SANDEEP KUMAR

Laboratory of Experimental and Computational Biology,

NCI - Frederick, Building 469, Room 147, Frederick, MD-21702, USA.

Phone: (301)-846-6542, Fax: (301)-846-5598, e-mail: kumarsan@ncifcrf.gov

URL: www-lecb.ncifcrf.gov/~kumarsan

Permanent Address: 995/21C, Faridabad, Haryana – 121001, India. Phone: +91-129-5423331

Date of Birth: 13th November, 1968

Nationality & Citizenship: Indian

Employment

1999-Present: Postdoctoral Visiting Fellow, Laboratory of Experimental and Computational Biology, Center for Cancer Research, Division of Basic Science, National Cancer Institute – Frederick, National Institutes of Health, Frederick, Maryland, USA.

1998-1999: Scientist, SAIC-Frederick, NCI-Frederick, Maryland, USA.

Education

Ph.D. Molecular Biophysics from Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India, in 1998.

Ph.D. Thesis title: Geometry and sequence correlation studies on α-helices in globular proteins.

M.Sc. Molecular Biology and Biotechnology with overall grade point average of 4.545 on 5.0 scale (83.175% marks) from G. B. Pant University of Agriculture & Technology, Pant Nagar, India, in 1992.

M. Sc. Thesis title: Computerized mathematical approach for kinetics and production of Glutamine synthetase by *B. brevis* BbG1.

B.Sc. (Honors) Physics with 74.77% marks from University of Delhi, Delhi, India, in 1989.

Honors, awards and Fellowships

1. Listed in 57th edition of Marquis Who's Who in America in November 2002.
2. Postdoctoral Visiting Fellowship from National Institutes of Health, USA.
3. Senior Research Fellowship (SRF) from Council of Scientific and Industrial Research (CSIR), India during 1994-97.
4. International travel award from CSIR, India, in 1996 for participation in an international conference in 1996.

5. Junior Research Fellowship (JRF) from CSIR, India during 1992-94.

6. Department of Biotechnology fellowship during 1990 - 92 for pursuing M. Sc. (Molecular biology and Biotechnology).

Research Interests

Computational molecular biophysics, sequence-structure-function relationship in biological macromolecules, protein folding, binding and thermodynamic stability, electrostatic interactions in proteins, *de novo* protein design, bioinformatics, computational biology, structural genomics, protein—protein interactions, analysis of gene expression data and cellular pathways.

Professional activities

Experience with writing grant applications. Formal and informal reviews of several manuscripts submitted to international journals in the field and of grant applications.

Research Experience

Postdoctoral research (1998 - Present): My research focuses on issues in protein folding, binding, flexibility and thermodynamic stability. This research involves mixing of concepts from computational chemistry, molecular modeling, protein structure and conformation, molecular biophysics and biochemistry with those from statistics, mathematical modeling and classical electrostatics. An advanced level of understanding of issues in protein folding, structure and function as well as strong computer programming skills are essential for this work. Please see appendix A for a brief description. Further information is available at URL: http://www.lecb.ncifcrf.gov/~kumarsan/.

M. Sc. & Ph.D. research: My Ph.D. research (1992-1998) was based on statistical analysis of sequence and structural relationship in α-helices. I had developed an algorithm called HELANAL to characterize α-helix geometries in proteins. This work also involved an interdisciplinary approach with concepts derived from several fields such as statistics, molecular biophysics, protein conformation, chemistry, molecular biology and computer programming. My M. Sc. Research (1990-1992) focused on bacterial sporulation. This was mostly experimental research work that involved microbiological and biochemical techniques. Please see appendix B for a description of my past research experience.

Future Research plans

Please see appendix C.

Teaching experience

Please see appendix D for my present teaching experience and future plans

Publication list

1. **Kumar, S.** & Nussinov, R. 2002. Relationship between ion pair geometries and electrostatic strengths in proteins. ***Biophys. J.***, 83, 3, 1595 - 1612. *Research paper*.

2. **Kumar, S.**, Tsai, C. J. & Nussinov, R. 2002. Maximal stabilities of reversible two state proteins. ***Biochemistry***, 41, 17, 5359 - 5374. *Research paper*.

3. **Kumar, S.**, Barzilai, A., Haspel, N., Sham, Y. Y., Tsai, C. J., Wolfson, H. J. & Nussinov, R. 2002. Critical building blocks in proteins: A common theme for folding and function. ***Recent research developments in protein folding, stability & design.*** M. M. Gromiha & S. Selvaraj (Editors). Research signpost, Trivandrum, India, pp 207-217. *Book article*.

4. **Kumar, S.** & Nussinov, R. 2002. Close-range electrostatic interactions in proteins. ***ChemBiochem***, 3, 7, 604 – 617. *Review*.

5. Sinha, N., **Kumar, S.** & Nussinov, R. 2001. Inter-domain interactions in hinge-bending transitions. ***Structure***, 9, 1165-1181. *Research paper*.

6. **Kumar, S.**, Tsai, C. J. & Nussinov, R. 2001. Thermodynamic differences among homologous thermophilic and mesophilic proteins. ***Biochemistry***, 40, 47, 14152-14165. *Research paper*.

7. Tsai, C. J., Ma, B., **Kumar, S.**, Wolfson, H. J. & Nussinov, R. 2001. Protein folding: Binding of conformationally fluctuating building blocks *via* population selection. ***Crit. Rev. Biochem. Mol. Biol.***, 36, 399-433. *Review*.

8. **Kumar, S.** & Nussinov, R. 2001. How do thermophilic proteins deal with heat? ***Cell. Mol. Life Sci.,*** 58, 9, 1216-1233. *Review*.

9. Tsai, C. J., Ma, B., Sham, Y. Y., **Kumar, S.** & Nussinov, R. 2001. Structured disorder and conformational selection. ***Proteins***, 44, 4, 418-427. *Review*.

10. **Kumar, S.,** Sham, Y. Y., Tsai, C. J. & Nussinov, R. 2001. Protein folding and function: The N-terminal fragment in adenylate kinase. ***Biophys. J.***, 80, 5, 2439-2454. *Research paper*.

11. **Kumar, S.** & Nussinov, R. 2001. Fluctuations in ion pairs and their stabilities in proteins. ***Proteins***, 43, 4, 433-454. *Research paper*.

12. **Kumar, S.**, Wolfson, H. J. & Nussinov, R. 2001. Protein flexibility and electrostatic interactions. ***IBM J. Res. Dev.***, 45, 3 / 4, 499-512. *Invited research paper*.

13. Ma, B., **Kumar, S.**, Tsai, C. J., Wolfson, H. J., Sinha, N., & Nussinov, R. 2001. Protein-Ligand interactions: Induced fit. In Robertson, S. (Ed.): ***Encycl. Life Sci.***, London, Macmillan Reference Limited, Nature publishing group / www.els.net. *Book article*.

14. Tsai, C. J., Ma, B., Yuk Sham, **Kumar, S.**, Wolfson, H. J. & Nussinov, R. 2001. A hierarchical, building blocks based computational method for protein structure prediction. ***IBM J. Res. Dev.***, 45, 3 / 4, 513-523. *Invited research paper*.

15. **Kumar, S.** & Nussinov, R. 2000. Fluctuations between stabilizing and destabilizing electrostatic contributions of ion pairs in conformers of the c-Myc-Max leucine zipper. ***Proteins***, 41, 4, 485-497. *Research paper*.

16. Bansal, M., **Kumar, S.** & Velavan, R. 2000. HELANAL - A program to characterize helix geometry in proteins**. *J. Biomol. Struct. Dyn.***, 17, 5, 811-819. *Research paper*.

17. **Kumar, S.**, Ma, B., Tsai, C. J., Sinha, N. & Nussinov, R. 2000. Folding and binding cascades: Dynamic landscapes and Population shifts. ***Protein Sci.***, 9, 1, 10-19. *Review*.

18. **Kumar, S.**, Tsai, C. J. & Nussinov, R. 2000. Factors enhancing protein thermostability. ***Protein Eng.***, 13, 3, 179-191. *Research paper.*

19. **Kumar, S.**, Tsai, C. J., Ma, B. & Nussinov, R. 2000. Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. ***Proteins***, 38, 4, 368-383. *Research paper*.

20. Ma, B., **Kumar, S.**, Tsai, C. J., Zeng Jian Hu & Nussinov, R. 2000. Transition state ensemble in enzyme catalysis: possibility, reality or necessity? ***J. Theoret. Biol.***, 203, 4, 383-397. *Review*.

21. **Kumar, S.**, Tsai, C. J., Ma, B. & Nussinov, R. 2000. Contribution of salt bridges toward protein thermostability. In Sarma, R. H. & Sarma, M. H. (Eds.): ***J. Biomol. Struct. Dyn.***, Conversation 11, Issue 1, Proceedings of the 11th conversation, University of Albany, SUNY, Adenine Press, pp 79-85. *Conference proceedings*.

22. **Kumar, S.** & Nussinov, R. 1999. Salt bridge stability in monomeric proteins. ***J. Mol. Biol.***, 293, 5, 1241-1255. *Research paper.*

23. Ma, B., **Kumar, S.**, Tsai, C. J. & Nussinov, R. 1999. Folding funnels and binding mechanisms. ***Protein Eng.***, 12, 9, 713-720. *Review*.

24. **Kumar, S.**, Ma, B., Tsai, C. J., Wolfson, H. J. & Nussinov, R. 1999. Folding funnels and conformational transitions via hinge bending motions. ***Cell Biochem. Biophys.***, 31, 2, 141-164. *Review.*

25. Tsai, C. J., **Kumar, S.**, Ma, B. & Nussinov, R. 1999. Folding funnels, binding funnels and protein function. ***Protein Sci.***, 8, 6, 1179-1188. *Review*.

26. **Kumar, S.** & Bansal, M. 1998. Geometrical and sequence characteristics of α-helices in globular proteins. ***Biophys. J.***, 75, 4, 1935-1944. *Research paper*.

27. **Kumar, S.** & Bansal, M. 1998. Dissecting α-helices: Position specific sequence determinants of α-helices in globular proteins. ***Proteins***, 31, 4, 460-476. *Research paper*.

28. **Kumar, S.** & Bansal, M. 1996. Structural and sequence characteristics of long α-helices in globular proteins. ***Biophys. J.***, 71, 3, 1574-1586. *Research paper*.

Manuscripts in preparation or communicated

1. **Kumar, S.** and Nussinov, R. (2002). The role of electrostatics in heat and cold adaptation by citrate

synthase. Submitted. *Research paper.*

2. **Kumar, S.,** Tsai, C. J. and Nussinov, R. (2002). Temperature range of thermodynamic stability for the native state of reversible two-state proteins. Submitted. *Research paper*.

Selected Conference/Symposia presentations

(Total Number of conference presentations is 14)

1. Role of electrostatic interactions in protein thermostability. **Kumar, S.** & Nussinov, R. The protein society: 15th symposium, 2001, Philadelphia, PA, USA. Abstract published in ***Protein Sci.***, 10, Suppl. 2, 413.

2. How do thermophilic proteins deal with heat? **Kumar, S.** & Nussinov, R. Gordon research conference on proteins. June 24 – 29, 2001. Holderness School, Plymouth, New Hampshire, USA.

3. Most salt bridges are stabilizing towards proteins. **Kumar, S.** & Nussinov, R. The protein society : 14th symposium,2000, San Diego, CA, USA. Abstract published in ***Protein Sci.***, 9, Suppl. 1, 241-S.

4. Salt bridge stability in monomeric proteins. **Kumar, S.** & Nussinov, R. 44th annual meeting of biophysical society, 2000, New Orleans, Louisiana, USA. Abstract published in ***Biophys. J.***, 78, 1, part 2, 2505-Pos.

5. Why salt bridges in thermophiles? **Kumar, S.**, Ma, B., Tsai, C. J. & Nussinov, R. The protein society: 13th symposium,1999, Boston, MA, USA. Abstract published in ***Protein Sci.***, 8, Suppl. 1, 211-S.

6. Why salt bridges in thermophilic proteins? **Kumar, S.**, Ma, B., Tsai, C. J. & Nussinov, R. 43rd annual meeting of biophysical society, 1999, Baltimore, MD, USA. Abstract published in ***Biophys. J.***, 76, 1, part 2, Su-Pos451.

7. Position dependent sequence characteristics of α-helices in globular proteins. **Kumar, S.** & Bansal, M.. 24th Aharon Katzir-Katchalsky conference (Bioinformatics - Structure), 1996, Jerusalem, Israel. Abstract published in ***Fold. & Des.*** I:A95.

8. Spectrofluorometric comparison of different forms of Glutamine synthetase from *B. brevis*. G. K. Garg, R. P. S. Kushwaha, Suja Abraham & **Kumar, S.**. Molecular mechanisms of enzyme action, XVI IUBMB, 1994, Indian Institute of Science, Bangalore, India.

Referee Names and Contact Information

Prof. Ruth Nussinov, Laboratory of Experimental and Computational Biology, Building 469, Room 151, NCI-Frederick, Frederick, MD –21702, USA. Email: ruthn@ncifcrf.gov, phone: +1-301-846-5579, fax: +1-301-846-5598.

Prof. Jacob V. Maizel, Jr., Laboratory of Experimental and Computational Biology, Building 469, Room 151, NCI-Frederick, Frederick, MD –21702, USA. Email: jmaizel@ncifcrf.gov, phone: +1-301-846-5532, fax: +1-301-846-5598.

Dr. Nagarajan Pattabiraman, Room W224A, Department of Oncology, Lombardi Cancer Center, 3900

Reservoir Road, Georgetown University, Washington DC, USA. Email: np47@georgetown.edu, phone: +1-202-687-5946.

Dr. Thomas D. Schneider, Laboratory of Experimental and Computational Biology, Building 469, Room 151, NCI-Frederick, Frederick, MD –21702, USA. Email: toms@ncifcrf.gov, phone: +1-301-846-5581, fax: +1-301-846-5598.

Appendix A: Present Research Activity

I joined the laboratory of experimental and computational biology in NCI-Frederick as a postdoctoral researcher in January 1998. Since then, I am focusing on protein stability, nature of electrostatic interactions such as salt bridges in proteins, protein flexibility and coupling between protein folding andW function.

Protein stability and electrostatic interactions

Stability of a protein depends on its sequence, structural and thermodynamic properties as well as its environmental conditions *e.g.* solvent, pH, temperature, salt concentration and presence/absence of denaturants, ligands, cofactors or metal ions. For a protein or protein domain, which folds in a reversible two-state (N↔U) manner, the variation in its thermodynamic stability with respect to temperature can be described by Gibbs-Helmholtz equation:

$$\Delta G(T) = \Delta H_m (1-T/T_m) - \Delta C_p[(T_m - T) + T\ln (T/T_m)]$$

Where, $\Delta G(T)$ is Gibbs free energy change at temperature, T, between native (N) and unfolded (U) state of the protein. T_m is the melting temperature of the protein. ΔH_m and ΔC_p are the enthalpy change at melting temperature and heat capacity change between native (N) and unfolded (U) states respectively. A plot of this equation yields protein stability curve. Comparison of protein stability curves among homologous thermophilic and mesophilic proteins can facilitate analysis of relationship between macroscopic thermodynamic properties and microscopic sequence / structural differences in these proteins. Recently, I have compared protein stability curves among five families of thermophilic and mesophilic proteins. Among the several possible mechanisms, the thermophilic proteins in the database achieve greater thermal stability by broadening and up-shifting their stability curves (Figure 1). This result shows a direct correlation between maximal protein stability with melting temperature. These observations predict that thermophilic proteins contain additional specific interactions, such as electrostatic interactions, as compared to their mesophilic homologues. An independent analysis of thermodynamic parameters for 31 non-homologous reversible two state folding proteins from thermophilic, mesophilic and psychrophilic organisms has supported this conclusion. This analysis has also revealed that most of the reversible two state proteins with significantly large hydrophobic

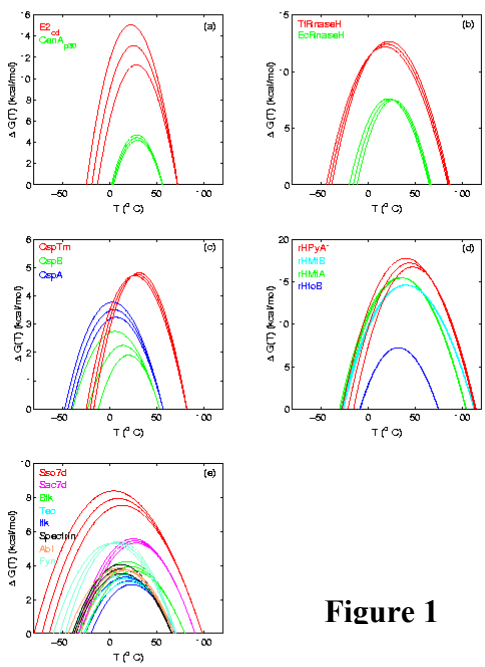

Figure 1

cores are maximally stable around the room temperature, irrespective of their amino acid sequence, structural fold and living temperatures of the source organisms.

The studies on temperature dependent variations in protein stability have been greatly facilitated by discovery of the thermophilic and psychrophilic organisms. Presently, complete genome sequences are available for several thermophilic organisms. While crystal structures for several thermophilic proteins are available in the protein data bank (PDB), those for psychrophilic proteins are relatively fewer. I have recently compiled a database of 18 non-redundant families of proteins from thermophiles and mesophiles and compared all sequence and structural factors among the homologues. The thermophilic and mesophilic proteins have similar hydrophobicities, compactness, oligomeric states, polar and non-polar contribution to surface areas, main-chain and side-chain hydrogen bonds. Amino acid insertions / deletions and proline substitutions do not show consistent trends. However, the thermophilic proteins consistently showed increased and better optimized electrostatic interactions. Recent analyses of complete genomes of the thermophilic organisms also support this observation. In particular, increase in the number of salt bridges and their networks may explain greater thermal stability of glutamate dehydrogenase from hyperthermophile *Pyrococcus furiosus*. Continuum electrostatics based calculations indicate that the salt bridges in the thermophilic glutamate dehydrogenase are highly stabilizing. In contrast, the salt bridges in the mesophilic glutamate dehydrogenase from *Clostridium symbiosum* are only marginally stabilizing. The presence of a larger number of salt bridges and their networks renders an electrostatically more favorable protein environment in the thermophilic glutamate dehydrogenase.

At the same time, whether electrostatic interactions stabilize or destabilize proteins is a controversial issue. Recent literature indicates that the electrostatic free energy contributions by close range electrostatic interactions (*e.g.* salt bridges) towards protein stability vary from being highly stabilizing to insignificant to even destabilizing. The salt bridges can become destabilizing due to their inability to overcome energetically unfavorable desolvation of the charged residues in a folded protein. I have performed continuum electrostatics calculations on a dataset of 222 non-equivalent salt bridges derived from 36 non-homologous high-resolution monomeric protein crystal structures. Most of the salt bridges are stabilizing, regardless of whether they are buried or exposed, isolated or networked, hydrogen bonded or non-hydrogen bonded. Geometrical orientation of the side chain charged groups in the salt bridging residues with respect to each other is crucial in determining the electrostatic strength of the salt bridges. Polarizabilities of the protein molecules also affect the electrostatic strengths of the salt bridges.

Protein flexibility and electrostatic interactions

Based on lattice modeling studies and statistical mechanical concepts, a "new view" of protein folding has been proposed. It describes protein folding as a multiple pathway process with different protein molecules

gliding down their energy landscapes along different routes to reach the native conformation. With other members in the group, I have been exploring the implications of the "new view" for folding and binding mechanisms of the real proteins. We have suggested that energy landscapes of the proteins are dynamic, changing with the changes in the protein's environment. Redistribution of conformer populations and conformer selection is the fundamental mechanism for protein function. These concepts are also intimately related with protein flexibility. There are two kinds of protein flexibilities, namely, segmental and systemic flexibility. Segmental protein flexibility refers to slow time-scale motion (*e.g.* hinge-bending) in the protein molecule in response to function. Systemic flexibility refers to fast time-scale motion in protein backbone and side chain atoms. Unlike segmental flexibility, the motion due to systemic flexibility is distributed throughout the molecule. The electrostatic interactions are affected by the protein flexibilities and *vice versa*. An analysis of protein structures shows that formation of salt bridges is avoided across fragments, domains and subunits which move with respect to one another due to segmental flexibility. Electrostatic stabilities of ion pairs show considerable variations in the NMR conformer ensembles of proteins due to systemic flexibility. These studies have also provided further insights into the relationship between geometry and electrostatic strengths of close range electrostatic interactions.

Another aspect of protein flexibility relates to the molecular adaptation of proteins to cold environments. Enzymes from cold loving (psychrophilic) micro-organisms have high catalytic activities at low temperatures. The greater catalytic efficiency in these enzymes is though to be due to greater flexibility of these proteins. Recently, I have carried out extensive sequence and structural comparison of citrate synthase from hyperthermophile *Pyrococcus furiosus*, mesophile chicken and psychrophile Antarctic bacterium DS2-3R. The hyperthermophilic and psychrophilic citrate synthases show greater similarities to each other than to the mesophilic citrate synthase. Both hyperthermophilic and psychrophilic citrate synthases show increased number of charged residues and electrostatic interactions. Continuum electrostatic studies of the two enzymes indicate the role for electrostatics both in heat and cold adaptation by citrate synthase. Electrostatics stabilize the hyperthemophilic citrate synthase, particularly in active site and dimmer interface regions. In psychrophilic citrate synthase, the role of electrostatics is to ensure proper protein hydration at low temperatures and to impart flexibility, particularly, in active site regions *via* highly destabilizing charged residues (Our unpublished work).

Protein folding and function

Protein folding can be viewed as hierarchical arrangement of protein building blocks. A building block is a contiguous segment of protein sequence (~ 30 residues) that may (or may not) maintain its native protein conformation in solution by itself. One (or more) building blocks in the proteins may be critical for acquisition of their native protein structure. I have developed a computational method to detect critical building blocks in

protein structures. Application of this method to adenylate kinase shows an N-terminal building block to be the putative critical building block. Molecular dynamics simulations show that the rest of the adenylate kinase forms a compact non-native conformation in the absence of this critical building block. This building block is also important to protein function and contains the "P-loop" which is conserved in all nucleotide monophosphate kinases. These observations indicate a possible coupling between protein folding and function, especially for those proteins which show complex non-sequential folding pattern. Using a non-redundant database of sequence and structure wise non-homologous protein chains in the protein data bank (PDB), I have developed a library that contains predicted critical building block in proteins. Presently, I am co-advising a graduate student to study this coupling in hen egg white lysozyme and lipase.

My postdoctoral research at NCI-Frederick has been quite productive. So far, it has contributed towards 24 publications (please see publication list). Additionally, one research papers is communicated and more papers are being prepared. This work has also been presented at several conferences/symposia and is very well cited. Further description of my current research work along with the reprints/preprints of the published articles is available at my web site (http://www.lecb.ncifcrf.gov/~kumarsan).

Appendix B: Past Research Experience

My research interests lie in the area of molecular biophysics. My past research experience consists of two parts. First part is the research work carried out during **1990-1992** in partial fulfillment of the requirements for M. Sc. degree in Molecular biology and Biotechnology at G. B. Pant University of Agriculture and Technology. The research work focused on sporulation in *Bacillus brevis* BbG1. I had characterized the growth kinetics of *Bacillus brevis* BbG1 in nutrient and minimal media by mathematical modeling. Glutamine synthetase is considered a biochemical signal for onset of the sporulation process in response to nutrient depletion. Using the fermenter, I had isolated glutamine synthetase at different growth stages of *Bacillus brevis* BbG1 nutrient and minimal media, in presence and absence of inhibitors, Mg^{2+} and alanine. The enzyme was purified using affinity chromatography and its fluorescence spectra were recorded using a fluorescent substrate analog ethenoADP as the extrinsic probe. This work involved mostly microbiological, biochemical and biophysical techniques. This work was useful in probing the conformational changes of Glutamine synthetase at different stages of bacterial growth and in sporulation. My M. Sc. thesis, entitled "Computerized Mathematical approach for kinetics and production of Glutamine synthetase by *B. brevis* BbG1" submitted to G. B. Pant University of Agriculture and Technology, describes this work in details. This work was performed at the departments of Biochemistry and Molecular Biology & Genetic Engineering under the exceptional guidance of Prof. G. K. Garg.

The second part of my research was performed at Molecular Biophysics Unit, Indian Institute of Science, Bangalore during **1992-1998** under the supervision of Prof. Manju Bansal. I had developed an algorithm, called **HELANAL**, to characterize α-helix geometries in proteins in terms of their local parameters. In this algorithm, geometry of an α-helix is characterized by computing local helix axes and local helix origins for four contiguous C^{α} atoms, using the procedure of Sugeta and Miyazawa (Biopolymers, 1967, 5, 673 - 679), and by sliding this window over the length of the helix in steps of one C^{α} atom. The angles between successive local helix axes can identify local bends or kinks as well as occurrence of smooth curvature in the helix. The local helix origins trace out the path described by the helix in three dimensional space. The local helix origins are reoriented in X-Y plane and the reoriented points are used to fit a circle as well as a line, by least squares method. Based on the relative goodness of line and circle fit to the local helix origins, the helix is classified as being linear or curved. An α-helix is classified as being kinked, if atleast one local bending angle in the middle of the helix is greater than 20 degrees. This procedure was applied to 1174 α-helices with non-identical amino acid sequences found in 225 non-homologous protein chains. These α-helices contain 9 or more residues. My analysis has indicated that most of the α-helices in proteins are smoothly curved. Amino acid composition analysis of the α-helices in various geometrical classes has

revealed several interesting relationships. This indicates that fine geometrical details of protein secondary structural elements are encoded by their amino acid sequences. Furthermore, I had utilized the database of 1174 α-helices to analyze statistical preferences for different amino acids to occur at 15 different positions, namely, N", N' Ncap, N1, N2, N3, N4, Mid, C4, C3, C2, C1, Ccap, C', C", in and around α-helices. Each position has its own characteristic preferences and avoidances for different amino acids. The Ncap position is most exclusive. It prefers six amino acids and avoids eleven others. This work also led to characterization of several sequence and structural helix capping motifs at α-helix termini. The research is described in details in my Ph. D. thesis entitled "Geometry and sequence correlation studies on α-helices in globular proteins" submitted to Indian Institute of Science, Bangalore, in 1997. I was awarded the Ph.D. degree by Indian Institute of Science in 1998. My Ph. D. research work involved an interdisciplinary approach with concepts derived from several fields such as statistics, molecular biophysics, protein conformation, chemistry, molecular biology and computer programming. This work has resulted in four research articles in international journals (please see bibliography) and is well cited. **HELANAL** is freely available to academic community *via* internet (www.lecb.ncifcrf.gov/~kumarsan).

Appendix C: Future Research Plans

I plan to carry out further statistical analyses and computer modeling studies to take advantage of increasing availability of data on protein sequences, structures and thermodynamics. Genome wide comparisons of the protein sequences among thermophiles and mesophiles shall be useful in developing further understanding of the molecular basis of protein thermal stability. Inclusion of protein sequences and structures from psychrophilic organisms will greatly facilitate such studies. Besides these analyses, there is a deeper need to develop a theoretical framework that ties protein sequence/structure data with thermodynamic (calorimetric) data on protein stability. The hydrophobic effect and the electrostatic interactions both contribute towards protein stability. However, the hydrophobic effect is the major force driving protein folding. This dominant nature of the hydrophobic effect is responsible for heat and cold denaturation of the proteins and explains why most proteins, including those from the psychrophiles, are maximally stable around the room temperature. The importance of electrostatic interactions lies in fine-tuning the protein stability. Hence, the differences in stabilities of homologous thermophilic, mesophilic and psychrophilic proteins are expected to be mostly due to the differences in their electrostatics, rather than hydrophobicity. This scheme predicts that altering electrostatics can be useful in manipulating protein stability. To better understand the role of salt bridges and their networks in proteins, I intend to further carry out systematic statistical analyses of the electrostatic free energy contributions by close range electrostatic interactions as well as contribute towards methodology development efforts to provide better models for studying protein electrostatics. Furthermore, it shall be useful to construct electrostatic profiles of the proteins and compare these among thermophilic, mesophilic and psychrophilic proteins. Conformer ensembles generated from long molecular dynamic simulations of proteins would yield important information about the role of electrostatic interactions in protein folding, binding, and stability. Since electrostatic interactions are also important for protein flexibilities, I plan to analyze protein conformer ensembles to understand systemic and segmental flexibilities in protein structure, and their role in protein function.

The discovery that a given protein segment may be critical both for protein folding and function is exciting. Such a coupling is evolutionarily attractive because the proteins that contain such critical building blocks need to guard against mutations in a single segment to preserve both their fold and function. I intend to further probe this relationship between protein folding and function. Initial studies in this direction would be based on multiple sequence and structural alignments for the well-studied proteins in our critical building blocks library. Experimental inputs from fragment complementation studies shall also be useful. The structural consequences of removing critical building blocks from the proteins would be studied by molecular dynamics simulations.

The technique of micro-array has found many applications in mapping gene networks, analyzing differential gene expression in normal and diseased (e.g. cancerous) cells or studying development. I intend to collaborate with scientists involved in such research to develop efficient statistical tool to analyze micro-array data. One of the aims of this research would be interpret the gene networks in terms of metabolic pathways involving protein—protein interaction.

Besides these plans, I intend to learn and collaborate with other researchers, particularly those involved in understanding relationships among sequence, structure and function of proteins *via* crystallography, structural bioinformatics, molecular modeling and computational chemistry. I am also enthusiastic about writing grant proposals. Initially, I would like to start with proposals either based on my own research experience or in collaboration with more senior scientists.

Appendix D: Teaching experience and plans

In 1991, I had supervised undergraduate laboratory experiments in general biochemistry course offered to Agriculture, Veterinary and Home Science students at G. B. Pant Univ. of Agriculture and Technology. In 1994, I had acted as a teaching assistant for a course in Fortran Programming (entitled 'Computers in Chemistry') offered to Integrated Ph.D. students at Indian Institute of Science. Presently, I am co-advising a graduate student for her M. Sc. thesis on protein folding and function. In future, I see teaching as one of my major academic activity. I have deep interest in teaching at graduate level. I can contribute towards courses related to molecular biophysics, macromolecular conformation, structure and function, use of computers and molecular graphics in biology, physics of protein folding/binding and structural bioinformatics either independently or in teams. I would also like to guide students for their Ph.D. research. Postdoctoral fellows shall be provided with opportunities to develop academic independence. Since teaching also provides learning opportunity, I shall be enthusiastic about it if given a chance.