RNA Secondary Structure Prediction in Soft Computing Framework: A Review

Shubhra Sankar Ray1, 2, Munia Bachhar1, and Sankar K. Pal1, 2, Fellow, IEEE,

Center for Soft Computing Research, 2Machine Intelligence Unit Indian Statistical Institute, Kolkata 700108, India
shubhra@isical.ac.in, munia.bachhar@gmail.com, sankar@isical.ac.in

***Abstract*—This article provides an overview of the application of certain soft computing tools namely, genetic algorithms (GAs), simulated annealing (SA), and artificial neural networks (ANNs) in certain tasks of RNA secondary structure prediction. Different tasks like prediction of helix, bulge, hairpin curve, internal loop, and multiloop are, first of all, described along with their basic features. The relevance of using soft computing tools to these problems is then mentioned. These are followed by different approaches along with their merits for addressing some of the aforesaid tasks. Finally some limitations of the current research activity are provided.**

Keywords—RNA, combinatorial optimization, dynamic programming, soft computing, genetic algorithms, simulated annealing, neural networks.

I. Introduction

Throughout the last few decades determining RNA structure has gained very much importance, as it is invaluable in creating new drugs and understanding genetic diseases and helps biologists to understand the role of the molecule in the cell [1]. The RNA secondary structure prediction problem (2°RNA) is a critical one in molecular biology. By x-ray diffraction secondary structure can be determined directly [2], but this is difficult, slow, and expensive [3]–[5]. Moreover, most RNAs are currently impossible to crystallize. That is why developing mathematic and computational methods to predict the secondary structure of RNA is very necessary [6].

This article provides an overview of the certain soft computing based techniques that have been developed over the past few years for RNA secondary structure prediction. First we describe the biological basics along with the basic tasks in structure prediction. Next different soft computing tools like genetic algorithm, simulated annealing and artificial neural networks to address them are explained. Finally, some conclusions and future research directions are presented.

II. Biological basics and Different Tasks in RNA secondary structure prediction

An RNA molecule represents a long chain of monomers called nucleotides and each nucleotide consists of a base (any one of adenine (A), cytosine (C), guanine (G) and uracil (U))), a phosphate group and a sugar group [2]. Traditionally, an RNA secondary structure was modeled as a tree. Later, since 1995, RNA structures is treated as a special string and called as string model [7]. The specific sequence of the bases along the chain is called primary structure of the molecule. The structure is usually modeled as a word over the alphabets 'A', 'C', 'G', and 'U'. Through the creation of hydrogen bonds the two groups of complementary bases 'A-U' and 'C-G' form stable base pairs, and are known as the Watson-Crick base pairs [6] while, A-U pairs form two hydrogen bonds, C-G pairs form three hydrogen bonds and tend to be more stable then A-U pairs. Other bases also sometimes pair, especially G-U pair. The G-U pairs are known as 'wobble base pairs and form one hydrogen bond only. The base-pair structure is referred to as the secondary structure of RNA. Generally, the secondary structure is determined discretely by observing whether each base is either paired or not. Base pairs almost always occur in a nested fashion in RNA secondary structure. More formally, a base pair between positions i and j and a base pair between positions i' and j' are nested if and only if i<i'<j'<j or i'<i<j<j'. When non-nested base pairs occur, they are called pseudoknots. For a clear description of the 2°RNA problem, some definitions [6] of RNA structure are needed :

Definition 1: Four-letter alphabet is used to represent an
RNA sequence, which is the primary structure of RNA: R =
$r_1r_2r_3....r_n$, where r_i є {A,U,G,C} and i = 1, 2......n.

Definition 2 (Canonical Base Pairs): In an RNA secondary structure, base pairs are formed as one of the three kinds of pairs, C-G (G-C), A-U (U-A), and G-U (U-G). Base pairs CG (G-C) and A-U (U-A) are called Watson-Crick base pairs. The base pair G-U (U-G) is referred to as a wobble base pair. These three types of pairings are referred to as canonical base pairs.

Definition 3: (i,j) is used to represent the base pair formed by the ith base and the jth base, then the subset of set s={(i, j),1≤i≤j≤n} is called RNA secondary structure if s satisfies the following conditions:
(i,j) is a canonical base pair.
for (i, j) є s, (i', j') є s , if i ≤ i'≤j≤j' , then i=i'.
if (i,j) є s, then j-i>3;

Definition 4: We can call the two base pairs i.j and i'. j' compatible if

(a) i = i' and j = j' (they are the same base pair),
(b) i <j <i' <j' (i.j precedes i'. j'), or
(c) i <i' <j' <j (i.j includes i'. j').

There are six recognized secondary substructure prediction tasks exist and these are 1) Helix, 2) Bulge, 3) Hairpin curve, 4) Internal loop and 5) Multiloop, and 6) External single-stranded regions. A schematic view of various substructures are available in Fig. 1. Base pairs are almost always stacked onto other base pairs in an RNA structure. Contiguous stacked base pairs are called stems (see Fig. 2) and single stranded subsequences bounded by base pairs are called loops [4] (see Fig. 1). A loop at the end of a stem is called a hairpin loop. Single stranded bases occurring within only one side of a stem are called a bulge loop. In an internal loop there are single stranded

978-1-4244-5540-9/10/$26.00 ©2010 IEEE

bases interrupting both sides of a stem. The loops [8] [9] from which three or more stems radiate are called multibranched loops. Single stranded RNA like rRNA folds back itself, forming helical areas interspersed with unpaired, single-stranded areas. The helices are formed when Watson-Crick complementary nucleotides are paired in addition to Guanine and Uracil pairs. Helix generation proceeds first, by iterating through all canonical base pairs (see definition 2) for a given RNA sequence and then attempting by stacking additional base pairs on top of existing ones. Stacked pairs, which form helices, provide stability in the secondary structure. There are two specified constraints for helices. First, each helix must have at least three stacked base pairs. Second, the sequence or loop connecting the two strands must be at least 3 nucleotides long (see definition 3). Since the generation of a helix terminates at the first mismatched base pair, other secondary structures are implicitly defined in the various bulges and loops which remain outside of the stacked pairs. Thus, the determination of helices alone is considered sufficient, in some investigations [3], [5], [10], to account for all other secondary structure elements. The different structural elements which can manifest themselves in the resulting secondary structure include different loops and external bases [4]. Most of the works in RNA secondary structure prediction is based on free energy minimization of a single RNA sequence.

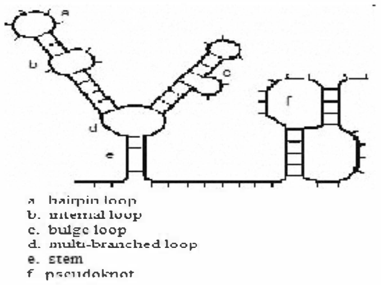


Fig. 1. Different types of secondary substructures in RNA

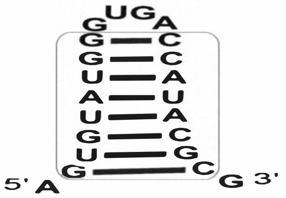


Fig. 2. The region inside the box is an example of a stem. At the top, the unbound nucleotides forms a hairpin loop.

The process, of free energy minimization, predicts the secondary structure based on different thermodynamic models [3] and has been studied since the early 1970s [11]. Assumptions are that the natural fold is a low energy structure and the contributions of RNA secondary structure components, such as stems and loops, are independent and additive. Studying every possible structure for a sequence would solve the folding problem, but it is not feasible, and needs searching techniques to find the minimized energy structure. In an alternative approach, RNA secondary structures can be predicted by comparative sequence analysis using functionally related sequences. In this method, a structure is predicted by searching an alignment for base-pairings that are common to all sequences in the dataset and requires multiple sequences and large sample sizes (typically 1,000 structures). When the number of available sequences with high similarity is small or when there is only a single RNA molecule, prediction of RNA structure based on free energy minimization is the most widely used approach.

RNA tertiary structure is governed by interactions between secondary structures through formation of additional hydrogen bonds or hydrophobic interactions. The interactions that determine secondary structure are generally significantly stronger than those governing tertiary structure. It is generally assumed that the influence of tertiary structure on secondary structure is negligible; consequentially, secondary structures can be determined independently of tertiary structures.

III. Soft computing in RNA secondary structure prediction

One of the first attempts to predict RNA secondary structure using dynamic programming by maximizing the number of base-pairs and using a simple nearest-neighbor energy model is presented in [12]. However, the time and memory requirements of the dynamic programming based algorithms (DPA) [13], [14] are prohibitive. This has prompted researchers to use soft computing tools like GAs [15], SA [16] and ANNs [17] to achieve near optimal results in less computational time and memory requirement. Soft computing based methods try to find low energy stable structures as these structures are most likely to be found naturally. In order to calculate the free energy of the complete structure, the free energy contribution from each substructure is summed. Although it is expected that the lowest energy structure is the natural fold, it is not always true. Often, external interactions affect the resulting structure. Although, DPAs traditionally yield only one optimal structure with minimum free energy, the natural structure may not be the one with minimum free energy. On the other hand, soft computing tool like GAs provide a population of solutions as sub-optimal structures and also makes it possible to investigate not only the minimum free energy structure but also other low energy structures that may be closer to the natural fold.

A. ANN for RNA Secondary Structure Prediction

Artificial Neural Network (ANN) based algorithm to predict the RNA secondary structure by maximizing base pairs are described in [6] and [18]. In [6] Hopfield neural networks (HNN) is used, where, base-pairings configuration of 2° RNA are determined from RNA primary structure, based on the assumption that the most favorable structure is similar to the energetically most stable structure. The more base pairs are found, the more stable is the secondary structure. In order to maximize the base pairs of RNA, all possible base pairs of certain RNA are mapped into an adjacent graph. The vertices of the adjacent graph are the base pairs and the edge of two vertices means that the two base pairs are not compatible (deff4). Maximizing the base pairs is equivalent to find the maximum set of vertices in a graph that each two vertices are not edged. The problem is to find the Maximum Independent Set (MIS) of this graph. An independent set in a graph is a set of vertices, no two of which are adjacent. A MIS is an independent set whose cardinality is the largest among all independent set of a graph [6]. The problem of finding a MIS is NP-complete.

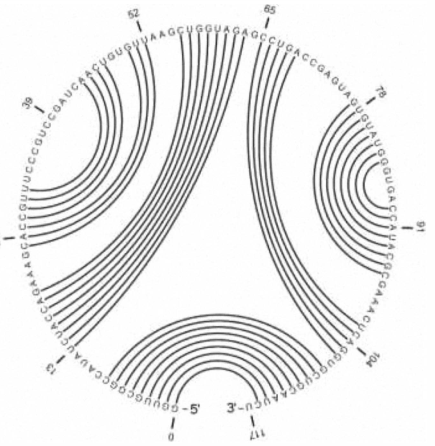


Fig. 3. A circular representation of RNA secondary structure [6].

Considering an adjacent graph of RNA base pairs, it is equivalent to certain circle graph of RNA bases. The nucleotides here are stretched out uniformly along the circumference of a circle and the base pairs are represented by circular arcs that link paired bases and meet the circle at right angles. A circular graph representation of RNA secondary structure can be seen in Fig. 3. Finding a MIS in the adjacent graph is equivalent to finding the maximal planar subgraph of a corresponding circle graph, in which an arc stands for a base pair. The transition of this problem is also illustrated in [6]. HNN is used to find the maximal planar subgraph of a corresponding circle graph. A cost function, termed energy, which is a measure of system-wide constraint violation is used here. A unit's contribution to the networks energy can be computed locally by the following equation:

$$\Delta E_k = E(a_k = 0) - E(a_k = 1) = (\Sigma_i a_i \omega_{ki}) \qquad (1)$$

Where a_k is the activation level of the *i*th unit, and ω_{ki} is the connection weight between the *i*th and *j*th units. The unit turns on/off depending on which state lowers the networks energy. Since the absolute value of the energy is bounded by the weights, the search is guaranteed to converge, if asynchronous node updating is used.

In the algorithm, *n* neurons are used to represent the n arcs of a circle graph, where each neuron performs the following binary function:

$V_i = 1$ if $U_i > 0$, 0 otherwise

Where V_i and U_i are the output and input of the *ith* neuron.

V_i=1 means that the *i*th arc is not embedded in the circle graph, V_i=0 indicates that the *i*th arc is embedded in the circle graph. The motion equation of *i*th neuron is given as:

$$dU_i/dt = A(\Sigma^n{}_i\, d_{ij}(1 - V_j)(distance(i))^{-1}(1 - V_j)p(i)^{-1} - Bh(\Sigma_i{}^n\, d_{ij}(1 - V_j))V_i p(i)) \qquad (2)$$

where, $d_{xy}=1$ if *x*th arc and the *y*th arc intersect each other in the circle graph, 0 otherwise.

In [18], class information of RNA in the initialization of Hopfield network is introduced as secondary structures of non-coding RNAs are believed to be conservative on the same class. The work is otherwise similar to that in [6]. As the initialization is improved with class information, experimental results are also found superior to the related work.

B. GAs for RNA Secondary Structure Prediction:

The possibilities of using GAs for the prediction of RNA secondary structure were investigated in [7], [19]. The implementations used a binary representation for the solutions (chromosomes in GAs). The algorithm, using the procedure of stepwise selection of the most fit structures (similarly to natural evolution), allows different models of fitness for determining RNA structures. The analysis of free energies for intermediate foldings suggests that in some RNAs the selective evolutionary pressure suppresses the possibilities for alternative structures that could form in the course of transcription. The algorithm had inherent incompatibilities of stems due to the binary representation of the solutions.

Wiese et al. [10] used GAs to predict the secondary structure of RNA molecules, where the secondary structure is encoded as a permutation of helices to overcome the inherent incompatibilities of binary representation for RNA secondary structure prediction. At first, a set of all potential helices, H, is generated from a given primary RNA sequence by a helix generation algorithm using a thermodynamic model. The RNA structure prediction problem then becomes a combinatorial optimization problem of finding the subset of helices from a set of feasible helices leading to a minimum energy structure. Each helix in H is indexed with an integer ranging from 0 to n−1, n being the total number of generated helices. Each chromosome of GA is then encoded by a permutation of these integers and provides a solution for RNA secondary structure. For example, assuming set H contains five helices and 0, 1, 2, 3, 4 and 3, 1, 4, 0, 2 are two possible structures. Depending on how the individual helices conflict, both permutations could result in vastly different structures. Helix conflicts are eliminated by decoding the permutation from left to right. The helix specified at each point in the permutation is checked for conflicts with helices to its left. If no conflict is found, the helix is retained; otherwise, it is discarded. This process ensures that each predicted helix does not share nucleotides with any other helix in the subset. In order to calculate fitness of a chromosome, i.e. the free energy of the complete structure, the free energy contribution from each substructure is summed. Three different operators of GAs, i.e. selection, crossover, and mutation are then applied to a population of chromosomes in a elitist model framework. Different combinations of crossover and mutation probabilities ranging from 0.0 to 1.0 in increments of 0.01 and

0.1 were tested for 400 generations with a population size of 700 (maximum). At the end of the process, the program yields a set of chromosomes with high scores (number of base pairings), containing with high probability structures analogous to the native RNA structure. Experimental results on RNA sequences of lengths 76, 210, 681, and 785 nucleotides were provided. It was shown that the Keep-Best Reproduction operator has similar benefits as in the traveling salesman problem domain. A comparison of several crossover operators was also provided.

The work in [3], is similar to [10] where, a population of chromosomes evolves by selection, crossover, and mutation. The main difference between [10] and the recent method [3] has been the use of better crossover and mutation operators and incorporating state-of-the-art thermodynamic models to calculate the free energies. In [3], experimental results are provided by comparing the predicted structures with 19 known structures from four RNA classes.

A massively parallel GA for the RNA folding problem has been used in [20]–[22]. The authors demonstrated that the GA with improved mutation operator predicts more correct (true-positive) stems and more correct base pairs than what could be predicted with a dynamic programming algorithm.

Related works are available in [23] and [1].

C. SA for RNA Secondary Structure Prediction:

A stochastic optimization algorithm like Simulated Annealing (SA) [16] is also used for solving the RNA secondary structure prediction problem [5]. As an iterative search optimization algorithm, it is based on successive update steps (either random or deterministic) where the update step length is proportional to an arbitrarily set parameter which can play the role of a temperature. In an analogy with the annealing process of metals, the temperature is made high in the early stages of the process for faster minimization or learning, then it is reduced for greater stability.

It was first described in [24], how to use SA for identifying RNA secondary structures without considering the free energy minimization approach. In this work an algorithm, using SA, for aligning multiple RNA sequences to identify possible secondary structure, is presented. Dot matrices generated from intra-sequence comparisons are used to obtain possible common secondary structures. A hit probability for dot matrices is calculated and a score function based on this hit probability is defined. Simulated annealing is applied to optimize the score function. A solution set of multiple sequence alignment is also introduced, and the effects of increasing the number of alignment gaps and the alignment length on the solution set are analyzed. An optimized transition rule, which moves two positions in a sequence with each iteration, is applied to increase the rate of convergence.

Schmitz and Steger's [25] used SA for RNA secondary structure prediction using free energy minimization approach.

However, their research able to provide limited results from a single RNA sequence without any quantitative results. Whereas, SARNA-Predict [5] employs a modified SA as its search engine, combining a novel mutation operator, permutation-based encoding for RNA structure and different annealing schedules. Experiments on 33 individual known structures from eleven RNA classes (tRNA, viral RNA, anti-genomic HDV, telomerase RNA, tmRNA, rRNA, RNaseP, 5SrRNA, Group I intron 23SrRNA, Group I intron 16SrRNA, and 16SrRNA) are also shown. The method accepts all decreased energy structures and probabilistically accepts increased energy structures in order to avoid local minima in the search space. The decision to either accept or reject a new structure is based upon the change in structure (ΔEnergy) between new and current structure. If ΔEnergy = 0, the new structure will be accepted. However, if ΔEnergy >0, the new structure will also be accepted with some probability. The Boltzmann distribution is used to determine this probability. The probability of accepting the new structure, when ΔEnergy >0, is given by Eq. 1, where temperature T is the current temperature (a control parameter in the annealing process) and E is the energy state. This distribution expresses the idea that a system in thermal equilibrium at temperature T has its energy probabilistically distributed among all different energy states (or values of ΔEnergy). Even at low temperature, there is a chance of the system being accepted with a probability as follows:

$$Probability[Accept] = e^{-(E_{new} - E_{old})/T} = e^{-\Delta Cost/T} \quad (3)$$

Again when $E_{new} < E_{old}$, this probability is greater than unity; in such cases the change is arbitrarily assigned a probability P=1 (i.e. the system always takes such an option). As a result, this general scheme will most often accept a downward step while sometimes accepting an upward step. Also, if T is decreased slowly enough, SA is guaranteed to reach the best solution. However, it will take an infinite number of moves. If T is high, the algorithm is in an exploratory phase (all moves have about the same probability), and if T is low, the algorithm is in an exploitation phase (the greedy moves are most likely).

IV. Conclusion

An overview of different tasks regarding RNA secondary structure prediction and the relevance of soft computing to handle them are provided. Soft computing, specially GAs, appears to be a very powerful artificial intelligence paradigm to handle the structure prediction tasks. Even though the current approaches in structure prediction are very helpful in identifying patterns and functions of RNAs, the output results are still far from being perfect as simplified models are only considered in most of the works. There are some general characteristics that might appear to limit the effectiveness of soft computing. For example, in GAs, the basic selection, crossover and mutation operators are common to all applications; so researches are now focused to design problem specific operators to get better results and to reduce computational time.

Acknowledgement

Support of the Dept. of Science and Technology, Govt. of India to the Center for Soft Computing Research through its IRHPA scheme is acknowledged.

References

[1] M. Neethling and A.P. Engelbrecht, "Determining rna secondary structure using set-based particle swarm optimization," *IEEE Congress on Evolutionary Computation*, pp. 6134–41, 2006.

[2] E. W. Steeg, *Artificial Intelligence and Molecular Biology*, chapter Neural Networks, Adaptive Optimization, and RNA Secondary Structure Prediction, pp. 121–60, American Association for Artificial Intelligence, Menlo Park, CA, USA, 1993.

[3] K. C. Wiese, A. A. Deschenes, and A. G. Hendriks, "Rnapredict-an evolutionary algorithm for rna secondary structure prediction," *IEEE/ACM Trans. on Computational Biology and Bioinformatics*, vol. 5, no. 1, pp. 25–41, 2008.

[4] Kay C. Wiese, Andrew Hendriks, Alain Deschnes, and Belgacem Ben Youssef, "P-rnapredicta parallel evolutionary algorithm for rna folding: Effects of pseudorandom number quality," *Nucleic Acids Research*, vol. 4, no. 3, pp. 219–27, 2005.

[5] H. H. Tsang and K. C. Wiese, "Sarna-predict: Accuracy improvement of rna secondary structure prediction using permutation based simulated annealing," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. electronic version, pp. 1–14, 2008.

[6] Q. Liu, X. Ye, and Y. Zhang, "A hopfield neural network based algorithm for rna secondary structure prediction," *Proc. of the First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06)*, pp. 1–7, 2006.

[7] V. Batenburg, A. P. Gultyaev, and C. W. A. Pleij, "An APL-programmed genetic algorithm for the prediction of RNA secondary structure," *Journal of Theoritical Biology*, vol. 174, no. 3, pp. 269–280, 1995.

[8] C. B. Do, D. A. Woods, and S. Batzoglou, "Contrafold: Rna secondary structure prediction without physics-based models," *Bioinformatics*, vol. 22, no. 14, pp. 9098, 2006.

[9] J. P. Abrahams, M. Berg, E. Batenburg, and C. Pleij, "Prediction of rna secondary structure, including pseudoknotting, by computer simulation," *Nucleic Acids Research*, vol. 18, no. 10, pp. 3035–44, 1990.

[10] K. C. Wiese and E. Glen, "A permutation-based genetic algorithm for the RNA folding problem: a critical look at selection strategies, crossover operators, and representation issues," *Biosystems*, vol. 72, no. 1-2, pp. 29–41, 2003.

[11] I. Tinoco, O. C. Uhlenbeck, , and M. D. Levine, "Estimation of secondary structure in ribonucleic acids," *Nature*, vol. 230, pp. 362- 267, 1971.

[12] R. Nussinov and A. B. Jacobson, "Fast algorithm for predicting the secondary structure of single-stranded rna," *Proc. of the National Academy of Sciences*, vol. 77, no. 11, pp. 6309–6313, 1980.

[13] T. Akutsu, "Dynamic programming algorithms for rna secondary structure prediction with pseudoknots," *Discrete Applied Mathematics*, vol. 104, pp. 45–62, 2000.

[14] M. Waterm and T. Smith, "Rapid dynamic programming algorithms for rna secondary structure," *Advances in Applied Mathematics*, vol. 7, no. 0196-8858/86, pp. 455–64, 1986.

[15] S. K. Pal, S. Bandyopadhyay, and S. S. Ray, "Evolutionary computation in bioinformatics: A review," *IEEE Transactions on Systems, Man, and Cybernetics, Part-C*, vol. 36, no. 5, pp. 601–615, 2006.

[16] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, "Optimization by simulated annealing," *Science 13 May 1983*, vol. 220, no. 4598, pp. 671–80, 1983.

[17] S. S. Ray, S. Bandyopadhyay, P. Mitra, and S. K. Pal, "Bioinformatics in neurocomputing framework," *IEE Proc. Circuits Devices & Systems*, vol. 152, pp. 556–564, 2005.

[18] Yang Liu Maozu Guo Quan Zou, Tuo Zhao, "Predicting rna secondary structure based on the class information and hopfield network," *Computers in Biology and Medicine*, vol. 39, no. 3, pp. 206–214, 2009.

[19] A. P. Gultyaev, V. Batenburg, and C. W. A. Pleij, "The computer simulation of rna folding pathways using an genetic algorithm," *Journal of Molecular Biology*, vol. 250, pp. 37–51, 1995.

[20] B. A. Shapiro and J. Navetta, "A massively parallel genetic algorithm for RNA secondary structure prediction," *J. Supercomputing*, vol. 8, pp. 195–207, 1994.

[21] B. A. Shapiro and J. C. Wu, "An annealing mutation operator in the genetic algorithms for RNA folding," *Computer Applications in the Biosciences*, vol. 12, pp. 171–180, 1996.

[22] B. A. Shapiro, J. C. Wu, D. Bengali, and M. J. Potts, "The massively parallel genetic algorithm for rna folding: Mimd implementation and population variation," *Bioinformatics*, vol. 17, no. 2, pp. 137–148, 2001.

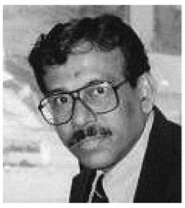
[23] K. C. Wiese, A. A. Deschenes, and A. G. Hendriks, "Rnapredictan evolutionary algorithm for rna secondary structure prediction," *Nucleic Acids Research*, vol. 5, no. 1, pp. 25–41, 2008.

[24] J. Kim, J. R. Cole, and S. Pramanik, "Alignment of possible secondary structures in multiple rna sequences using simulated annealing," *Comput Appl Biosci.*, vol. 12, no. 4, pp. 259–67, 1996.

[25] M. Schmitza and G. Steger, "Description of rna folding by "simulated annealing"," *Journal of Molecular Biology*, vol. 255, no. 1, pp. 254–66, 1996.

Shubhra Sankar Ray (www.isical.ac.in/~shubhra) is an Assistant Professor at the Indian Statistical Institute, Kolkata, India. He received the M.Sc. in Electronic Science and M.Tech in Radiophysics & Electronics from University of Calcutta, Kolkata, India, in 2000 and 2002, respectively, and received a Ph.D. in Computer Science and Engineering fro the Jadavpur University in 2008. He worked as a Visiting Research Fellow at the Center for Soft Computing Research: A National Facility, Indian Statistical Institute, Kolkata, India, in 2006-2008, and Post Doctoral Fellow at the Biophysics Division, Saha Institute of Nuclear Physics, Kolkata, India, in 2008-2009. His research interests include soft computing, bioinformatics, evolutionary computation, neural networks, and data mining.

Munia Bachhar is a Visiting Research Fellow at the Center for Soft Computing Research: A National Facility, Indian Statistical Institute, Kolkata, India. She received the B.Tech in Computer Science from West Bengal University of Technology, Kolkata, India, in 2007 and M.Tech in Computer Science fromUniversity of Calcutta, Kolkata, India, in 2009. Her research interests include bioinformatics and Soft Computing.



Sankar K. Pal (www.isical.ac.in/~sankar) is the Director and Distinguished Scientist of the Indian Statistical Institute. Currently, he is also a J.C. Bose Fellow of the Govt. of India.He received a Ph.D. in Radio Physics and Electronics from the University of Calcutta in 1979, and another Ph.D. in Electrical Engineering along with DIC from Imperial College, University of London in 1982.

He worked at the University of California, Berkeley and the University of Maryland, College Park in 1986-87; the NASA Johnson Space Center, Houston, Texas in 1990-92 & 1994; and in US Naval Research Laboratory, Washington DC in 2004. Since 1997 he has been serving as a Distinguished Visitor of IEEE Computer Society (USA) for the Asia-Pacific Region.

Prof. Pal is a Fellow of the IEEE, TWAS, IAPR, IFSA and all the four National Academies for Science/Engineering in India. He is a co-author of fifteen books and more than three hundred research publications in the areas of Pattern Recognition and Machine Learning, Image Processing, Soft Computing, Data Mining, Web Intelligence, and Bioinformatics. He has received several coveted awards in India and abroad. He serves(ed) in the editorial board of several International journals including IEEE, IET and Science Direct