



Protein Fold Recognition Algorithms-A General Survey

T. KALACHELVI¹ and P. RANGARAJAN²

¹Research Scholar, Sathyabama University, Chennai, India.

²Professor and Head/IT, R.M.D.Engineering College, Kavaraipettai,India.

(Received: January 03, 2013; Accepted: January 17, 2013)

ABSTRACT

Protein folding is considered as a significant confrontation in biological and protein research. This confrontation is interrelated to the fact that the conventional computational approaches are not potent enough to search for the appropriate structure in the large conformational space of protein. This insufficiency of the computational methods is a major hindrance in facing the protein folding problem. Trying to solve the problem, many researchers have examined the efficiency of the protein threading technique. Perhaps, parallel evolutionary methods for protein fold recognition have also been used. Typically, protein folding accredits to how a protein amino acid sequences with respect to the physiological conditions, folds into a three-dimensional structure called as native state. In this paper, various algorithms that have been framed for protein structure prediction. Furthermore, a survey of parallel evolutionary models for protein fold recognition has also been provided. The result of this survey showed that evolutionary methods can be effectively used to resolve the protein folding problems.

Key words: Protein folding model, Protein threading, Evolutionary methods, Optimization, Genetic algorithm.

INTRODUCTION

It is a great confront for nowadays biologists to envisage the three-dimensional structure of a protein with respect to its linear sequence. As is well-known, proteins are the substantial key molecules in all life processes. Proteins are made up of amino acid chains that constitute 20 different amino acids that are folded into distinctive three dimensional protein structures. These structures are predicted by their sequence of amino acids. Moreover, knowing and understanding the protein structures will have a

remarkable impact on perceptive of biological processes, biotechnological inventions and medical discoveries.

Meanwhile, there are two experimental methods avail for the determination of the three-dimensional protein structure from its amino acid sequence: Nuclear Magnetic Resonance (NMR) and X-ray crystallography. Regrettably, these methods are not adequate enough and that is because of the fact that they are time-consuming and expensive. As a result, there is a need for a hasty and reliable computational method to

determine structures from protein sequences, particularly since the number of completely-sequenced genomes is mounting very fast.

Biologists have accepted that proteins could have similar structural folds yet if they have no sequence or functional similarity. Factually, the total number of structural folds in nature is extremely small when compared to the number of recognized protein sequences. It can be elaborately stated as, fold recognition methods try to identify the structural fold of a protein from a structure template library, which has given its sequence information then produce an alignment between the recognized template protein and the query, from which the structure of query protein can be determined. Fold recognition approaches are much efficient specifically in the following cases:

- When the protein sequence does not have any primary sequence similarity to any other sequences with a known structure.
- When some model from the protein structure library represents the true fold of the protein

sequence.

Though there have been many evaluations and developments of the diverse fold recognition methods, researchers have established two main points:

1. Current energy functions are not accurate enough to determine the free energy of a definite conformation;
2. There is no direct computational procedure that can identify the conformation.

The size of the protein conformation space is huge. The authors^{13, 1} have discussed that the protein threading problem is MAX-SNP-hard and NP-complete. Many techniques; such as Molecular Dynamics, Monte Carlo, Neural Network and Genetic Algorithms, have been utilized in protein folding in order to face the computational difficulty. Additionally, researchers of the paper^{14, 28, 12, 24} were used evolutionary methods to solve the protein fold recognition problems. On the other side, researchers^{2, 5, 8, 19} have used some parallel methods to solve the same problem.

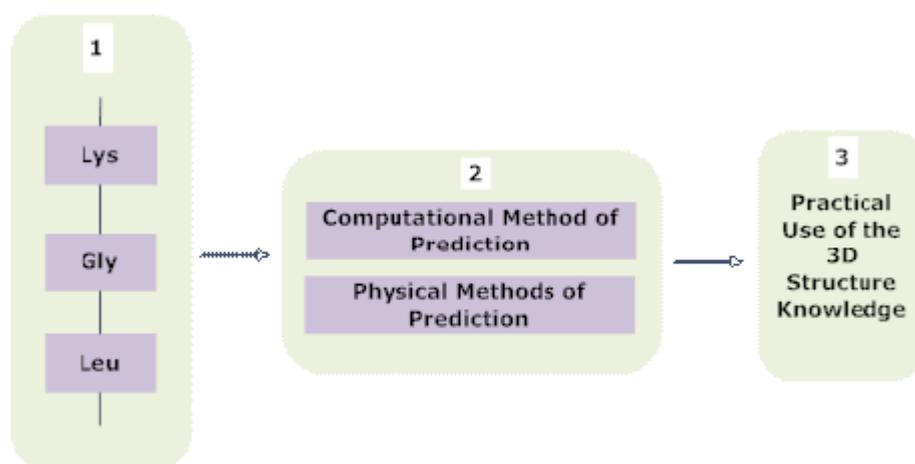


Fig. 1: Protein Structure Prediction

The figure presented above demonstrates the protein structure prediction method from a primary protein structure to the 3D structure. Protein fold recognition methods effort to recognize the appropriate template from a structure template collection for a query protein and produce an alignment between the query and

the identified template protein, through which the structure of query protein can be determined. Moreover, protein fold recognition using the protein threading technique has established a great success⁹. There are four phases for the threading technique for protein folding for an amino acid sequence.

- Phase 1 Construct a protein structure template library
- Phase 2 Design a scoring function to determine the fitness between the template and target sequence
- Phase 3 Design a proficient algorithm for searching
- Phase 4 Find the best alignment between the template and target sequence by minimizing the scoring function.

Further, aligning the query to the template is the key element of the protein threading problem. The following step is to make out the best alignment among all possible alignments between the template and the query protein, and that is by looking for an alignment that generates a proper score function²⁹. Apparently, a query can be defined as a sequence of amino acids of a protein. However, a template is the three-dimensional coordinates of all atoms for each amino acid in the protein sequence which is termed as a series of cores (such as α -helix, β -sheet), links, loops and turns. The process of threading a query against a template is to find out which basic folds the amino acids of the protein query can fit and then calculate the free energy of the query⁸. Generally, the word threading implies that the sequence is dragged step by step through each location on each template, but in fact the process is based on searching for the best alignment of the sequence on that template, as determined by some scoring function. Figure 2 shows the protein threading process.

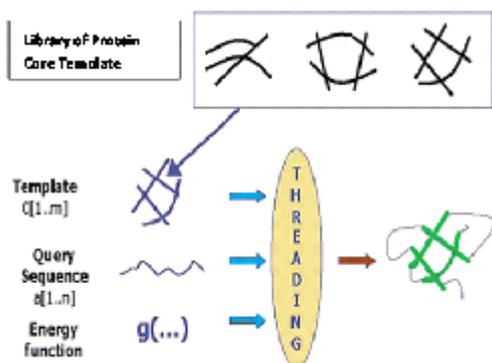


Figure 2: Protein Threading Process

Threading is a complicated computational problem and has been illustrated and proved to be NP-complete¹³ and hence should be directed by effectual heuristics. Also it has been evidenced that the protein threading problem is MAX-SNP-hard, which represents that it cannot be estimated to an arbitrary precision in polynomial time¹.

There are many recent studies that have been focusing on protein structure prediction⁷ and protein folding with optimized methods for acquiring appropriate results. The major intention of this survey is to analyze those studies and project the adept methods, pros and cons in that.

METHODS AND METRIALS

Protein Fold Recognition Method – Training Models

Many researchers have tried variant techniques such as Molecular Dynamics, Genetic Algorithms, Monte Carlo and Neural Network in order to overcome the computational difficulty of protein fold problem. However, this section discusses different successful methods for protein fold recognition.

Neural Network

In¹⁰, the authors introduced a novel method for fold recognition. This method uses a conventional sequence alignment algorithm to produce alignments, which are then determined by a method derived from threading techniques. Moreover, in order to generate a single measure of confidence in the developed prediction method, each threaded model is evaluated by a neural network. Here, the study of authors can be divided into three phases:

1. Sequence Alignment
2. Calculation of salvation terms and pair potential
3. Alignment evaluation using a neural network

Further, the author¹⁰ implemented GenTHREADER Protocol and GenTHREADER program. This approach has been applied to the genome of *Mycoplasma genitalium*. The results show that more than 46% of the proteins are

derived from the identified protein coding regions that have a significant relationship to a query protein of known structure. Perhaps, only one domain of the protein can be predicted that is giving a total coverage of 30% when evaluated as a fraction of the number of amino acid residues in the entire proteome.

The authors claimed that the speed of this method, along with its low false-positive rate and sensitivity makes it ideal for routinely predicting the structure of all the proteins in a translated proteome.

In accordance with further developments, the authors claimed that this approach could be extended easily to take into consideration any number of input metrics and any sources of sequence-structure information. It can be noted that GenTHREADER is able to produce structurally comparable models for one-half of the targets, but notably precise sequence-structure alignments were produced for only one-third of the target protein sequences. Another note is that it is capable to predict the correct answer for the vast majority of the facile targets, if a structurally related fold was present in the server's fold libraries. Nonetheless, among the hard targets it is capable to produce parallel models for only 40% of the cases, half of which had a significantly precise sequence-structure alignment.

Kuang Lin *et al.*¹¹ have trained an Artificial Neural Network model to predict the compatibility of amino acid sequences with structural environment. The authors called their program as TUNE (Threading Using Neural nETwork). But, their model is not concerned the training procedure of native protein structures discrimination. Further, they tested their model on the discrimination of native 3D structure and protein decoy, its performance is equivalent to pseudo-energy functions with atom level structural discrimination, better than the two functions in accordance with residue level structural descriptions. Moreover, they used the protein structure classification CATH to select training and test the sample sets. All the native structures given in the decoy sets are used for assessing ANN models.

In¹⁵, the authors have enhanced and benchmarked GenTHREADER method. Their improvements augment the number of remote homologies that can be identified with a low error rate that imply a higher reliability of score which also enlarge the quality of the models enhanced. Nan Jiang *et al.*,¹⁷ proposed a novel fold recognition model with mixed environment-specific substitution mapping (MESSM) with three key features:

- (i) A structurally-derived substitution score is produced using neural networks.
- (ii) A mixed environment oriented substitution mapping is developed by integrating the structural-derived substitution score with sequence contour from well-developed sequence substitution matrices.
- (iii) A support vector machine is incorporated to measure the significance of the sequence-structure alignment.

They examined their model on two benchmark problems namely, Wallner's Benchmark and Fischer's Benchmark, the model MESSM was perceived to lead to a good performance on protein fold recognition.

Support Vector Machine (SVM)

Xu²⁷ presented a Support Vector Machine (SVM) regression method to directly predict the alignment precision of a sequence template alignment. The authors invoked experiments on a large-scale benchmark using the Support Vector Machine (SVM) regression approach. The authors claimed that experimental results show that SVM regression technique has much better performance in both specificity and sensitivity than the composition corrected Z-score method and SVM regression technique also performs better than SVM classification method. Additionally, SVM regression method facilitates the threading program to execute faster than the composition-corrected Zscore method. Sangjo Han *et al.*,⁶ presented an alternative method for estimation of importance of the alignments. They took a query of a protein and arranged it to a template of length n in the protein fold library, and then this alignment is given into a feature vector of length $n+1$, which is then evaluated by Support Vector Machine (SVM). The output of SVM is further transformed to

Table 1: Technique based protein fold recognition approaches

Technique	Predominant Contribution of Papers
Neural Networks	GenTHREADER: An adept and reliable protein fold recognition approach for genomic sequences [19] Enhancement of the GenTHREADER approach for genomic fold recognition [15] Protein Fold Recognition using Neural Networks and SVM [17], introduced a novel method called Mixed Environment-Specific Substitution Mapping (MESSP)
Support Vector Machine (SVM)	Fold Recognition by Alignment Accuracy [27] Fold recognition by blending profile- profile alignment and support vector machine[6] Protein Fold Recognition Using Networks and SVMs[17]
Bayesian Networks	Bayesian Network model for protein fold and remote & super family recognition [22]
Structural Pattern-based Method	A method (SPREK) was developed for the evaluation of protein models based on residue packing interactions [20]
Genetic Algorithm	Potential of genetic algorithms in protein folding and protein engineering simulation [4] Protein folding simulations using GA [25] A framework of genetic algorithms for protein structure prediction [24] Chaotic Clonal Genetic Algorithm for Protein folding model[30] A new median selection strategy for GAs has been implemented for protein folding problem [16]
Monte Carlo	Evolutionary Monte Carlo approach for protein folding Simulations [14]

Table 2: Parallel Evolutionary Method Based Protein Fold Approaches

Method	Predominant Contribution of Papers
Parallel Genetic Algorithm	A parallel hybrid GA for 3-D peptide structure prediction[3] Multiple molecular sequence alignment by parallel hybrid genetic algorithm [2] Alignment of multiple protein sequences by parallel hybrid genetic algorithm [19] Parallel evolution strategy for protein threading[8] Parallel evolution strategy on grids for the protein threading problem [18]
Evolutionary Approach	Multi-class protein fold recognition using multi-objective evolutionary algorithms[5] Artificial Bee Colony for Two-dimensional Protein folding [31]
P-Rna Predict Approach	A detailed analysis of parallel speedup in P-RnaPredict [26]
Probabilistic Roadmap Methods	Parallel protein folding with STAPL[23]

a subsequent probability that a query sequence is correlated to a template, given SVM output. With respect to their results, the new technique gave considerably better performance than PSI-BLAST and profile-profile arrangement with Z - score scheme. The authors stated that the cause that SVM worked so well is related to the transitional sequence prediction and its capability to identify the essential features among alignments of remotely related proteins.

Bayesian Networks

Raval *et al.*,²⁰ demonstrate a Bayesian network approach for protein fold and super family assimilation. The Bayesian network approach is a structure, which combines probability theory and graphical representation that includes, as a particular case, hidden Markov models²⁰. Further, the authors introduced a new implementation of a Bayesian network that can be trained amino acid sequence, secondary residue accessibility and structure for proteins of known three-dimensional structure. They stated that the cross validation examinations using Bayesian classification present that the Bayesian network model which employs structural information outperforms a hidden Markov model trained only on amino acid sequences.

Structural Pattern-based Methods

In²², Taylor and Jonassen developed an approach for assessment of protein models dependent on residue packing interactions. Moreover, their method was described to estimate the register of a sequence on a structure dependent on the matching of structural patterns against a library framed from the protein structure databank. The computer program that incorporated the method is called SPREK (Sequence-structure Pattern-matching by Residue Environment Comparison). The authors alleged that the performance of SPREK on the decoy approaches was correspondent to those acquired with more complex methods. Compared to earlier methods, their approach is very undemanding. There are no oversized tables of potentials or any large weight parameters. Despite its minimalism, their method did not abandon structural information as occurs in the majority of methods that take only pair-wise residue interactions on the account. The authors

maintained a portrayal of the structure environment around a residue that includes the sequential order of the residues and their secondary structure state in the environment. A main advantage of their method is its aptitude to operate using only with the α -carbon atom positions.

Evolutionary Methods

Genetic Algorithms (GA)

The first study to initiate genetic algorithms to the field of protein structure prediction was given in^{4,25}. The authors introduced GAs as a novel tool to study proteins. Their research revealed that the genetic algorithm simulation which categorized the important folding constrictions as overall hydrophobic packaging and betaphilic propensity of the residues for transpositions attained a unique fold.

Unger and Moul²⁴ have framed a genetic algorithm search procedure appropriate for use in protein folding process. Moreover, they used genetic algorithms to fold proteins on a two-dimensional square lattice in the protein HP model. Figure 3 reveals the sample protein residue chain with energy -4. The white square presents hydrophilic residue, while the black represents the hydrophobic. The solid line depicts the protein sequences, whereas the dashed line identifies hydrophobic-hydrophobic (HH) contacts.

They also maintained a population of arrangements of the polypeptide chain and altered the conformations by the process of mutation, in the aspect of conventional Monte Carlo steps and crossovers, in which the parts of polypeptide chain are exchanged between conformations. For protein folding on a simple two-dimensional lattice it was brought that the genetic algorithm is dramatically greater to conventional Monte Carlo methods. Further, Schulze-Kremer^S and Tiedemann^U²¹ used a genetic algorithm to explore energetically and structurally favorable conformations. Hybrid protein representation is used in the proposal that comprises three operators manage the protein genes and its fitness function in accordance with the simple force field.

Yadgari *et al.*,²⁸ directed the genetic algorithm model used to carry out sequence to

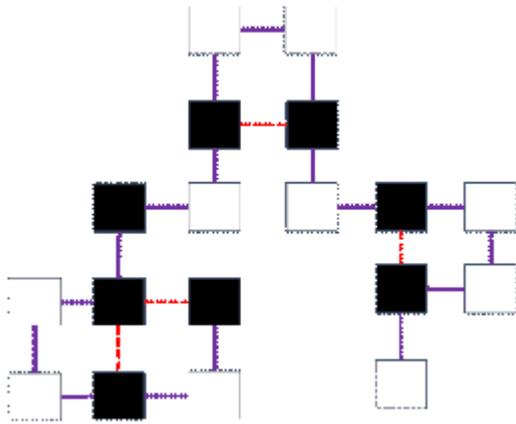


Fig. 3: Sample Protein Chain

structure alignments. In their investigation, the sequence-structure pairs were taken from a database of structural alignments where the sequence of a protein was threaded through the structure of the other. In this process, an appropriate representation has been described in which genetic operators can be implemented effectively. Their representation usually consists of numbers zeros and ones or any integer number. The authors gave that the algorithm performance is evaluated for a set of sequence-structure pairs. The consequences of changing operators and parameters are explored and evaluated. The data they have demonstrated designate that the Genetic Algorithms technique is an efficient and feasible approach for threading.

Further, the authors stated that genetic algorithms threading is moderately robust and does not overly dependent on the particular assortment of parameter or operators. Unger²⁴ described the problem of protein structure prediction and protein alignments by using GAs. It is broadly recognized that one of the major obstacles in representing this question is that the "standard" computational methodologies are not influential enough to search for the correct structure in the large conformational space. Genetic algorithms, a supportive computational method, have been successful in various difficult computational tasks. Thus, it is not astonishing that in recent years several studies were accomplished to explore the possibility of using GA to concentrate on the protein structure

prediction problem. This study describes how a general framework of genetic algorithms can be used for protein structure prediction. With this framework, the major studies that were published in recent years are conferred and compared. Applications of genetic algorithms to the related protein alignment problems are also mentioned in this paper. The rationale of the necessity genetic algorithms are suitable for protein structure prediction is also presented. The author described that GAs are effective general search algorithms and as such are suitable for any optimization problem that includes problem related to protein folding. The author added some improvements to be made to GA techniques to progress performance. An interesting option to explore within the GA framework is to compose a distinction between the energy function and the fitness function. In this way it might be potential to accentuate different aspects of the fitness function in dissimilar stages of folding. Another option is to introduce unambiguous memory into the emerging substructure, such that substructures that have been beneficial to the structures that furnish them will get more level of resistance from changes. Protein structure prediction is defined as the determination of tertiary protein structure by using the information of its primary structures³². There described that there are two important issues in protein structure prediction. The first issue is designing a structure model and the second is the design of optimal technology.

M.V.Judy and K.S.Ravichandran¹⁶ developed a new intermediate selection policy for genetic algorithms and invoked it for protein folding problem. Further, the authors proposed a new transitional selection step, which is called as Modified Keep-Best Reproduction (MKBR) in order to overcome the problem that the parents may be inferior to the children as it is recognized in GA in practice. The novel selection method assures that new genetic data is entered into the gene pool, as well as good preceding genetic material is being conserved. They have also presented the dominance of modified keep-best reproduction on numerous instances of the protein folding problem, which not only finds the best solution, but also determines them faster than the standard generational replacement systems. While

considering about structure model, Amino acids are the building blocks of proteins and that is defined as the molecule contains an amine and carboxile groups. Depending upon the structure, size, electric charge and solubility constraints of amino acid side chains, they can be classified under either hydrophobic or hydrophilic. The hydrophobic and hydrophilic can also be termed as the residues of proteins. The energy determination for protein structure model is based on the counting of every two hydrophobic residues that are non-successive in the protein sequence and adjacent neighbors on the lattice³⁰. The work in³¹, introduced ABC (Artificial Bee Colony) optimization for 2D protein folding by applying it to HP lattice model. The reliability of the process could be further improved by banding some efficient conceits.

Evolutionary Monte Carlo

Monte Carlo methods have conventionally been employed to attend to the protein folding problem. The algorithm is based on minimizing the energy function, using a path that does not fundamentally follow the natural folding pathway. The genetic algorithm technique incorporates many Monte Carlo conceits²⁴. Traditional molecular-dynamics and Monte Carlo tends to get trapped in local minima, so that the native structure cannot be displaced and the thermodynamic quantities cannot be determined precisely¹⁴. To resolve this problem, Liang and Wong¹⁴ proposed an Evolutionary Monte Carlo (EMC) methodology for protein folding simulations. Further, the authors demonstrated that EMC can be employed successfully in the simulations of protein folding on simple lattice models and to find the ground status of a protein.

Parallel Evolutionary Methods (PEM) for protein fold recognition

Many researchers used parallel approaches to solve the protein fold recognition problem in current studies. Perhaps, some researchers also used parallel methods to resolve RNA sequence problem. Basically, there are three domains of biological sequences: DNA, RNA and protein. Some researchers mainly focus on the alignment in one domain. Nevertheless, the method can be easily extended to handle other

domains. Thus, in the following sections, some parallel evolutionary techniques for biological structure prediction will be described.

Parallel Hybrid Gas

Carpio *et al.*,³ were initially demonstrated a parallel hybrid genetic algorithm for three-dimensional polypeptide structure determinations. Their earlier research is based on a simple genetic algorithm, which was inadequate to produce better fit conformers, thus the authors have proposed an improvement in two significant aspects:

- ✓ Parallelization of the inventive procedure to enrich the assortment of conformers in the population
- ✓ Hybridization of the plain genetic algorithm has been developed to process the atoms of the side chains of protein.

In³, it is claimed that a comparison of the best fit individual after the 500th generation attained by the hybrid genetic algorithm shows more appropriate level of evolution of the method. Further, Nguyen *et al.*¹⁹ proposed a parallel hybrid genetic algorithm for illustrating the multiple protein sequence arrangement problem. Moreover, they demonstrated a new GA-based method for multiple protein sequence alignment. The authors described that experimental results of benchmarks from the BALiBASE revealed that the proposed method is superior to MSA, SAGS and OMA methods with consideration to quality of running time and solution. It can be to determine multiple sequence alignment as well as evaluating cost functions.

Island Parallel Gas

Anbarasu *et al.*,² developed an evolution-based approach for multiple molecular sequence arrangement. It is stated in the paper that the approach is completely based on the island Parallel Genetic Algorithm (iPGA) that confides on the fitness distribution over the population of protein sequence alignments. Further, the algorithm explores for an alignment among the independent evolving populations by optimizing weighted amount of pairs-objective function which determines the alignment quality.

Some of the most broadly used multiple

molecular sequence alignment packages like Mutal, Pileup and ClustalW are based on dynamic programming. They have some advantages of being simple and fast as well as logically sensitive, but their main problem is the local minimum problem. In their study, the authors depict an iPGA strategy that executes on a distributed network of workstations.

Their parallel GA technique was implemented on PARAM 10000; a parallel machine established at the Center of Development of Advanced Computing, Pune and is revealed to consistently perform well than the sequential genetic algorithm. The algorithm produced alignments that were considerably better than a different method, ClustalW.

Multi-objective fmGA

Earlier research using the Simple Genetic Algorithm (GA), fast messy GA (fmGA), messy GA (mGA) and Linkage Learning GA (LLGA) has made improvement on this problem. However, past study used off-the-shelf software such as GENESIS, GENOCOP and mGA⁵. Day *et al.*⁵ demonstrated a modified fmGA as multi-objective implementation of the fmGA (MOfmGA) and a farming approach for the parallel fmGA for protein structure prediction. The authors concentrated on tuning fmGA in an attempt to augment the effectiveness and efficiency of the procedure in solving a protein structure and in finding enhanced ways to identify secondary structures.

Problem definition, mapping to algorithm domain, protein model representation, tool selection modifications and conducted experiments were determined in this study. They claimed that their improvement of using MOfmGA have been manipulated to scale its efficiency to 4.7 times a computational results and serial run time support their hypothesis that the MO version affords more acceptable results.

Parallel Evolution Strategy

Islam and Ngom[8] framed an evolution strategy for protein threading and also developed two parallel methodologies for fast threading based on an evolution approach for protein threading. The parallelization is completely based on master-

slave architecture. Their approach for protein threading is based on evolution strategy. Moreover, the Single Query Single Template Parallel ES Threading (SQSTPEST) approach threads one query beside each template. The Single Query Multiple Templates Parallel ES Threading (SQMTPEST) technique threads one query over a set of templates.

Alioune Ngom¹⁸ proposed a new evolution strategy for solving the protein threading problem using the test strategy called EST. The author revealed that with recombination, his EST method gave much better results, both in threading time and energy, than an existing genetic algorithm based method. Without recombination, EST is equivalent to the GA based strategy but much faster. The paper also proposed a parallel approach for fast threading; his parallel EST was employed on Grid-enabled platforms for High- Performance Computing paradigms.

RnaPredict Approach

Wiese and Hendriks²⁶ declared a parallel evolutionary algorithm called PRnaPredict for RNA secondary structure prediction. PRnaPredict is a fully parallel accomplishment of a coarse-grained distributed EA for efficient RNA secondary structure prediction and is totally dependent on RnaPredict, a serial EA for the matching principle which encodes RNA secondary structures in permutations and comprises two thermodynamic models based stacking-energy. Those sets of experiments were accomplished on five known structures of 3 RNA classes. The authors claimed the results that P-RnaPredict was exposed to possess good prediction accuracy, specifically on shorter sequences and P-RnaPredict acquired in predicting structures with greater true positive base pair counts and lesser false positives than mfold on definite sequences.

Probabilistic RoadmapMethods

Thomas and Amato²³ presented a new computational approach for studying protein folding that was dependent on probabilistic roadmap approaches for motion planning method. Further, the authors claimed that their method yielded an approximate map of a protein's energy landscape that consists thousands of feasible

folding pathways. Other simulation techniques such as Monte Carlo methods or molecular dynamics needed many orders of magnitude more time to generate a single or partial trajectory. They stated their experiments parallelizing their technique using STAPL, that is being established in the Parasol Lab at Texas A&M. Using STAPL, they were capable to easily parallelize their sequential code to attain scalable speed ups.

RESULTS AND DISCUSSION

It is obvious from the previous section that many researchers used evolutionary methods to resolve protein fold recognition problem. Further, the results of the survey are given in two separate tables presented below. Table 1 depicts the technique based protein fold recognition approaches and the Table 2 shows the protein fold recognition approaches based on parallel evolutionary methods.

CONCLUSION

The paper comprises the survey results in the developing area of protein folding that proffers many computational and mathematical problems. Further, the analysis have been done regarding various methods and techniques such as evolutionary algorithms, genetic algorithms, Support Vector Machine (SVM), Neural networks, Bayesian Network, etc., It is obvious from the research that a vital element for a structure survey is a library of protein folds that aligns all the known or defined structures into fold-families. From the results, the authors claimed that the protein fold recognition for long pattern protein sequences is a great confrontation for many years. However, the computational complexity can be solved by effective parallelization of evolutionary methods, which can also afford better performance in protein folding. As a future work, protein folding using some extended genetic algorithms along with evolutionary conceits is of great and valuable interest.

REFERENCES

1. Akutsu, T. and S. Miyano, "On the approximation of protein threading," *Theoretical Computer Science*, **210**(2): pp. 261-275 (1999).
2. Anbarasu, L.A., P. Narayanasamy and V. Sundararajan, "Multiple molecular sequence alignment by island parallel genetic algorithm," *Curr. Sci.*, **78**: 858-863 (2000).
3. Carpio, C.A.D., Sasaki, S.I., L. Baranyi and H. Okada, "A parallel hybrid GA for peptide 3D structure prediction," Proceedings of the Workshop on Genome Informatics (1995).
4. Dandekar, T. and P. Argos, "Potential of genetic algorithms in protein folding and protein engineering simulations," *Protein Eng. Des. Select.*, **5**: 637-645 (1992).
5. Day, R.O., G.B. Lamont and R. Pachter, "Protein structure prediction by applying an evolutionary algorithm," Proceedings of the International Symposium on Parallel and Distributed Processing, 155-162 (2003).
6. Sangjo Han, Byung-chul Lee, Seung Taek Yu, Chan-seok Jeong, Soyoung Lee and Dongsup Kim, "Fold recognition by combining profileprofile alignment and support vector machine," *Journal of Bioinformatics*, **21**: pp. 2667-2673 (2005).
7. Fischer, D., L. Rychlewski, R.L. Dunbrack, A.R. Ortiz and A. Elofsson, "CAFASP3: The third critical assessment of fully automated structure prediction methods," *Journal of Proteins*, **53**: pp. 503-516 (2003).
8. Islam, R. and A. Ngom, "Parallel evolution strategy for protein threading," Proceedings of the 25th International Conference on Chilean Computer Science Society, pp. 74 (2005).
9. Moulton, J.J., F. Fidelis, A. Zemla and T. Hubbard, "Moulton, J.J., F. Fidelis, A. Zemla and T. Hubbard, Critical assessment of methods on protein structure prediction

- (CASP)-round," *Journal of Proteins*, **53**: pp. 334-339 (2003).
10. Jones, D.T., "GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences," *Journal of Molecular Biology*, **287**: pp. 797-815 (1999).
 11. Lin, K., A.C.W. May and W.R. Taylor, "Threading Using Neural NETWORK (TUNE): The measure of protein sequence-structure compatibility," *Journal of Bioinformatics*, **18**: pp. 1350-1357 (2002).
 12. Krasnogor, N., B.P. Blackburne, E.K. Burke and J.D. Hirst, "Multimeme algorithms for protein structure prediction," Proceedings of the International Conference on Parallel Problem Solving from Nature, pp. 769-778 (2002).
 13. Lathrop, R., "The protein threading problem with sequence amino acid interaction preferences is NP-complete," *Protein Eng. Des. Select.* **7**:pp. 1059-1068 (1994).
 14. Liang, F. and W.H. Wong, "Evolutionary monte carlo for protein folding simulations," *Journal of Chemi. Phys.*, **115**: 3374-3380 (2001).
 15. Mcguffin, L.J. and D.T. Jones, "Improvement of the GenTHREADER method for genomic fold recognition," *Journal of Bioinformatics*, **19**: pp. 874-881 (2003).
 16. Judy, M.V. and K.S. Ravichandran, "A solution to protein folding problem using a genetic algorithm with modified keep best reproduction strategy," Proceeding of IEEE Congress on Evolution Computation, pp. 4776-4780 (2007).
 17. Jiang, N., W.X. Wu and I. Mitchell, "Protein fold recognition using neural networks and support vector machines," Proceeding of the 6th International Conference on Intelligent Data Engineering and Automated Learning-IDEAL, pp. 462.469 (2005).
 18. Alone, N., "Parallel evolution strategy on grids for the protein threading problem," *Journal of Parallel Distributed Computing*, **66**: 1489-1502 (2006).
 19. Nguyen, D.H., Yoshihara, I. Yamamori, K. and Yasunaga, M., "Aligning multiple protein sequences by parallel hybrid genetic algorithm," *Genome Inform.*, **13**: 123-132 (2002).
 20. Raval, A., Z. Ghahramani and D.L. Wild, "A Bayesian network model for protein fold and remote homologue recognition," *Journal of Bioinformatics*, **18**: 788-801 (2002).
 21. Schulze-Kremer, S. and U. Tiedemann, "Parameterising genetic algorithms for protein folding simulation," Presented at Colloquium on Molecular Bioinformatics, 1994, Institute of Electrical Engineers, IEE Press, London, UK.
 22. Taylor, W.R. and I. Jonassen, "A structural pattern-based method for protein fold recognition," *Proteins Struct. Funct. Bioinformat.*, **56**: 222-234 (2004).
 23. Thomas, S. and N.M. Amato, "Parallel protein folding with STAPL," Proceedings of the 18th International Parallel and Distributed Processing Symposium, pp. 189 (2004).
 24. Unger, R., "The genetic algorithm approach to protein structure prediction.," *Struct. Bond.*, **110**: pp. 153-175 (2004).
 25. Unger, R. and J. Moul, "Genetic algorithms for protein folding simulationsn," *Journal of Molecular Biology*, **231**: pp. 75-81 (1993).
 26. Wiese, K.C and A. Hendriks, "A detailed analysis of parallel speedup in P-RnaPredict-an evolutionary algorithm for RNA secondary structure prediction," Proceeding of the IEEE Congress on Evolutionary Computation, pp: 2323-2330 (2006).
 27. Xu, J., "Fold recognition by predicted alignment accuracy," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2**: pp. 157-165 (2005).
 28. Yadgari, J., A. Amir and R. Uunger, "Genetic threading. Constraints," **6**: 271-292 (2001).
 29. Yanev, N. and R. Andonoy, "Solving the protein threading problem in parallel," Proceedings of the 17th International Symposium on Parallel and Distributed

- Processing, pp. 157-164 (2003).
30. Yudong Zhang, Lenan Wu, Yuankai Huo and Shuihua Wang, "Chaotic Clonal Genetic Algorithm for Protein folding model," In the Proceedings of International Conference on Computer Application and System Modeling, 3: 120-124 (2010).
31. Yudong Zhang, LenanWu, "Artificial Bee Colony for Two Dimensional Protein Folding," *Advances in Electrical Engineering Systems*, 1(1): pp. 19-23.
32. R. F. Mansour, "Applying an Evolutionary Algorithm for Protein Structure Prediction," *American Journal of Bioinformatics Research*, 1(1): pp. 18-23 (2011).