Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.983 – 993

RESEARCH ARTICLE

Ab-initio Algorithms for 3D-Protein Structure Prediction

Mohan Kumar[1], Satish Pandey[2], Dr. K. L. Jaiswal[3], Pramod Yadav[4]

[1]Birla Corporation Limited, Satna, MP. mohankumar@satnacement.com

[2]Ph.D. research Scholar, MGCGVV, Satna, M.P., India, satishpandey189@reddiffmail.com

[3]Assistant Professor, Department of Physics, Govt. P.G. Science College-Rewa, M.P., India,drkanhaiyalaljaiswal@gmail.com

[4]Ph.D. research Scholar, MGCGVV, Satna, M.P., India, yadav.pramod181@gmail.com

Abstract: ***The protein folding problem remains one of the most challenging open problems in computational biology. Simplified models in terms of lattice structure and energy function have been proposed to ease the computational hardness of this optimization problem. There are three computational methods of protein structure prediction. These are namely homology modeling, threading and ab initio. Homology modeling and threading methods are knowledge based and requires libraries of experimentally predetermined structures, so a new reliable approach of protein structure from amino acid sequences is required. In this paper, comparative study of different methods of ab initio approach of protein structure prediction along with their limitations has been discussed. Special emphasis has been put on constraint based approach, which is potential approach for suggesting more efficient algorithm for 3D protein structure prediction.***

Key Words: Optimization problem, threading, homology modeling, HP Model, native state, conformational energy function, optimization

1. Introduction:

Protein folds spontaneously and instantly into native shapes. There are many computational methods [1] to predict Protein folding Pathways and Structure of Proteins from Homology modeling and threading [2]. These methods take longer times and do not give exact structures. Although much progress has been made in energy-based methods, Successful prediction of protein structure solely from the potential energy function still remains as a challenging problem. For this reason, most of recent new fold prediction methods use information on known structures to some degree. So, there is requirement of ab initio methods to predict protein structure solely based on solvent conditions and amino acid sequences.

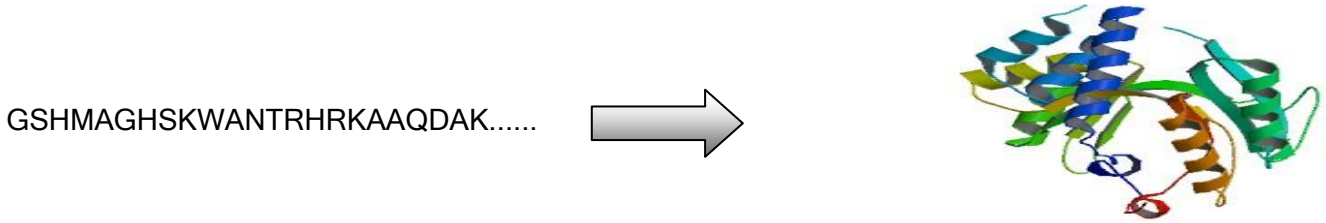


Figure 1. Sequence of amino acids of any proteins folds into unique native state conformation.

Experimental Protein structure determination procedures are time consuming and expensive. Earlier, many researchers have tried to understand processes of protein folding[3][4]. Earlier researchers developed HP model of protein folding. This model is thermodynamic based. Constraints based approaches for protein folding is remarkably more successful[5][6].

Anfinsen's dogma: Anfinsen's dogma (also known as the **thermodynamic hypothesis**) is a postulate in molecular biology championed by the Nobel Prize Laureate Christian B. Anfinsen. The dogma states that, at least for small globular proteins, the native structure is determined only by the protein's amino acid sequence. This amounts to say that, at the environmental conditions (temperature, solvent concentration and composition, etc.) at which folding occurs; the native structure is a unique, stable and kinetically accessible minimum of the free energy.

What is Protein: Protein is common term used to describe any sequence of polymerized amino acids. There are 20 Amino Acids. For a protein to be in functional state at least there should be a collection of at least 40 amino acids bonded together in polypeptide backbone bond. Examples of protein include Blood, Enzymes, Hormones, and Antibodies. DNA is creating RNAs which is transformed into proteins. Each gene in Genome is responsible for production of one particular protein. Local interactions, like hydrogen bonding, gives shape to secondary structure elements while hydrophobicity basically controls over all shape of protein structure. It is to be noted that only some valid sequences of amino acids are stable proteins.

Function of Proteins:
Antibodies: are specialized proteins involved in defending the body from antigens.
Contractile Proteins: are involved in muscle contraction and movement.
Enzymes: are proteins that facilitate biochemical reactions. They are often referred to as catalysts because they speed up chemical reactions.
Hormonal Proteins: are messenger proteins which help to coordinate certain bodily activities.
Structural Proteins: are fibrous and stringy and provide support. Examples include keratin, collagen, and elastin. Keratins strengthen protective coverings such as hair, quills, feathers, horns, and beaks.
Storage Proteins: store amino acids. Examples include ovalbumin and casein. Ovalbumin is found in egg whites and casein is a milk-based protein.
Transport Proteins: are carrier proteins which move molecules from one place to another around the body. Examples include hemoglobin and cytochromes.

How is protein produced: Protein is formed from transcription of DNA. DNA is first converted into mRNA which changes into protein. DNA is almost 2 meter long and consists of many segments(genes) which encode proteins. A single DNA is capable of producing many types of proteins. Protein synthesis information is encoded in DNA. A single DNA carries all characteristics of an organism.

Amino Acid sequences, structure and Cell Function: Primary Protein sequence comprising of chain of amino acids residues folds into tertiary structure carrying out several biological functions in a cell.

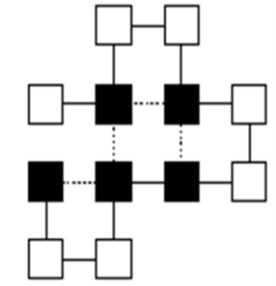
Types of Protein Structures:

1. Primary Structure
2. Secondary structure
3. Super secondary structures
4. Tertiary structure
5. Quaternary structure

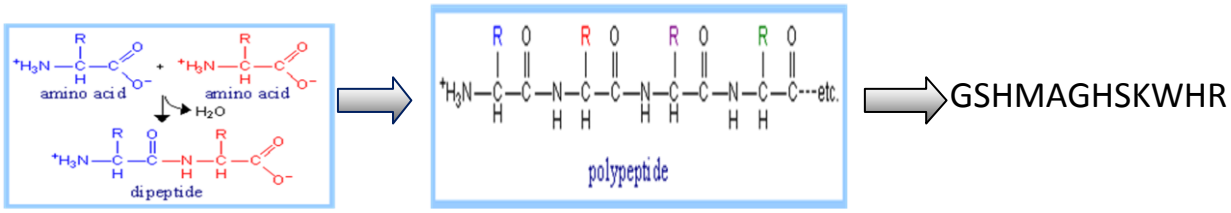
HP Model of Protein Folding: In this model amino acids, which are components of proteins, have been categorized into two classes.

1. **Hydrophobic:** Hydrophobic amino acids are non polar. These atoms are placed near center of Protein structure. These forces determine overall structure of proteins

2. **Hydrophilic**: These amino acids are polar in nature. Folding programs are designed to keep these residues away from core of structure. In all HP model algorithms, common agenda is to minimize total energy of structure. Every non continuous HH contacts are awarded energy contribution of -1



2D -Square Lattice for HP Model.



The primary structure or amino acid sequence is unique for each protein. The sequence in which the amino acids are attached to one another ultimately is dependent on the genetic code from DNA. This primary structure dictates the function of the protein indirectly through additional levels of structure.

2. Different approaches for protein structure determination: 3D Structure of Protein can be experimentally determined by Nuclear Magnetic Resonance (NMR) and X-Ray Crystallographic Diffraction Methods. These Methods are expensive and takes a lot of time. Researchers have also developed computational Methods. This is categorized into three groups namely Ab initio, Homology modeling and Threading methods.

Different Folding Algorithms for Protein folding by *ab*-initio Methods:

1. Monte Carlo Method for Protein Folding.
2. Genetic Algorithm by Unger and Moult[12].
3. Ant Colony System Approach for Protein Folding[13].
4. Prune Enriched RosenBluth Method (PERM)[14].
5. Hydrophobic Zipper Method[15].
6. Memetic Algorithm for Protein Folding[16].
7. Filter and Fan Method of HP Protein Folding[17].
8. Evolutionary Monte Carlo for protein folding simulations[18].
9. PFOLD-P-ACO Protein Folding Algorithm[19].
10. A replica exchange Monte Carlo algorithm for protein folding in the HP model[20].
11. Branch and Bound Algorithm for Protein Folding[21].
12. Heuristic approaches e.g. Rosetta[22].
13. Constraint based approach by Rolf Backofen[23].

2.1 Monte Carlo Method for Protein Folding: It consists of following steps.

(1) Start from a random coil conformation.

(2) Format a conformation S1 with energy E1, make a single random change of the conformation to conformation S2 and evaluate its energy E2.

(3) If E2 <= E1, then accept the change to conformation S2, otherwise decide, nodeterministcally, whether to accept the change according to the energy increase with the change. Usually the criteria is of the form: accept if :

Rnd<exp { (E1-E2) /Ck},

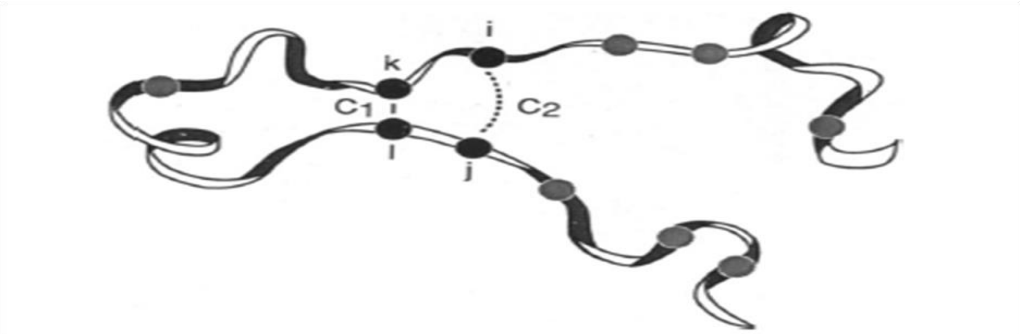
Where Rnd is a random number between 0 and 1 and Ck is gradually decreased (cooled) during the simulation to archive convergence. If the change was not accepted, then retain the former conformation S1.

2.2 Genetic Algorithm by Ron Unger and John Moult[12]: This algorithm is based on biologically inspired principles of genetics. In it, a population of some best conformations is maintained. Some of these conformations are applied mutation and crossover operations. If energy of new conformation is lower than previous ones then population is updated with this new one. At some steps mutation is also applied in this process. After finite number of steps this loop terminates and minimum energy conformation is selected as best conformation. It is to be noted that initial conformations are complete conformations.

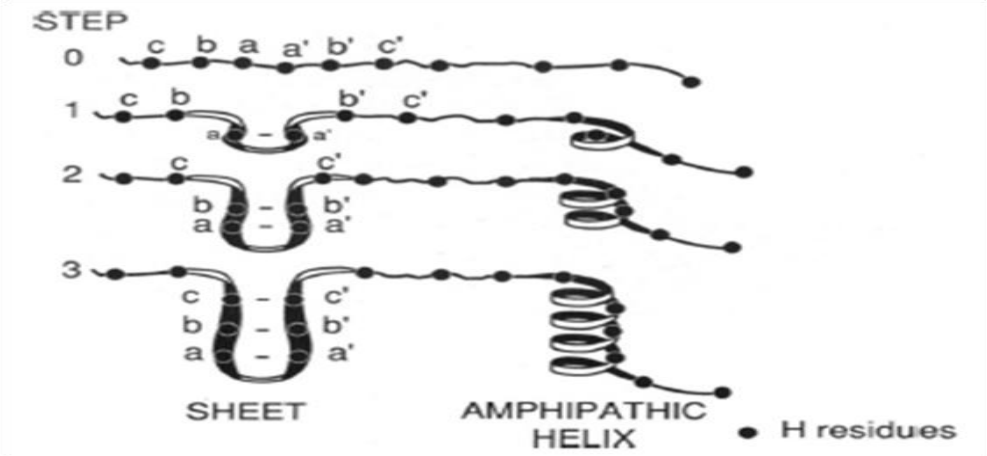
2.3 Ant Colony System Approach for Protein Folding[13]: The ACO algorithm uses a colony of artificial ants that behave as co-operative agents in a mathematical space where they are allowed to search and reinforce path ways (solutions) in order to find the optimal ones. The problem is represented by graph and the ants walk on the graph to construct solutions. After initialization of the pheromone trails, ants construct feasible solutions and the pheromone trails are updated. The higher the value of the pheromone and the heuristic information, the more profitable is to select this move and resume the search.

2.4 Prune Enriched Rosen Bluth Method[14]: PERM is a biased chain growth algorithm with resampling ("population control") and depth-first implementation. The chain grows by adding monomers, PERM suppresses these fluctuations by ''pruning'' configurations with too low weight and by ''enriching'' the sample with copies of high-weight configurations. PERM can be viewed as a special realization of a ''go-with the-winners'' strategy. It is found to outperform all other fully blind general purpose stochastic algorithms (Genetic Algorithms) which have been employed on this problem

2.5 Hydrophobic Zipper Method[15]: In this Method an initial process for folding starts at some point. Folding process is similar to zipping of zipper (chain in gent's paint for locking). This initiation is added by additional binding of monomers to form either α-helix or β-sheet. This is totally local effect. This local cooperativity produces α -helix and β-strands and β-sheets.



Origin of cooperativity: In above figure, Given C1(k, l), monomer I explores mainly the possibility of pairing with j, not with all other H monomers; thus, this search is nonexhaustive. It is cooperative because the probability of forming contact C2= (i, j) is much higher if C1 is formed.



In HZ model protein folding there are some fold initiating points. These points grow into definite helix and sheets. E.g. in the above figure, the closest hydrophobic (H) residues (solid dots) in sequence Pair together first. E.g. a and a' in step 0. They constrain the chain and bring other H monomers, such as the (b, b') pair into spatial proximity. Now (b, b') further constrain the chain and brings the (c, c') pair into spatial proximity, etc. as H contacts form and develop a core of helices and sheet zip up provided they have appropriate H sequences.

Strong Points: Helix-coil cooperativity is driven by interactions that are local in the sequence and leads only to helical structures, relatively independently of the amino acid sequence. HZ cooperativity for heteropolymer collapse is driven by nonlocal (but T-local) interactions mediated by the solvent. It leads to helical, sheet, and irregular structures, depending on the amino acid sequence. The folding process is postulated to be a fast nonexhaustive search, in which, for appropriate sequences, the early stages of H core formation are concurrent with formation of helices and sheets. The folding pathways involve formation of H-H contacts that are brought into close spatial proximity by virtue of preceding contact constraints.

Weak Points: HZ folding pathways lead to the conformations of global minimum in free energy for many but not all HP sequences. It is based on oversimplified HP model. It can be extended for exact structure prediction. β-Sheet formation can be well explained by cooperativity. α-Helix formation cannot be well explained by origin of cooperativity.

2.6 Memetic Algorithm for Protein Folding[16]: In this Protein Folding of HP Model, different self adaptive local search operators have been applied to optimize energy of 3D protein structure. These are namely following:

1. End Move.
2. Corner Move.
3. End Corner Move.
4. Crank Shaft Move.

Authors have also developed speciation operator to speed up minimum energy native shape of protein.

2.7 Filter and Fan Approach for Protein Folding [17]: The filter and fan (F&F) method was initially proposed by Glover (1998) as a method for refining solutions obtained by scatter search, and was further extended by Rego and Glover (2002). Graphically, the F&F model can be illustrated by means of a neighborhood tree where branches represent submoves and nodes identify solutions produced by these moves. The neighborhood tree is explored breadth-first, level by level. Each level is governed by the filter candidate list strategy that selects an optimum and a filter and fan search to explore subset of moves induced by the fan candidate list strategy.

The implementation of a pull move neighborhood requires three basic functions associated respectively with: (a) detecting all valid pull moves available, (b) executing the pull move, and (c) computing the energy value of the resulting conformation.

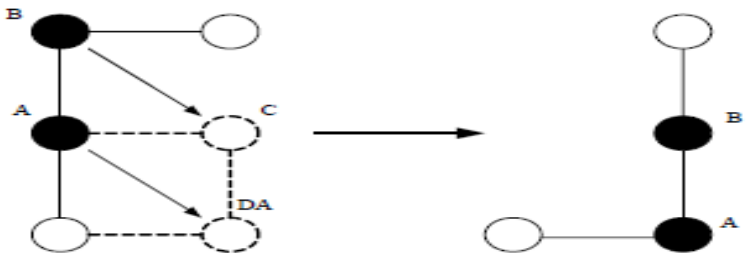


Figure: Move set for filter and Fan Method.

2.8 Evolutionary Monte Carlo for protein folding simulations [18]: It works on full developed conformations. It is not like chain growth algorithms. It combines good features of genetic algorithm and montecarlo sampling techniques. This method employs mutation, exchange, and crossover operators to speed up for search for optimal energy native state conformations. It has also focused on identifying the subsequences which will possibly fold to secondary structures in the native state of a given protein. Common steps in this method include following steps.

1. Make random move and produce a new conformation.
2. Calculate the energy change ΔE for the new conformation.
3. Accept or reject the move based on the Metropolis criterion.

$$P = \exp(-\frac{\Delta E}{kT}) \longrightarrow \text{Boltzmann factor}$$

If $\Delta E<0$, P>1, accept new conformation; Otherwise: P>rand(0,1), accept, else reject.

To apply EMC to the 2D HP model, each conformation of a protein is represented by a vector. The energy function of the conformation is denoted by H(z) and the Boltzmann distribution is defined by

$$f(z) \alpha \exp\{ -H(z)/t \}$$

Where t is called the temperature of the system. In EMC, to simulate from the target distribution f (z), a sequence of distributions $f_1 ... f_n$ are first constructed.

2.9 PFOLD-P-ACO Protein Folding Algorithm [19]: In this approach of HP protein folding, Authors have used Hybrid and heuristic approach to solve protein folding problem in shortest possible Time. Protein folding problem is NP hard, so heuristic and optimization techniques are required to find optimum solution. Four constraints are used in this heuristic method:

1. Ant Colony optimization algorithm is applied (ACO).
2. Population based i.e. 10 conformations are reserved.
3. Growing chain Length of protein is kept below half of full length.
4. P-rich subsequences are kept near to each other.

2.10 A replica exchange Monte Carlo algorithm for protein folding in the HP model [20]: Replica Exchange Monte Carlo (REMC) search maintains χ independent replicas of a potential solution. Each of the χ replicas has an associated temperature value ($T_1, T_2 ... T_\chi$). Each temperature value is unique and the replicas are numbered such that $T_1 < T_2 < ... < T_\chi$. X conformations are maintained by the algorithm with the replica numbers ($1 . . . \chi$,) and always associate temperature T_j with replica j (for all j such that $1 \leq j \leq \chi$). Thus, the exchange of replicas is equivalent to (and is commonly implemented as) the swap of replica labels. All other steps in this algorithm are like genetic algorithms.

2.11 Branch and Bound Algorithm for Protein Folding [21]: In this algorithm, the benefit of each possible location of hydrophobic monomers is evaluated and only promising nodes are kept for further branching at each level. This algorithm is chain growth based i.e. monomers are added one by one and energy of partial conformation is evaluated. If there is overall improvement in energy of partial conformation, then that movement in that direction is considered good. Here size of tree grows exponentially, as monomers are added to partial conformations.

2.12 Heuristic Approaches: Macro Molecular modeling with Rosetta [22]: Molecular modeling problems are treated in Rosetta by following the same steps of defining an atom tree, initial conditions, accessible rotamers, and a schedule of Monte Carlo moves affecting the flexible degrees of freedom. Modeling the binding of small molecules to proteins is very similar to the problem of protein-protein docking. *Modeling of membrane* proteins follows a protocol similar to the modeling of soluble proteins, albeit with different low-resolution and high-resolution energy functions.
Weakness of these Methods: This is knowledge based Methods. It requires collection/library of huge collection of protein structure fragments. Natural Protein folding is very simple and spontaneous. Natural Protein folding occurs instantaneously as soon as protein is generated by DNA in biological cell. Prediction results of Rosetta are accurate but folding process in Rosetta is time consuming and extensively computational.

2.13 constraint based approach by Rolf Backofen [23]: This method is based on applying different folding rules as constraints to constraint solver. Constraint solver is designed such that additional constraint can be added without

modifying existing program scheme. In this method following are Structural Constraints for protein tertiary structure prediction:

i) Domain bounds: Bounds all the variables X_i, Y_i, Z_i to 0 . . . 2N. Moreover, exploiting some lattice properties we also force $X_i + Y_i + Z_i$ to be even.

ii) Avoid Self Loops: Avoid self loops implements the non-overlapping property, i.e., forces all points to be distinct. This is realized converting each coordinate triple into an unique number and calling the built-in predicate all different on the list $[B_1, .., B_n]$ where Bi #= (Xi ∗ P ∗ P)+(Yi ∗ P)+Zi for a suitable value P).

iii) Distance Constraints: Distance constraints implement the side chain occupancy property and thus defines that two non-consecutive points are at a lattice distance greater than 1. This is achieved by adding, for each pair of non-consecutive points i and j, the equivalent constraint:

- Dx #= (X1-X2)*(X1-X2),
- Dy #= (Y1-Y2)*(Y1-Y2),
- Dz #= (Z1-Z2)*(Z1-Z2),
- Dx+Dy+Dz #>2.

iv) Avoid Symmetries: Avoid symmetries remove admissible conformations that are equivalent to others modulo symmetries and/or rotations. The predicate selects three particular consecutive amino acids. We set the position of the first of them removing all folding equivalent modulo translations. We fix also the positions of the second and the third point in order to remove all equivalences modulo rotations.

v) Structural Constraints- Angular Values: Angles defines the admissible angles formed by three consecutive amino acids. Recall that the bend angles property imposes that the admissible bend angles between amino acids i, i+1 and i+2 are 90◦ and 120◦ degrees.

vi) Compactness Constraints: Compact constraints introduces the user-defined compact factor (cf), since Euclidean distance has a very poor propagation effect, we implement here a relaxation. In fact, we impose that each coordinate of the pair of amino acids is in distance at most cf. Instead of using a sphere of radius cf, we impose a box of side 2cf to limit the distance of the amino acids.

vii) Secondary Information: Secondary information encodes the Secondary Structure information as constraints in the program. The Secondary structure is described by a list of elements of the type:

- Helix (i, j): s_i, s_{i+1}. . . s_j form an α-helix.
- Strand(i, j): s_i, s_{i+1}, . . . , s_j are in a β-strand.
- Ssbond(i, j): presence of a disulfide bridge between si and sj .

viii) Lattice Model: This algorithm generally uses cubic lattice and Face Centered Cubic (FCC) lattice for placement of monomers of the given protein. For FCC lattice Xi+Yi+Zi is even number.

Strong Points of Above Algorithm: It is based on all constraints and rules available for protein folding. Additional constraints can be added to Constraint solver without affecting over all program architecture. Improvements in these methods may give highly accurate results.

3. How Much Successful are Current Folding Algorithms?

1. Folding algorithm for protein structure Prediction based on HP model are important for study of processes of protein folding pathways. These methods ca not give exact structure as it is found in PDB. For accurate structure prediction macromolecular modeling will be more suitable. Macro molecular modeling based on full atomic model of atoms is very complicated. It is why researchers are focusing on oversimplified HP models.

2. Folding algorithm does not take into account other cellular environmental and protein and protein interaction effects in folding. At present in protein folding, secondary structures of protein does not appear in folding process.

3. Structure Prediction is very far from real structures of Proteins even in simple proteins.

4. In earlier days of protein structure prediction, researchers focused on finding native state in shortest possible time. They developed various optimization techniques like annealing, self adaptive local search methods, Monte Carlo simulations and genetic algorithms.

5. At present most of the methods are based on fragment assembly. Some of prediction methods are based on homology modeling. There is need of prediction of structures from purely amino acids sequences only. So that, folding process may be well formulated. It will save a lot of database space and computational time.

4. Advantages of Protein Folding Study:

1. Knowledge of folding pathways and native structure.
2. Design of new proteins with pre-specified characteristics.
3. Controlling protein folding process to prevent undesirable Consequences.
4. Understanding of different biological cell processes.
5. Design of new drugs based on active site of proteins.

5. Comparative Study of Major 3D Protein Structure Prediction Servers/ Authors:

Authors/Server	Main Features	Strong Points	Weak Points	Algorithm
Faming Lianga, Wing Hung Wongb[20]	Evolutionary Monte Carlo for protein folding simulations	Uses secondary structure elements for faster prediction	Used only for 2D protein folding	Approximation
Andrea Bazzoli, Andrea G.B. Tettamanzi[18]	A Memetic Algorithm for Protein Structure	Uses Memetic Operators for optimization	Prediction in a 2D-Lattice HP Model only	Aproximation
Ken A. Dill , Klaus M. Fiebig , and Hue Sun Chan[17]	Cooperativity in protein-folding kinetics	strong connectivity between folding and cooperativity	Not very realistic, Takes longer time to predict 2D structures	Non Exhaustive search
Ron Unger and John Moult [12]	Genetic Algorithm for Protein folding. This algorithm is many times faster than Monte Carlo.	Helpful in understanding protein folding	It does not give exact structure.	approximation

Rolf Backofen & Sebastian Will[23]	Constraint-Based Approach to Fast and Exact Structure Prediction in	Three-Dimensional Protein Models	Needs Computation of hydrophobic cores before prediction	Exact Method
Shuai Cheng Li, In a thesis presented to the University of Waterloo in 2009[11]	Structural fragment library have been developed and maintain to best predict 3D structures	Based on comparative and homology modeling	Needs collection of structural fragment library	Heuristic approach and lengthy procedure

Tables of prominent contributors for Protein Structure Prediction.

6. Suggestions for Different new approaches that may improve accuracy in protein structure prediction:

1. Object oriented and Event driven programming language paradigm to contract all atoms towards core of protein. Original covalent bonds in amino acids will remain unaffected but topological contacts may influence over all shape of protein.
2. Lattice free model to predict structure prediction.
3. Choosing another possibility that may help in realistic protein structure prediction.
4. Full atom model based on solvent and electrostatic interactions may give exact and fast structure.
5. Heuristic based approach to combine all available facts about protein folding process.

7. Targets to be achieved by new Modeling:

1. More realistic and accurate 3D- protein structure predictions.
2. Proposed algorithm will predict 3D structures of all proteins. It will predict structures of globular, fibrous as well as non crystalline water soluble proteins.
3. It will include cellular environmental conditions as well as effect of protein-protein interactions.
4. Proposed algorithm will be fast and as simple as spontaneous folding of proteins takes place in nature.

8. Conclusion:

There are a lot of researchers working for determination of protein structure. But none of the methods are predicting exact structures. Present knowledge of proteins is like tip of iceberg. There needs a lot of further research work to improve prediction methods of protein structure. As we know more and more about cellular biology of protein folding, we will be able to predict relatively fast and exact structure of proteins. This knowledge will be very helpful for simulating internal processes occurring in biological cells.

References

[1] Rabindra Ku. Jena, Musbah M. Aqel, Pankaj Shrivastava, Prabhat K. Mahanti (2009). Soft Computing Methodologies in Bioinformatics, European J. of Scientific Research Vol.26 No.2 (2009) pp.189-203,

[2] Mingfu Shao, Sheng Wang, Chao Wang, Xiongying Yuan, Shuai Cheng Li, Weimou Zheng, DongoBu , Incorporating Ab Initio energy into threading approaches for protein structure prediction, From The Ninth Asia Pacific Bioinformatics Conference (APBC 2011) Inchon, Korea. 11-14 January 2011.

[3] Yue K and Dill KA, Forces of tertiary structural organization in globular proteins. Proc of Natural Academy of Sciences USA 1995, 92:146–150.

[4] Sebastian Will, Constraint based Hydrophobic core construction for protein structure prediction in the face centered cubic lattice , pacific symposium on biocomputing 7:661-672 (2002).

[5] Alessandro Dal Palù, Agostino Dovier and Federico Fogolari,2004,Constraint Logic Programming approach to protein structure prediction, BMC Bioinformatics

[6] Dal Palù A, Dovier A and Pontelli E: Heuristics, Optimizations and parallelism for protein structure prediction in CLP(FD). International Conference on Principles and Practice of Declarative Programming NY: ACM Press; 2005, 230–241.

[7] Alfonso Valencia, Multiple sequence alignments as tools for protein structure and function prediction, Protein Design Group, National Centre for Biotechnology, CNB-CSIC, Madrid, Spain.

[8] Deb Bardhan, Jingjing Hu, Vinay Nadimpally, Ganesh Ramesh, Chris Bystroff, Mohammed J. Zaki**,**2004**,** Predicting Protein Folding Pathway, Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY.

[9] Alena Shmygelska, Holger H Hoos, 2005,An ant colony optimization algorithm for the 2D and 3D hydrophobic polar protein folding problem, BMC Bioinformatics.

[10] Abu Dayem Ullah and Kathleen Steinhofel,2010,A hybrid approach to protein folding problem integrating constraint programming with local search, BMC Bioinformatics, The Eighth Asia Pacific Bioinformatics Conference Bangalore, India 18-21 January 2010.

[11] Shuai Cheng Li, A New Approach to Protein structure prediction, Thesis presented to the University of Waterloo, Ontario, Canada, 2009.

[12] Ron Unger and John Moult, Genetic Algorithms for Protein folding simulations, 1993, J. Mol. Biol. 231, 75-81.

[13] A Stefka Fidanova, Ivan Lirkov, 2004,Ant Colony System Approach for Protein Folding, Institute for Parallel Processing, Bulgarian Academy of Sciences, Acad G. Bonchev, bl. 25A, 1113 Sofia, Bulgaria.

[14] Hsiao-Ping Hsu, Vishal Mehra, Walter Nadler, and Peter Grassberger, Growth-based optimization algorithm for lattice heteropolymers, 2003, John-von-Neumann Institute for Computing, Forschungszentrum Julich, D-52425 Julich, Germany.

[15]Ken A. Dill, Klaus M. Fiebig, and Hue Sun Chan, Cooperativity in protein-folding kinetics, 1992, Proc. Natl. Acad. Sci. USA Vol. 90, pp. 1942-1946, Biophysics.

[16]Andrea Bazzoli, Andrea G.B. Tettamanzi, 2004,A Memetic Algorithm for Protein Structure Prediction in a 3D-Lattice HP Model ,EvoWorkshops 2004, LNCS 3005, pp. 1–10.

[17]César Rego, Haitao Li, Fred Glover ,A Filter-and-Fan Approach to the 2D HP Model of the Protein Folding Problem,2006, School of Business Administration, University of Mississippi, University, MS 38677, USA.

[18]Faming Lianga, Wing Hung Wongb, 2001, Evolutionary Monte Carlo for protein folding simulations, Journal of Chemical Physics, Volume 115, Number 7.

[19]Torsten Thalheim, Daniel Merkle, and Martin Middendorf, Protein Folding in the HP-Model Solved With a Hybrid Population Based ACO Algorithm. 2008, IJCS 35:3.

[20]Chris Thachuk, Alena Shmygelska, Holger H Hoos, 2007,A replica exchange Monte Carlo algorithm for protein folding in the HP model, BMC Bioinformatics,

[21]Mao Chen* and Wen-Qi Huang, A Branch and Bound Algorithm for the Protein Folding Problem in the HP Lattice Model, Geno. Prot. Bioinfo. Vol. 3 No. 4 2005.

[22]Julian Lee, Seung-Yeon Kim, Jooyoung Lee, Protein structure prediction based on fragment assembly, Biophysical Chemistry,115 (2005) 209– 214.

[23]Rolf Backofen & Sebastian Will, 2006, A Constraint-Based Approach to Fast and Exact Structure Prediction in Three-Dimensional Protein Models, Springer Science.

