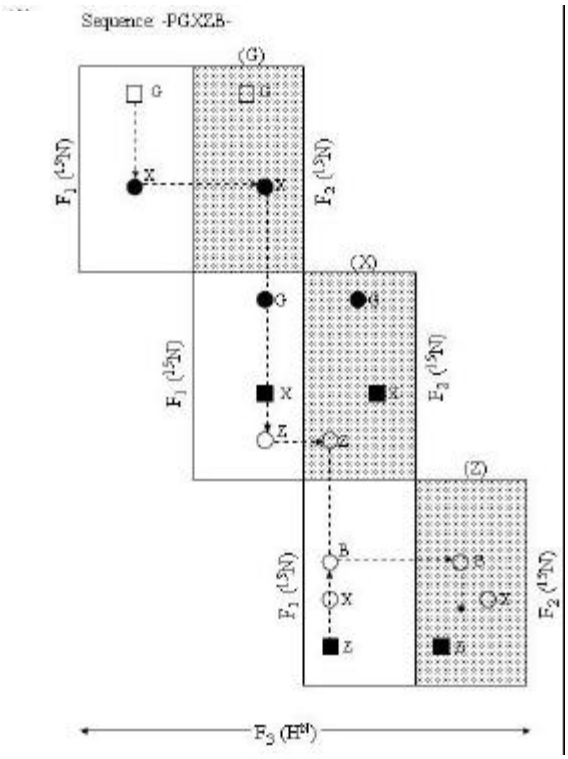

Fig 1: The protocol for sequential walk through the HNN spectrum using an illustrative sequence PGXZB where X, Z, B can be any residue other than glycine and proline. Squares are diagonal peaks and circles are sequential peaks. Filled and open symbols are positive and negative peaks respectively.

I have successfully applied the new approach to different proteins in the MW range 10-29 kDa both in the folded and in the unfolded forms - the two ends of the folding funnel. It was also used to get assignment of flexible regions in the 160 kDa molten globule-like aggregate of Barstar at low pH, which is the first aggregate for which resonance assignments have been obtained. Table below lists the proteins for which complete resonance assignments was obtained using this approach.

Proteins	Size (kDa)
HIV-1 protease-acetyl pepstatin complex	22
Unfolded HIV-1 protease (in 6M Gdn, 8M urea)	22
Barstar aggregate	160
Barstar unfolded (in 8M urea)	10
Folded FK506 binding protein (FKBP)	12
Folded SUMO	11
Unfolded SUMO	11
SUMO-GED complex	26
Unfolded TFR-HIV-1 pro-protease dimer	29
Unfolded TFR-HIV-1 pro-protease monomer	18

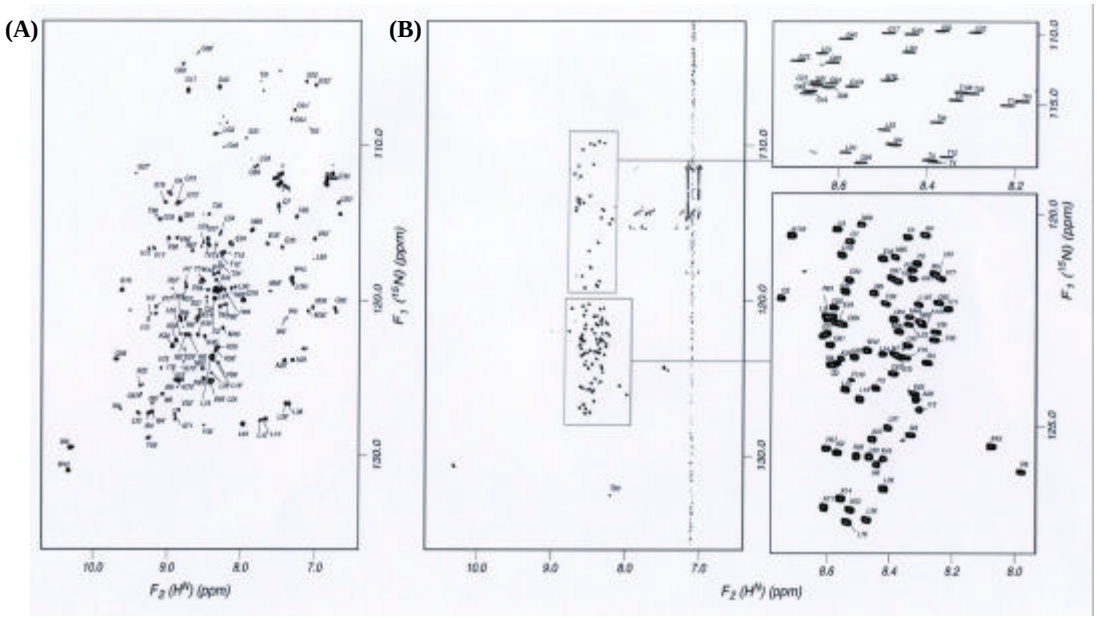

Fig 2*:* HSQC spectra with H^N and ^{15}N assignments for folded **(A)** and unfolded **(B)** HIV-1 protease tethered dimer. The resonance assignments have been obtained using the approach described above.

A methodology for measurement of $^1J_{C^\alpha N}$ *coupling*

A novel approach developed by me which uses the ratio of the intensities of the sequential and diagonal peaks in the F_2-F_3 planes of the HN(C)N spectrum for measurement of one bond N-C^α *J*-couplings in medium size labeled proteins. The reliability of the approach has been demonstrated using a short peptide and then its application to proteins has been demonstrated using doubly labeled FK506 binding protein (FKBP, molecular mass ~ 12 kDa). This quantitative application extends the significance of the HN(C)N experiment, in the context of structural genomics, since the single experiment, not only provides a great enhancement in the speed

of resonance assignment, but also quantitative structural information in the form of dihedral restraints for precise determination of the structures of the proteins .

Characterization of unfolded state of HIV-1 protease and folding Studies

Folding studies on proteases by the conventional hydrogen exchange experiments are severely hampered, due to interference from the autolytic reaction in the interpretation of the exchange data. In this background, NMR identification of the hierarchy of early conformational transitions (folding propensities) in HIV-1 protease was done, by systematic monitoring of the changes in the state of the protein as it is subjected to different degrees of denaturation by guanidine hydrochloride. Secondary chemical shifts, H^N-H^α coupling constants, ^1H – ^{15}N NOEs and ^{15}N transverse relaxation parameters have been used to report on the residual structural propensities, motional restrictions, conformational transitions etc, and the data suggest that even under the strongest denaturing conditions (6M guanidine) hydrophobic clusters as well as different native and non-native secondary structural elements are transiently formed. These constitute the folding nuclei, which include residues spanning the active site, the hinge region and the dimerization domain. Interestingly, the proline residues influence the structural propensities, and the small amino acids, gly and ala enhance the flexibility of the protein. On reducing the denaturing conditions partially folded forms appear. The residues showing high folding propensities are contiguous along the sequence at many locations or are in close proximity on the native protein structure suggesting a certain degree of local cooperativity in the conformational transitions. The dimerization domain, the flaps and their hinges seem to exhibit the highest folding propensities. The data suggest that even the early folding events may involve many states near the surface of the folding funnel. The results obtained are summarized in the figure 3.

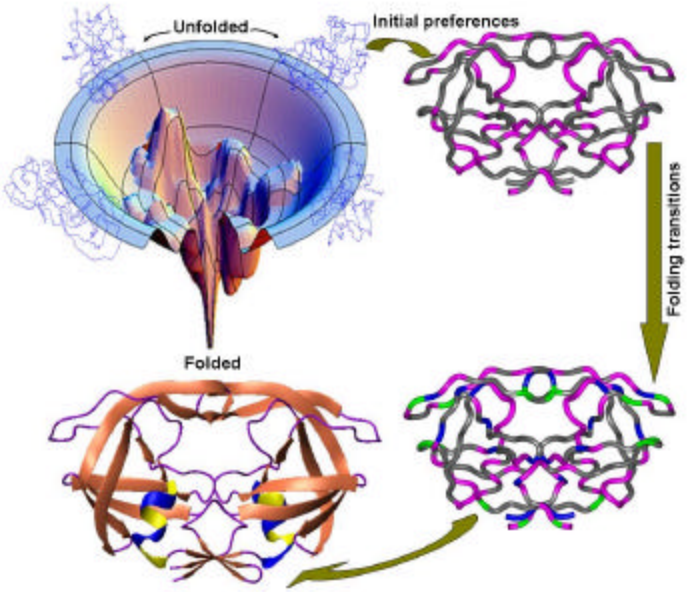

Fig 3: Initial folding hierarchy in HIV-1 protease.

Structure and dynamics of DNA Quadruplexes containing C-tetrad

In addition to proteins, structural polymorphism; folding and dynamism in DNA are of paramount importance for understanding of structure function relation in biology. While DNA has been known to form folded and multistranded structures, G-quadruplexes being the most prominent ones, very little is known about the dynamic characteristics of these structures. In this background, I have also investigated the structure and dynamics of DNA quadruplexes formed by a SV40 repeat sequence GGCGGG in aqueous solution containing Na^+ ions at neutral pH. This has led to observation of a novel C-tetrad in the middle in the structure of the sequences d-TGGCGGGT and d-TGGGCGGT. The C-tetrad is formed by symmetrical pairing of four Cs in a plane via NH_2-O2 H-bonds (see fig 4). ^{13}C relaxation measurements at natural abundance for C1' sugar carbons provided valuable insight into the dynamism of G and C-tetrads in the quadruplex.

In another sequence d-TGGCGGC containing two GGC triplet repeats related to the disease Fragile X-Syndrome an interesting conformational switch was observed. At neutral pH, the solution structure is a parallel-stranded quadruplex in presence of K^+ ions. In contrast, the sequence forms an antiparallel duplex in Na^+ containing solutions. As the pH of the Na^+ sample is lowered, an equilibrium mixture of a duplex and a quadruplex appears, and at pH 2.2, the molecule exists entirely as a quadruplex. These results would be of significance from the point of view of recognition and regulation by different helicase enzymes, which have been found to discriminate between different types of quadruplex structures.

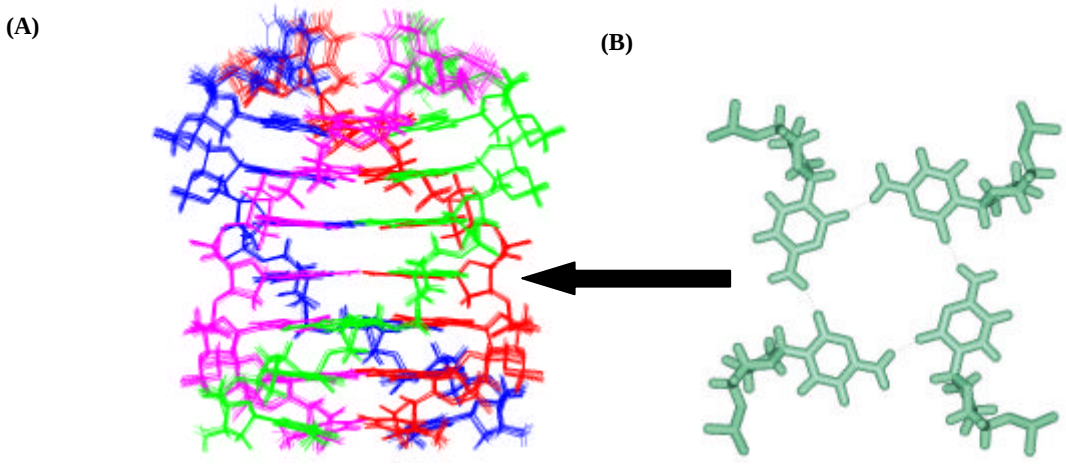

Fig 4:**(A)** Superposition of ten relaxation matrix refined quadruplex structures of d-TGGGCGGT. **(B)** The configuration of C-tetrad in the average structure. (**PDB id 1EVO**)

Publications

A. (In refereed journals)

1. **Bhavesh NS,** Sinha R, KrishnaMohan PM and Hosur RV 2003 NMR elucidation of early folding hierarchy in HIV-1 protease *J. Biol. Chem.* **278**, 19980-19985

2. **Bhavesh NS**, Chatterjee A, Panchal SC and Hosur RV 2003 Application of HN(C)N to rapid estimation of $^{1}J(N\text{-}C^{\alpha})$ coupling constants correlated to y?torsion angles in proteins: Implication to structural genomics. *Biochem. Biophy. Res. Comm***. 311**, 678-684.

3. **Bhavesh NS** and Hosur RV 2003 Exploring 'unstructured' proteins. *Proc. Ind. Natl. Sci. Acad.* (*In Press*).

4. **Bhavesh NS** Patel PK and Hosur RV 2003 Novel structural motifs in telomere DNA. *Proc. Ind. Natl. Sci. Acad.* (*In Press*).

5. Juneja J, **Bhavesh NS**, Udgaonkar JB and Hosur RV 2002 NMR identification and characterization of the flexible regions in the 160 kDa molten globule-like aggregate of Barstar at low pH. *Biochemistry* **41**, 9885-9899.

6. Kumar M, Kannan KK, Hosur MV, **Bhavesh NS**, Chatterjee A, Mittal R.and Hosur RV 2002 Effects of remote mutation on the autolysis of HIV-1 Protease: X-ray and NMR investigations. *Biochem, Biophy. Res. Comm***. 294**, 395-401.

7. Chatterjee A, **Bhavesh NS**, Panchal SC and Hosur RV 2002 A novel protocol based on HN(C)N for rapid resonance assignment in (^{15}N, ^{13}C) labeled proteins: Implications to structural genomics. *Biochem. Biophy. Res. Comm***. 293**, 427-432.

8. **Bhavesh NS**, Panchal SC and Hosur RV 2001 An efficient high-throughput resonance assignment procedure for structural genomics and protein folding research by NMR. *Biochemistry* **40**, 14727-14735. (**Accelerated publication**)

9. **Bhavesh NS**, Panchal SC, Mittal R and Hosur RV 2001 NMR identification of local structural preferences in HIV-1 protease tethered heterodimer in 6 M guanidine hydrochloride. *FEBS Lett.* **509**, 218-24.

10. Panchal SC, **Bhavesh NS** and Hosur RV 2001 Real time NMR monitoring of local unfolding of HIV-1 protease tethered dimer driven by autolysis. *FEBS Lett*. **497**, 59-64.

11. Panchal SC, **Bhavesh NS** and Hosur RV 2001 Improved 3D triple resonance experiments, HNN and HN(C)N, for H^{N} and ^{15}N sequential correlations in (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. *J. Biomol. NMR* **20**, 135-147.

12. Patel PK, **Bhavesh NS** and Hosur RV 2000 Cation Dependent Conformational Switches in d-TGGCGGC Containing Two Triple Repeats of Fragile X syndrome: NMR Observations. *Biochem. Biophy. Res. Comm***. 278,** 833-838.

13. Patel PK, **Bhavesh NS** and Hosur RV. 2000 NMR Observation of a Novel C-Tetrad in the Structure of the SV40 Repeat Sequence GGGCGG**.** *Biochem. Biophy. Res. Comm***. 270**, 967-971.

B. (In conference proceedings (till Dec 2003))

- Hosur RV, Panchal SC, **Bhavesh NS**, Chatterjee A, Kumar A, Mishra RK and Mittal R 2003 High-throughput procedures for structural genomic research by NMR *Recent Trends in Biophysical Research* (Ed M. Maiti, G. Suresh Kumar and S.Das) 54-66

- **Bhavesh NS**, Sinha R, KrisnaMohan PM and Hosur RV 2003 NMR elucidation of hierarchy in the folding of HIV-1 protease. At National Magnetic Resonance Symposium at IISc. Bangalore **Oral presentation**

- **Bhavesh NS,** Juneja J, Udgaonkar JB and Hosur RV 2003 NMR assignment and identification of local structural preferences in Barstar in 8 M urea. At National Magnetic Resonance Symposium at IISc. Bangalore. **Poster presentation**

- Chatterjee A, **Bhavesh NS**, Panchal SC and Hosur RV 2003 New Strategies for Resonance Assignments and Measurement of $^{1}J(C^{\alpha}\text{-}N)$ coupling constants in Labeled Proteins: Application of 3D HN(C)N and its Extension to Four Dimensions. At National Magnetic Resonance Symposium at IISc. Bangalore. **Poster presentation**

- **Bhavesh NS,** Chatterjee A, Panchal SC and R.V. Hosur 2002 Rapid resonance assignment and measurement of $^{1}J(\text{N-C}^{a})$ coupling constants in labeled proteins from a single experiment. At 2nd International conference on structural biology and functional genomics. National University of Singapore, Singapore. **Poster presentation**

- **Bhavesh NS,** Sinha R, KrisnaMohan PM and Hosur RV 2002 Exploring structure in unstructured proteins. At National Magnetic Resonance Symposium at SGPGI. Lucknow. **Oral presentation**

- Chatterjee A, **Bhavesh NS**, Panchal SC and Hosur RV 2002 An algorithm for automated resonance assignment of labeled proteins by HNN and HN(C)N spectra. At National Magnetic Resonance Symposium at SGPGI. Lucknow. **Poster presentation**

- **Bhavesh NS,** Panchal SC, Mittal R and Hosur RV 2001 NMR assignment and identification of local structural preferences in HIV-1 protease tethered heterodimer in 6Mguanidine hydrochloride. At the Abdus Salam International Centre for Theoretical Physics (ICTP) Trieste, Italy, Oct 1–12, 2001. **Poster and oral presentation**

- Panchal SC, **Bhavesh NS** and Hosur RV 2001 New 3D triple resonance experiments, HN(CA)N and HN(COCA)N, for H^{N} and ^{15}N sequential correlations in labeled proteins. At National Magnetic Resonance Symposium at IIT, Chennai. **Oral presentation**

- **Bhavesh NS**, Panchal SC and Hosur RV 2001 H^{N}, ^{13}C and ^{15}N assignments in unfolded HIV-I protease tethered dimer. At National Magnetic Resonance Symposium at IIT, Chennai. **Oral presentation**

- Panchal SC, **Bhavesh NS**, Pillai B, Hosur MV and Hosur RV 2001 Studies on HIV-1 protease tethered heterodimer: Structure, dynamics and inhibitor interaction. International Symposium on Structure, Dynamics and Folding of Proteins as Viewed from the Pressure Axis. At Otsu, Japan.

- Patel PK, **Bhavesh NS** and Hosur RV 2000 Structure and Dynamics of C-tetrad in DNA Quadruplexes. At XIXth International Conference on Magnetic Resonance in Biological System (ICMRBS) at Florence, Italy. **Poster presenation**

- Patel PK, **Bhavesh NS** and Hosur RV 2000 NMR Observation of a Novel C-Tetrad in DNA quadruplex. At National Magnetic Resonance Symposium at TIFR Mumbai **Oral presentation**

- Patel PK, **Bhavesh NS** and Hosur RV 1999 NMR Observation of a Novel C-Tetrad in DNA quadruplex. *J. of Biosciences*, **24**, 92. At XIIIth IUPAB conference at New Delhi. **Poster presentation**