_____

Computational Biology and Bioinformatics: A tinge of Indian spice

N. Srinivasan*

Molecular Biophysics Unit; Indian Institute of Science; Bangalore 560 012; India;
N. Srinivasan* - E-mail: ns@mbu.iisc.ernet.in; Phone: +91-80-2293 2837; Fax: +91-80-2360 0535;
* Corresponding author

Published online February 28, 2006

Editorial Message:
Since the birth of *Bioinformation* the proportion of papers published by the Indian authors is substantial. This coincidence made me think about the past and present scenario of Indian computational biology and bioinformatics.

If you had a chance to travel around recently in some of the main roads in the major cities in India such as Hyderabad, Bangalore and Chennai (Madras) in the recent years, you might have noticed banners and posters advertising for bioinformatics or, more broad-based, biotechnology organizations rather like advertisement boards of hotels, cellular phones and computers. Biotechnology and Bioinformatics are popular terms now in India and recently *Nature* released a special outlook section on activities on biological sciences in India. **[1]** The generic term "Biotechnology" which seems to encompass traditional and modern biology is very popular especially among Indian students aspiring for a career in biology. This bioinformatics boom, which looks somewhat subsided now, is compounded by the launch of bioinformatics companies and bioinformatics wings of popular software and pharmaceutical companies – few flourishing, some lingering and many dead.

Suppose you are an enthusiast of protein structures, the most obvious work that comes to the mind is the famous Ramachandran map [**2**], and, perhaps Venkatachalam's β-turn **[3]** which are now text book material. Clearly, the Ramachandran era marks a glorious time of Indian Computational Biology.

Old wine in the old bottle:
Ironically the Ramachandran map resulted out of the provocative criticism on the early model of collagen proposed by Ramachandran and Kartha [**4**], although the main point of the finding, that it is triple helical structure, is fundamentally correct. This criticism was based on the contention that their initial model of collagen contains non-bonded atom pairs that are too closely placed. The result of this provocation is the development of the contact criteria and the famous Ramachandran map **[2]** that was done in association with C. Ramakrishnan and V. Sasisekharan.

Apparently, Pauling and Corey understood the principles of Ramachandran map so well that they felt no need to explain them by a (ϕ,ψ) diagram. **[5]** However, the simplicity of the two dimensional representation is so striking and highly revealing with information content much more than the spectacular and monumental α-helix and β-sheet structures. The basic principle and data that went behind proposition of α-helix and β-sheet structures is the *stability* contributed by hydrogen bonding and the fibre diffraction patterns. The proposition of Ramachandran map emerged from the consideration of *feasibility* of formation of conformations for the minimal system of two-linked peptide units based on the limit to how proximal two non-bonded atoms can get. Considerations of stability and feasibility are fundamentally quite different that they resulted in two independent masterpieces for the structural biologists. Over the past four decades, Ramachandran map has been put to extensive and powerful use in understanding the principles of protein structure, stability and folding beyond the ideal α-helix and ideal β-sheet **[e.g., 6, 7, 8, 9, 10, 11, 12]** aside from its fool-proof use in validating protein structures. **[13]**

Seminal early work from Ramachandran and coworkers on peptide and protein conformation was followed by a number of important contributions in developments of new methods for biomolecular structural research although the visibility of the publications in the international arena was only modest. A robust work on hydrogen bonding defined definitive lengths and angles for considering an interaction as "Hydrogen bond". **[14]** Comprehension of complex protein structures has been a subject of intense study with some simplified viewpoints proposed from Madras and Bangalore. **[15,16]** Such different viewpoints enabled recognition of occurrence of unusual structural patterns such as single helix of collagen type observed for the first time in globular proteins **[17]** which is presently recognized as a motif for interaction between protein modules. Two groups, fittingly both from India, provided an appraisal for the occurrence of conformations in the disallowed regions of the Ramachandran map. **[18, 19]** Sidechain rotamer preferences have been extensively studied over many years by several groups and one of the earliest and well noted work emerged from India. **[20]** Disulfide

engineering has been a popular approach to enhance the protein stability. The MODIP procedure, which involves stereochemical modeling of disulfide bridges [21], is popularly used in several laboratories around the world.

In retrospect, it appears that while much of the Indian work derived inspiration from the intellect of Ramachandran, it also followed the path of structural data analysis provided particularly by the groups of Thornton [22], Chothia [23] and Richardson. [24] The analysis and modeling work was not only done for proteins and peptides, but, also for nucleic acids [eg., 25] and carbohydrates. [eg., 26] Aside the structural data analysis, modeling of molecular recognition [eg. 27] and molecular dynamics has also been significant component of computational biology. [eg. 28]

New spices in the traditional curry:
In the recent times, the strength of India in computing, information technology and software engineering in academia and industry has provided a fresh impetus to computational biology and bioinformatics in India. The current generation of Indian computational biologists effectively exploited these pivotal points and drifted steadily towards global trends such as integration of biological data, development of useful suites of software and databases in biology, generation of new hypotheses on form, functions and models of evolution of biomolecules especially using genome-wide analyses [29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40], neuronal simulations and systems biology.

In a remarkable combination of bioinformatics and biological work, it has been shown that some of the annotated fadD genes, located adjacent to the polyketide synthase genes in the *Mycobacterium tuberculosis* genome, constitute a new class of long-chain fatty acyl-AMP ligases. These proteins activate long-chain fatty acids as acyl-adenylates, which are then transferred to the multifunctional polyketide synthases for further chain extension. [41] In a work by the same combination of bioinformatics and experimental biologists based on precise identification of biological functions of proteins from Pps cluster, they have rationally produced a nonmethylated variant of mycocerosate esters. [42]

Implications of conserved regions in protein folding and stability [43] and their use in function prediction [44] have been recently proposed. The area of protein-protein interactions has been a subject of recent focus. [45, 46, 47] It has been shown for the first time that protein-protein interfaces are not topologically equivalent, in general, if the proteins are distantly related. It has also been shown that variation in protein-protein interactions in members within a superfamily could serve as diverging points in otherwise parallel metabolic or signaling pathways.

[48] Molecular dynamics simulations tailored with analysis of known 3-D structures [49, 50, 51] have provided new insights.

The main course - Post genome sequencing era:
Contribution of genomic data based on sequencing projects undertaken in India has been limited. Indian Initiative for Rice Genome Sequencing with member institutions of University of Delhi South Campus and Indian Agricultural Research Institute participated in the international project on the genome sequencing of rice. [52] Another recent notable contribution is from the Jawaharlal Nehru University Delhi which participated in the genome sequencing consortium of *Entamoeba histolytica.* [53]

One of the earliest analyses of the human genome data from India, along side the work of Tony Hunter and coworkers [54], is the investigations on the complete repertoire of human kinases and discovery of previously unknown kinases and annotation of their functions. [55] Some of the predictions are consistent with the results of subsequent experimental studies. Analysis of kinases from other organisms [56, 57], phosphatases [58, 59] and prediction of sub-cellular localization of proteins [60] have generated many new and experimentally testable hypotheses. Extensive functional annotation of human X chromosome has been made by a team involving a large number of Indian researchers based in Bangalore. [61] A Bangalore-based company, Jubilant Biosys, has developed several useful resources and a number of these are licensed to companies around the world. The resources include *PathArt* which builds molecular interaction networks from curated databases, *Kinase ChemBioBase* which is a comprehensive database of small molecules that focuses on kinase targets and *GPCR Ligand Database* which is a small molecule ligand database on GPCR agonists/antagonists.

Another Bangalore-based company, Strand Life Sciences, has been focusing on new suites of software. One of their products *Admetis* is a platform for modeling and predicting drug-relevant properties of molecules *In Silico*. They design and synthesize compound libraries focused on specific targets. It employs novel machine learning-based methods for designing these libraries and partners with leading chemistry companies for synthesis. Strand Life Sciences uses its proprietary prediction tools for annotating these libraries. *Acuris* is a tool for annotation and management of gene-related data and it can automatically gather and present gene-related public information and literature curation.

_____

New flavors:

Much of the Indian work described so far is confined to computational studies at the molecular level. This feature is perhaps consistent with the global bias of computational biology at the molecular level. Systems biology provides an attractive direction to move higher in the hierarchy. The signs of systems biology work from India are beginning to appear. A recent work forms an excellent example. [62] Flux balance analysis has been performed on the mycolic acid pathway of *Mycobacterium tubercolosis*, and this analysis has provided insights into the metabolic capabilities of the pathway. *In silico* systematic gene deletions and inhibitory studies provide clues about proteins essential for the pathway.

A recent nice example of work on computational neuroscience from India addresses the question of how cells maintain changes in the efficacy of synaptic connections between nerve cells, during memory formation, despite molecular turnover, traffic, and biochemical noise. [63] The authors show using computer simulations that there is a self-sustaining switch involving the movement of AMPA receptors to and from the synaptic membrane and more conductance states may arise through interactions with a biochemical switch involving a synaptic protein kinase.

It is very effective to combine computational analysis and modeling with experimental results. One of the ways is to generate reasonable and testable hypothesis from computing and subject it to experimental scrutiny. Irrespective of the proposition from computational work is 'right' or 'wrong', one learns something new and useful about theoretical possibilities. Alternatively, computational biologists could provide a new viewpoint, using modeling and analysis, to an interesting experimental observation. Combining experimental and computational results and to be able to synergize them should be rewarding. In fact, the first paper entirely from India in the prestigious *Cell* marks the effective combination of experimental and theoretical analysis. [64] The authors investigate, using both the kinds of techniques, the size of lipid-dependent organization of glycosyl-phosphatidylinositol-anchored proteins (GPI-APs). It has been shown that cell surface GPI-APs are present as monomers and a smaller fraction as nanoscale cholesterol-sensitive clusters. Although there are a few other interesting combinations of experimental and computational work [e.g., 41, 42, 65] it is something that computational biologists in India would like to improve upon. Achieving an effective collaboration between expert experimental and computational biologists requires enthusiasm from both the sides to interact with each other and look at the results with an excellent understanding of strengths and weaknesses of the techniques involved.

Strand Life Sciences has made many new developments to facilitate modern biological experimental research. *Sarani* automates large-scale design of optimal oligonucleotide probes for microarray experiments and *Chitraka* is an image analysis and management tool for semi-automatic recognition and quantification of expressed gene spots from microarray experiments. *Sphatika* is a crystal image classification tool for high throughput X-ray crystallography and it classifies protein crystals into two broad categories, one comprising crystal hits and harvestable crystals and the other comprising empty wells, clear drops and precipitates. Some of the well-established Indian software and other technology giants such as Tata Consultancy Services and Infosys have started computational life sciences components. For example, Infosys has a major initiative in drug discovery informatics. Given the high standing and reputation of such companies in the contribution to economy and technology growth in the country their venturing into the life sciences areas is extremely encouraging for world-wide visibility of Indian bioinformatics.

Nutrition & Dessert:

Department of Biotechnology (DBT), an organization closely connected to the central ministry, generously supports bioinformatics centers instituted in many academic organizations in India. These centers were initially primed towards services in bioinformatics and teaching/training. Some of these institution run one year postgraduate diploma program and even two-year Masters program in Bioinformatics. The high quality of these programs is reflected by good achievements of the students in their further academic and industry engagements. In the recent times these centers are encouraged to have a strong research component too. Many small private organizations offer bioinformatics training, which are popular among the undergraduate and masters students. But the quality of the training given by such organizations is a question. The fundamental problem that prevails in such organizations is the lack of teachers who are well-trained and experienced in bioinformatics. There are clear exceptions such as Institute of Bioinformatics and Applied Biotechnology in Bangalore that is supported by the local Government where high quality training is provided in bioinformatics.

In association with computational biologists in the academic institutions in India and with the support of Council of Scientific and Industrial Research,

Government of India the Tata Consultancy Services, a frontline and a long standing software company based in India, has produced a suite of programs called Bio-Suite. Bio-Suite package covers all the major functional areas of Bioinformatics. This package can be used to analyze, formulate, predict and provide solutions to areas such as genomics, protein modeling and structural analysis, simulation and drug design.

Department of Biotechnology offer competitive research grants to the projects in bioinformatics through their taskforce explicitly for bioinformatics. National Bioscience Award instituted by the DBT aims to recognize outstanding research work done in India in the broad area of biological sciences. Among the recent recipients of this award includes, for the first time, the computational biologists. The present day Indian computational biologists are beginning to be visible internationally too with inclusion of an Indian in the editorial board of the journal *Bioinformatics*. The British biomedical research funding agency, the Wellcome Trust, support Indian researchers in the form of Senior Research Fellowship. The Indian recipients during the last 6 years include at least two computational biologists. Such trends would hopefully continue providing good encouragement to the computational biology community in the country. It is hoped that the resurrection of vibrant computational biology [66] in India after a period of modest international visibility in the post Ramachandran period would flourish and grow to newer heights.

References:
[1] Outlook: India *Nature* 436:477 (2005)
[2] G. N. Ramachandran *et al.,* *J Mol Biol* 7:95 (1963) [PMID: 13990617]
[3] C. M. Venkatachalam, *Biopolymers* 6:1425 (1968) [PMID: 5685102]
[4] G. N. Ramachandran & G. Kartha, *Nature* 174**:**269 (1954) [PMID: 13185286]
[5] D. Eisenberg, *Proc Natl Acad Sci (USA)* 100:11207 (2003) [PMID: 12966187]
[6] K. D.Gibson & H. A. Scheraga , *Biopolymers* 4:709 (1966) [PMID: 5946123]
[7] A. S., Kolaskar & V. Ramabrahmam *Int. J. Biol. Macromol.* 3:171 (1981)
[8] M. B. Swindells *et al., Nature Struct Biol* 2**:**596 (1995) [PMID: 7664128]
[9] K. Gunasekaran *et al., J Mol Biol* 275: 917 (1998) [PMID: 9480777]
[10] A. J. Petrescu *et al, Protein Sci* 9:1129 (2000) [PMID: 10892806]
[11] M. C. Vega *et al., Protein Sci* 9:2322 (2000) [PMID: 11206053]
[12] A. E. Serov *et al, Biochemistry* (Mosc) 70:804 (2005) [PMID: 16097945]
[13] G. J. Kleywegt & T.A. Jones *Structure* 4:1395 (1996) [PMID: 8994966]
[14] J. Mitra & C. Ramakrishnan *Int J Pept Protein Res* 17:401 (1981) [PMID: 7309343]
[15] R. Srinivasan *et al., J Mol Biol* 98:739 (1975) [PMID: 1195407]
[16] C. Ramakrishnan & K.V. Soman, *Int J Pept Protein Res* 20:218 (1982) [PMID: 7129755]
[17] K. V. Soman & C. Ramakrishnan, *J Mol Biol* 170: 1045 (1983) [PMID: 6644813]
[18] K. Gunasekaran *et al., J Mol Biol* 264:191 (1996) [PMID: 8950277]
[19] D. Pal & P. Chakrabarti. *Biopolymers* 63:195 (2002) [PMID: 11787007]
[20] T. N. Bhat *et al., Int J Pept Protein Res* 13:170 (1979) [PMID: 429093]
[21] R. Sowdhamini *et al., Protein Eng* 3:95 (1989) [PMID: 2594728]
[22] J.M. Thornton, *Protein Sci* 10:3 (2001) [PMID: 11266588]
[23] C. Chothia, *Annu Rev Biochem* 53:537 (1984) [PMID: 6383199]
[24] J. S. Richardson *Adv Protein Chem* 34:167 (1981) [PMID: 7020376]
[25] G. Gupta *et al., Proc Natl Acad Sci USA* 77:6486 (1980) [PMID: 6935663]
[26] M. Biswas & V.S.R. Rao, *Biopolymers* 19**:**1555 (1980) [PMID: 7417692]
[27] K. Seshadri *et al., J Biomol Struct Dyn* 11:395 (1993) [PMID: 8286064]
[28] D. Mohanty & M. Bansal, *Biophys J* 69**:**1046 (1995) [PMID: 8519959]
[29] S. A. Fernando *et al., Nucleic Acids Res* 32:D125 (2004) [PMID: 14681375]
[30] A. Krupa *et al., Nucleic Acids Res* 32**:**D153 (2004) [PMID: 14681382]
[31] A. Bhaduri *et al., Nucleic Acids Res* 32:W602 (2004) [PMID: 15215459]
[32] V. B. Sreenu *et al., Nucleic Acids Res* 31:106 (2003) [PMID: 12519959]
[33] S. Balaji *et al., Nucleic Acids Res* 29:61 (2001) [PMID: 11125050]
[34] S. Sandhya *et al., J Biomol Struct Dyn* 23:283 (2005) [PMID: 16218755]
[35] B. Anand *et al, Bioinformatics* 21:2821 (2005) [PMID: 15817691]
[36] A. Bhaduri *et al., Proteins* 54: 657 (2004) [PMID: 14997562]
[37] S.B. Pandit *et al., Nucleic Acids Res* 30:289 (2002) [PMID: 11752317]

[38] V. Mallika *et al., Nucleic Acids Res* 30**:**284 (2002) [PMID: 11752316]

[39] A. Krupa & N. Srinivasan *Protein Sci* 11:1580 (2002) [PMID: 12021457]

[40] A. Krupa *et al., Trends Biochem Sci* 28:9 (2003) [PMID: 12517446]

[41] O. A. Trivedi *et al., Nature* 428:441 (2004) [PMID: 15042094]

[42] O. A. Trivedi *et al., Mol Cell* 17:631 (2005) [PMID: 15749014]

[43] T. Prakash *et al., J Mol Biol* 345:937 (2005) [PMID: 15644196]

[44] C. A. Innis *et al., J Mol Biol* 337:1053 (2004) [PMID: 15033369]

[45] R. P. Bahadur *et al., J Mol Biol* 336:943 (2004) [PMID: 15095871]

[46] S. De *et al., BMC Struct Biol* 5:15 (2005) [PMID: 16105176]

[47] M. Guharoy & P. Chakrabarti, *Proc Natl Acad Sci (USA)* 102:15447 (2005) [PMID: 16221766]

[48] N. Rekha *et al., Proteins* 58:339 (2005) [PMID: 15562516]

[49] S. Chakravarthy & R. Varadarajan, *Structure* 7:723 (1999)

[50] C. K. Reddy *et al., J Mol Biol* 314:619 (2001) [PMID: 11846571]

[51] A. Bhinge *et al., Structure* 12:1989 (2004) [PMID: 15530363]

[52] The International Rice Genome Sequencing Project. *Nature* 436:793 (2005) [PMID: 16100779]

[53] B. Loftus *et al., Nature* 433:865 (2005) [PMID: 15729342]

[54] G. Manning *et al., Science* 298:1912 (2002) [PMID: 12471243]

[55] A. Krupa & N. Srinivasan. *Genome Biol* 3:66 (2002) [PMID: 12537555]

[56] A. Krupa & N. Srinivasan *BMC Genomics* 6:129 (2005) [PMID: 16171520]

[57] Anamika *et al., Proteins* 58:180 (2005) [PMID: 15515182]

[58] A. Bhaduri & R. Sowdhamini. *J Mol Biol* 352:736 (2005) [PMID: 16095610]

[59] R .P. Metpally & R. Sowdhamini *BMC Evol Biol* 5: 41 (2005) [PMID: 16022726]

[60] A. Garg *et al., J Biol Chem* 280:14427 (2005) [PMID: 15647269]

[61] H. C. Harsha *et al., Nature Genetics* 37:331 (2005) [PMID: 15800640]

[62] K. Raman *PLoS Comput Biol* 1:e46 (2005) [PMID: 16261191]

[63] H. Hayer & U.S. Bhalla *PLoS Comput Biol* 1:e20 (2005)

[64] P. Sharma *et al., Cell* 116:577 (2004) [PMID: 14980224]

[65] R. Pudi *et al., J Biol Chem* 278:12231 (2003) [PMID: 12540850]

[66] V. S. Gowri & S. Sandhya ,*Bioinformation* 1:94 (2006) [DOAJ: 128565]