



Protein tertiary structure prediction by integrating ant colony optimization with path relinking and structure knowledge

Sivagnanam Rajamanickam Mani Sekhar¹ · Siddesh Gaddadevara Matt² ·
Vasantha Kumara Mahadevachar³

Received: 2 June 2022 / Accepted: 21 February 2023 / Published online: 7 March 2023

© The Author(s), under exclusive licence to Bharati Vidyapeeth's Institute of Computer Applications and Management 2023

Abstract The study of protein and its structure plays an essential role in computational bioinformatics research. Various computational methods were proposed for protein tertiary structure prediction, but still, it is a challenging task as the number of amino acids known is more than the predicted tertiary structure. In this work, the authors integrated Metaheuristic Ant Colony Optimization (MACO) with Path Relinking (PR) strategy and structure knowledge (SK). As MACO can generate local minima of torsion angle, the author has integrated PR with MACO. To validate the proposed model author tested them on six different protein sequences and later used the basic Ant Colony Optimization (ACO) algorithm for computation. Subsequently predicted tertiary structure is authenticated using Root Mean Square Deviation (RMSD). The result shows that out of six protein sequences proposed model MACO-PR outperform for 2P5K, 1AIL, 1L2Y, 2MTW & 1WQC protein ID when compared with Basic ACO & RMSD values.

Keywords Protein tertiary structure · Metaheuristic ant colony optimization · Path relinking strategy · Machine learning · Bioinformatics.

Abbreviations

MACO	Metaheuristic ant colony optimization
PR	Path relinking
SK	Structure knowledge
ACO	Ant colony optimization
RMSD	Root mean square deviation
CASP	Critical assessment of protein structure prediction
PDB	Protein data bank
CREF	Central-residue-fragment based technique
ABC	Artificial bee colony
PIO	Pigeon inspired optimization
TS	Tabu search
SVM	Support vector machine
NN	Neural network

1 Introduction

Proteins show an essential role in the biological processes of all living organisms. They are the essential building blocks of all living organisms and play a significant role in cell development. The constancy of proteins depends on the analysis factor of protein structure prediction. Later this information is used to generate protein secondary and tertiary structures. They play a significant role in oxygen conveyance in blood vessels.

Structural bioinformatics is a study of different types of problems associated with different biological systems like protein secondary structure prediction, tertiary structure prediction, etc., and they are the most significant problem statement in structural bioinformatics [1, 2]. Subsequently, it results in the development of the latest scientific and computational methods.

Equal contribution from SRMS, SGM, SSM authors.

✉ Sivagnanam Rajamanickam Mani Sekhar
manisekhars@gmail.com; manisekhars@msrit.edu

¹ Department of ISE, M S Ramaiah Institute of Technology, Bangalore, India

² Department of CSE(Cyber Security), M S Ramaiah Institute of Technology, Bangalore, India

³ Department of CSE, Government engineering college, Dairy circle, Hassan, India

The identification of protein function depends upon how accurate the tertiary structure fold occurs for a given protein. A misfold tertiary structure will affect drug design, energy drink, and vaccination design [3, 4], etc. The creation of the protein tertiary structure depends upon the chemical and physical properties of the amino acid sequence and its polypeptide. All these properties are integrated within the building blocks of proteins. If certain protein performs incorrect folding, it leads to the protein breakdown and incorrect structure prediction, later called misfolding. These misfolded proteins yield different types of diseases like diabetes, Alzheimer's, etc. [5–7].

Protein Structure Prediction (PSP) is a challenging task in the bioinformatics field as the amount of amino acids known is more than the predicted tertiary structure. Although there are several experimental methods like X-ray and Nuclear magnetic resonance (NMR) used in finding Protein structure and also they can produce a more effective result. These techniques are more expensive and time-consuming as they can take months in recognizing the structure, which results in generating an immense gap between the number of amino acid sequences and identified protein 3-D structures. In order to overcome this gap, computational techniques must be used in protein structure finding [8–11].

In the past several computational approaches were proposed as replacements for the PSP problem [2]. Based on the structural information from the protein data bank, four different types of computational groups can be defined: (a) principle approaches with database knowledge, (b) ab initio or principle approaches without database knowledge, (c) relative modeling approaches, (d) fold recognition approaches. However, these methodologies have limitations. Group (b) can generate the latest protein structures using the latest folds, but even for a small amino acid sequence, the search space is more complex and with a high dimension. Next, Group (c) can easily forecast the protein structure, which is alike to the know amino acid sequence of the given structure. Group (d) protein structure prediction is based on the available fold collection from the protein data bank. Whereas group (a) method performs well when the result is compared with the critical assessment of protein Structure Prediction [12]. Due to the excellent performance of group(a), the author has integrated Group (a) techniques with machine learning (ML) concepts to improve the protein structure method.

The rest of the paper is structured as follows: Section II gives a brief on protein structure representation, energy function, and protein templates. Next section III discusses the related works. Section IV gives an insight into the proposed work. The result and validation of the proposed model are discussed in section V. Finally, section VI briefs the conclusion part.

2 Background

To better understand the principle of bioinformatics, this section illustrates the background of the protein structure, amino acid, energy function, and protein templates.

2.1 Protein's structure and its amino acid residual

The computed tertiary structure does not only depend on the amino acid sequence but also on various other parameters. Solvent, temperature, and many other biological parameters define the relation among amino acid sequences. These features help in identifying the natural protein structure, and more the features quantity high will be the accuracy. The dihedral angle [13] can be used to define the tertiary structure of the given sequence. This method uses bond length, as the length of the peptide chain is almost constant.

Protein is a collection of subsequently linked amino acid residues with associated peptide bonds. Every amino is made up of a carboxyl set, a link to carbon bonds, and a successively linked side chain with detailed physicochemical properties. When a carboxyl set of residue interacts with an amino set of the peptide bond is generated, later a water particle is released. A torsion angles [14] phi (ϕ) and psi (ψ) of amino acid can be used to define polypeptide and protein backbone. These torsion angles of a protein structure can be represented by the Ramachandran plot [15]. Meanwhile, steric hindrance defines the value of rotational round the ϕ and ψ angles. A side chain of polypeptide must have dihedral angles (Chi angles- χ).

2.2 Energy function

To decrease an energy function, the protein structure prediction method must change the orientation of the protein structure atoms [16]. To assess the quality of the predicted structure different energy functions can be used. Here the author has incorporated the Rosetta energy function [17]. The Rosetta scoring function contains more than eighteen energy terms, some are inter-atomic interaction, hydrogen bond, Newtonian physics, knowledge built potential, etc. According to the critical assessment of protein Structure Prediction (CASP), Rosetta energy-function based methods performs better when compared with other energy functions [18].

2.3 Protein templates

A Protein template from a protein data bank (PDB) can be used to decrease the high dimensionality and also the difficulty of the search space formed during the ab-initio process, later this knowledge can be used to design conformations. In the proposed work author has used

Central-Residue-Fragment based technique (CREF) for extraction of protein templet from PDB. The CREF [19, 20] predicts the protein templet based on the phi (ϕ), psi (ψ), and torsion angles of the amino acid chain acquired from the PDB. These PSP techniques can work on small fragments well. The cluster method is incorporated in the identification of templates, next, each cluster is labeled against the conformational state specified by the sections in the Ramachandran graph. Finally, the clustering result is used in the development of conformation using the mapping function.

3 Related work

Li et al. [21] proposed a bio-inspired algorithm called artificial bee colony (ABC) with pigeon-inspired optimization (PIO). Here the computation of protein 3D structure is performed by integrating two hybrid algorithms namely an ABC and PIO. Thereafter Cauchy perturbation is used to improve the local fitness. This experiment is performed on ten short-length protein sequences. Whereas Backtracking Search Optimization (BSO) and Tabu Search (TS) algorithm [22] perform prediction of protein 3D structure using BSO and TS. They used TS to overcome the generated local optimum. The prediction can be further increased by incorporating a path linking strategy. In [23] authors showed that their model ResNet is able to predict twenty-six protein folds out of thirty-two targeted values whereas when they have not incorporated ResNet they are able to predict only eighteen protein folds. The prediction can be further increased by incorporating a path linking strategy.

In [24] author has used Mod- artificial bee colony (ABC) with PDB structure knowledge for the prediction of protein tertiary structure. The predicted solution is further computed under a crossover operation between two target solutions. Later, RMSD, distance, and energy functions are used in the analysis of all eight protein sequences. In [25] the author has used a memetic algorithm with structure knowledge for the prediction of protein 3D structure from amino acid sequences. Further, during computation search space is reduced by including an angle probability list. The accuracy can be further improved by integrating ML techniques with a PR strategy.

Yousef et al. used a hybrid model [26] for tertiary structure prediction, they combined the genetic algorithm with energy function. However, during the crossover process, the computed structure can be different due to steric hindrance, this can be validated by energy minimization. Results show that the hybrid models botched to computing the sidechain torsional angles. The prediction can be further improved by including PyRosetta with structure knowledge.

RMSD plays an essential role in protein structure prediction. The author [27] proposed a method to decrease RMSD

error by incorporating a feed-forward neural network and an adaptive neuro-fuzzy approach. Subsequently, the computed result is used in structure prediction. The prediction result can be further increased by altering the number of linguistic variables.

In this work [28], the authors combined a neural network and particle swarm optimization algorithms for protein 3D structure prediction from amino acid sequences. They extracted protein sequences using three different hydrophobic, composition, and Frequency amino acids. The result shows that system is able to classify only alpha and beta trends, whereas it failed to classify alpha+beta and alpha/beta values. To overcome this, drawback authors suggested building a tree classifier at the initial stage.

4 Materials and methods

This section describes the proposed MACO-PR Algorithm implementation. The steps followed in the proposed approach to predict protein tertiary structure is as follows:

Step 1: Extraction of amino acid sequences from PDB.

The above given steps in shown in Fig. 1. Here amino acid sequences are extracted from PDB [29]. Next, the extracted amino acid and template are supplied as input to the proposed MACO-PR model. Finally, the protein tertiary structure is validated using RMSD.

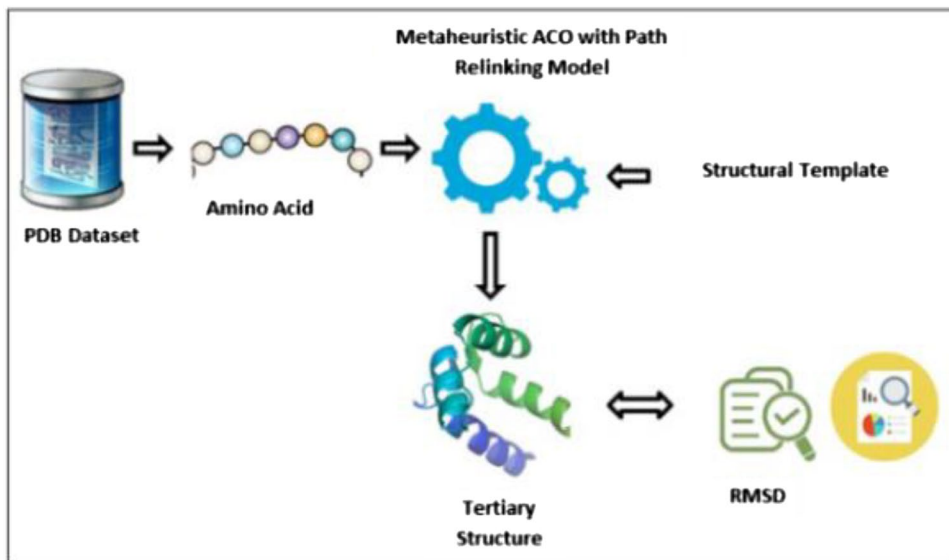
In this work, author developed an enhanced MACO for the protein tertiary structure prediction. Here MACO [30] is integrated with PR [31] strategy and SK. As MACO can generate local minima of torsion angle author has integrated PR with MACO named a MACO-PR. To validate the proposed model MACO-PR author tested them on six different protein sequences and with basic ACO, Support Vector Machine (SVM) [32] and Neural Network (NN) [33]. The ACO mainly goes through three stages: Construction Stage, Local Search Stage, and Pheromone Update Stage.

- Construction stage: Each labor ant starts at a random amino sequence position, and subsequently building a conformation of the given sequence by deploying a candidate solution.
- Local search stage: During this stage further Optimization of conformations folds occurs.
- Pheromone update Stage: based on the conformation energy and search stage each labor ant will start updating the pheromone matrix.

A generic ACO algorithm [30] working steps is illustrated below:

- Step 1 Initialize pheromone trails value to Nil.
- Step 2 Generate candidate confirmation.

Fig. 1 Proposed MACO-PR Architecture for Protein Tertiary Structure Prediction



- Step 3 Initialize labor ant search.
- Step 4 Update pheromone values.
- Step 5 Compute the best path.

During the path construction stage of the proposed model MACO-PR, each labor ant starts at a random position within certain amino acid sequences. Later each residual will get added when the sequence fold occurs in both directions either left or right. Subsequently for each labor ant path generation occur, resulting in protein conformation. Based on the pheromone values and direction, sequence position updating occurs for given position ‘i’. Here the total count of ants will be less than or equal to the number of nodes. Each labor ant will start from a random position [34] in search of the final node(n_i) with the resource (R_j). The computation will be carried out as shown in Eq. (1).

$$PF_{ij} = [(T_{ij})^{\alpha}(H_{ij})^{\beta}] / [\sum (T_{ij})^{\alpha}(H_{ij})^{\beta}] \tag{1}$$

Here, ‘T_{ij}’ denotes the pheromone values corresponding to node (n_i) and resource (R_j). ‘H’ is used to define the given heuristic value. Subsequently, ‘α’ value and ‘β’ denotes the identified pheromone and heuristic values.

PR [35] is a method used in the identification of trajectories linked leading to the best solution generated by heuristic methods. It can be considered as an enhanced version of scatter search. PR takes two or more efficient solutions generated by the original search and creates a path between them and also uses beyond the certain solutions in the given space.

Algorithm 1: Proposed MACO-PR Algorithm for protein tertiary structure prediction

```

1. Input: Amino acid sequences and energy function
2. Output: Optimized solution
3. begin
4.   Initialize pheno_path = null
5.   Initialize opt_path = null
6.   Define set X & Y
7.   while ant_no.: {1..n} do
8.     Find best path from set X
9.     Find best path from set Y
10.    OPn = compute (path)
11.    Apply equation 1
12.    if (OPn <= OPn+1)
13.      APn+1 = APn
14.      Update opt_path
15.    endif
16.    Apply pathrelinking (OPi, OPi+1)
17.  end while
18.  return opt_path
19.  end

```

Algorithm 1 illustrates the proposed MACO-PR prediction procedure. Initially, amino acid sequence and energy function are supplied as input to the proposed model. Subsequently, pheromone path and optimal path value have been assigned to Null as shown in line number 4 and 5. Line 6 defines the sets X and Y which contain the optimal paths. Next line numbers 7–11 compute the best path using Eq. (1), for sets X and Y. Line numbers 12–15 deal with the computation and comparison of the best path, meanwhile the best path is stored in ‘opt_path’ variable. Next, apply path-relinking method to compare the best path as shown in line number 16. The path-relinking process is applied

seeing that the latest result OP_i and OP_{i+1} from Y is randomly selected. It starts with OP_i value $i = 1.n$ and at each stage OP_i values is substituted by OP_{i+1} . Later all solutions should be visited through this process. Finally, line number 18 returns the best path. Here path-relinked is used to reduce local minima ensues during torsion angle adjustment as shown in line number 16.

5 Results

The proposed work was used to predict the protein tertiary structure. Here six different protein sequences have been used in the identification of protein structure. The six used protein sequences are 2P5K, 1AIL, 1AB1, 1L2Y, 2MTW, and 1WQC. The author has enhanced the ACO algorithm by incorporating path-relinking strategies for the effective retrieval of protein tertiary structure. Later lowest potential energy was used in structural analysis. Furthermore, Root Mean Square Deviation (RMSD) values were computed using PyRosetta [36], as shown in Eq. (2).

$$RMSD(x, y) = \sqrt{\sum_{i=1}^n \frac{\|r_{xi} - r_{yi}\|^2}{n}} \quad (2)$$

here r_{xi} and r_{yi} are the values representing the location of atom vectors representing the positions of the given atom. Tables 1 and 2 show the computed average energy and RMSD values for the proposed MACO-PR algorithm and normal ACO algorithm, here x-axis represents the PDB ID while y-axis represents computed average energy (negative values) or RMSD values (positive values) as displayed in Figs. 2 and 3.

For each six-protein sequence, minimum energy is computed using the proposed MACO-PR algorithm, and subsequently generated result is analyzed using ACO and RMSD methods. 2MTW protein ID generates the lowest energy $-17.45 \text{ kcal/mol}^{-1}$ using MACO-PR whereas $-16.83 \text{ kcal/mol}^{-1}$ by ACO method. Next 2P5K & 1AIL PDB ID's produces $-14.5 \text{ kcal/mol}^{-1}$ and $-13.0 \text{ kcal/mol}^{-1}$ energy when

Table 1 Average energy ($\text{Kcal} / \text{mol}^{-1}$) values computed through proposed MACO-PR algorithm and ACO algorithm

PDB ID	Avg. Energy with path-relinking (MACO-PR)	Avg. Energy without path-relinking (ACO)
2P5K	-14.5	-12.01
1AIL	-13.0	-12.55
1AB1	-16.5	-14.22
1L2Y	-14.29	-15.11
2MTW	-17.45	-16.83
1WQC	-15.06	-15.70

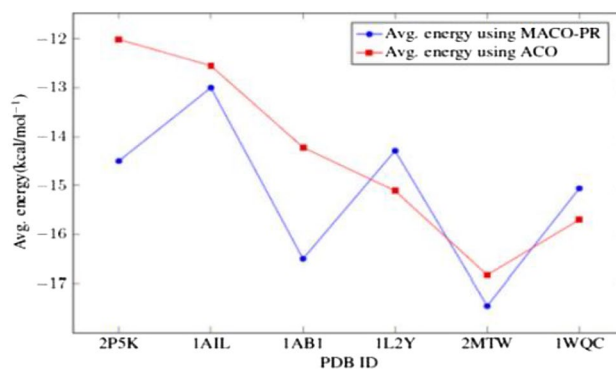


Fig. 2 Average energy ($\text{Kcal} / \text{mol}^{-1}$) values computed through proposed MACO-PR algorithm and ACO algorithm

computed using MACO-PR while $-12.01 \text{ kcal/mol}^{-1}$ and $-12.55 \text{ kcal/mol}^{-1}$ using ACO. Similarly, 1AB1 generates the average energy of $-16.5 \text{ kcal/mol}^{-1}$ and $-14.22 \text{ kcal/mol}^{-1}$ by incorporating MACO-PR and ACO methods. The result shows that out of six protein ID four protein ID generate the lowest energy when computed using the proposed model MACO-PR whereas the remaining two protein ID produces the lowest energy using ACO as shown in Fig. 2.

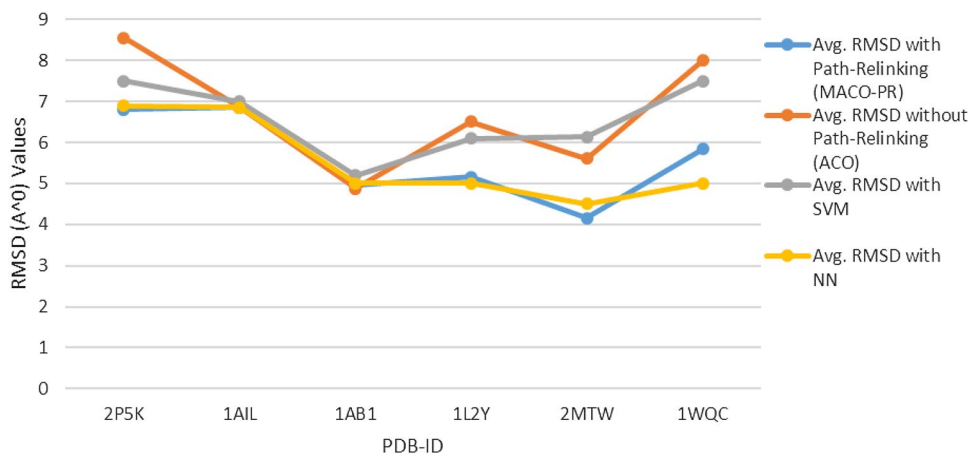
Figure 3, presents the calculated RMSD values using MACO-PR, ACO, SVM[32] and NN[33]. Here 2MTW PDB ID produces the lowest RMSD value of 4.16 using proposed MACO-PR while 5.61 using ACO and 6.5 using SVM. Similarly, 1L2Y generates a 5.16 RMSD values when computed using MACO-PR whereas 6.51 and 6.1 use ACO and SVM. Further 2P5K and 1AIL generate RMSD values of 6.81 and 6.85 when computed using MACO-PR whereas 8.55 and 6.85 using ACO and 8.2 and 7 using SVM. Finally, 1WQC produces an RMSD value of 5.84 when computed using MACO-PR and 8.00 using ACO. The result exemplifies that out of six different PDB ID's 2P5K, 1AIL, 1L2Y, 2MTW and 1WQC perform well when computed using the proposed model MACO-PR while ACO generate good result only for 1AB1 ID, similar case with SVM.

6 Conclusion and future scope

The study of protein and its structure is one of the key research areas in Computational Bioinformatics [1, 37, 38]. In this work, the author presents two different versions of the ACO method for protein tertiary structure prediction. In the first version, basic ACO, SVM and NN is used, whereas in the second version author has integrated MACO with PR strategy and SK. In the proposed work, Amino acid sequences are extracted from PDB. Next, the extracted amino acid and template are supplied as input to the proposed MACO-PR model. Here PR is used in

Table 2 RMSD (Å) values computed through proposed MACO-PR algorithm, ACO, SVM[32] and NN [33]

PDB ID	Avg. RMSD with Path-Relinking (MACO-PR)	Avg. RMSD without Path-Relinking (ACO)	Avg. RMSD with SVM	Avg. RMSD with NN
2P5K	6.81	8.55	7.5	6.9
1AIL	6.85	6.85	7	6.85
1AB1	4.96	4.88	5.2	5
1L2Y	5.16	6.51	6.1	5
2MTW	4.16	5.61	6.14	4.5
1WQC	5.84	8.00	7.5	5

Fig. 3 RMSD (Å) values computed through proposed MACO-PR, ACO, SVM and NN algorithms

overcoming the local minima of torsion angle generated during structure computation; however, SK is incorporated to handle the search space complexity.

The average energy measured using proposed model MACO-PR for 2P5K is $-14.5 \text{ Kcal/Mol}^{-1}$, 1AB1 is $-16.5 \text{ Kcal/Mol}^{-1}$, 1A1L is $13.0 \text{ Kcal/Mol}^{-1}$, and 2MTW is $-17.45 \text{ Kcal/Mol}^{-1}$ PDB ID while $-12.01 \text{ Kcal/Mol}^{-1}$, $14.22 \text{ Kcal/Mol}^{-1}$, $12.55 \text{ Kcal/Mol}^{-1}$, $-16.83 \text{ Kcal/Mol}^{-1}$ using basic ACO. Similarly, RMSD values computed using the proposed model MACO-PR for PDB ID's 2P5K is 6.81 Å, 1A1L 6.85 Å, 1L2Y is 5.16 Å, 2MTW 4.16 Å and 1WQC is 5.84 Å, whereas using ACO 2P5K is 8.55 Å, 1A1L 6.85 Å, 1L2Y is 6.51 Å, 2MTW 5.61 Å and 1WQC is 8.00 Å.

The above average energy value indicates that the proposed model MACO-PR performs well for 2P5K, 1AB1, 1A1L, & 2MTW PDB ID's but failed to do well for 1AB1 PDB ID. Also, MACO-PR performs well for RMSD values for 2P5K, 1A1L, 1L2Y, 2MTW & 1WQC PDB ID's whereas failed to predict 1AB1 PDB ID. Overall all results illustrate that MACO-PR outperforms when compared with basic ACO, SVM and NN concerning average energy and RMSD measures. As a part of future work, the model can incorporate deep learning techniques for more timely and accurate prediction.

Funding Not applicable.

Data availability Not available.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Chou KC (2004) Structural bioinformatics and its impact to biomedical science. *Curr Med Chem*. 11(16):2105–2134. <https://doi.org/10.2174/0929867043364667>
2. Dorn M, Silva MB, Buriol LS, Lamb LC (2014) Three-dimensional protein structure prediction: methods and computational strategies. *Comput Biol Chem*. 53:251–276. <https://doi.org/10.1016/j.compbiolchem.2014.10.001>
3. Pandi Veerapandian (1997) Structure-based drug design. Marcel Dekker, New York. <https://doi.org/10.1021/jm980006n>
4. Breda A, Valadares NF, de Souza ON, Garratt RC (2007) Protein structure, modelling and applications. *Bioinformatics in Tropical Disease Research. A Practical and Case-Study Approach*
5. Smith Adam (2003) Protein misfolding. *Nat Rev. Drug Dis*. 426(6968):78–102. <https://doi.org/10.1038/426883a>
6. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and Human Disease. *Annu Rev Biochem*. 75:333–366. <https://doi.org/10.1146/annurev.biochem.75.101304.123901>

7. Dobson CM (2003) Protein folding and misfolding. *Nature*. 426(6968):884–890. <https://doi.org/10.1038/nature02261>
8. Adjuik TA, Ananey-Obiri D (2022) Word2vec neural model-based technique to generate protein vectors for combating COVID-19: a machine learning approach. *Int J Inform Technol* 14(7):3291–3299
9. Sofi MA, Wani MA (2022) Protein secondary structure prediction using data-partitioning combined with stacked convolutional neural networks and bidirectional gated recurrent units. *Int J Inform Technol* 14(5):2285–2295
10. Sekhar SM, Siddesh GM, Raj M, Manvi SS (2021) Protein class prediction based on Count Vectorizer and long short term memory. *Int J Inform Technol* 13(1):341–348. <https://doi.org/10.1007/s41870-022-00978-x>
11. Raja G, Reddy US (2022) Solve DNA sequence assembly problem using hybrid crow search optimization and multi classification techniques. *Int J Inform Technol* 14(5):2541–2547
12. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using rosetta. *Acad Press*. 383:66–93. [https://doi.org/10.1016/S0076-6879\(04\)83004-0](https://doi.org/10.1016/S0076-6879(04)83004-0)
13. Neumaier A (1997) Molecular modeling of proteins and mathematical prediction of protein structure. *SIAM Rev*. 39(3):407–460. <https://doi.org/10.1137/S0036144594278060>
14. Lesk AM (2002) Introduction to Bioinformatics. Oxford University Press Inc., New York, USA
15. Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. *Adv Protein Chem*. 23:238–438. [https://doi.org/10.1016/S0065-3233\(08\)60402-7](https://doi.org/10.1016/S0065-3233(08)60402-7)
16. Desjarlais JR, Clarke ND (1998) Computer search algorithms in protein modification and design. *Curr Opin Struct Biol*. 8(4):471–475. [https://doi.org/10.1016/S0959-440X\(98\)80125-5](https://doi.org/10.1016/S0959-440X(98)80125-5)
17. DasGupta D, Kaushik R, Jayaram B (2015) From Ramachandran Maps to Tertiary Structures of Proteins. *J Phys Chem B*. <https://doi.org/10.1021/acs.jpcc.5b02999>
18. Kryshchuk A, Fidelis K, Moulton J (2014) Casp10 results compared to those of previous casp experiments, proteins: *Struct. Funct Bioinf*. 82:164–174. <https://doi.org/10.1002/prot.24448>
19. Dorn M, de Souza ON (2008) CReF: a central-residue-fragment-based method for predicting approximate 3-D polypeptide structures. *Proceed ACM Sympos Appl Comput*. <https://doi.org/10.1145/1363686.1363979>
20. Dorn M, Breda A, Norberto de Souza O (2008) A hybrid method for the protein structure prediction problem. *Lect Notes Bioinf*. https://doi.org/10.1007/978-3-540-85557-6_5
21. Li T, Zhou C, Wang B, Xiao Boxiang, Zheng X (2018) A hybrid algorithm based on artificial bee colony and pigeon inspired optimization for 3D protein structure prediction. *J Bionosci*. 12(1):100–108. <https://doi.org/10.1166/jbns.2018.1499>
22. Xu Y, Zhou C, Zhang Q, Wang B (2016) 3D Protein Structure Prediction with BSA-TS Algorithm. *Intern Conf Indust Eng Appl Appl Intell Syst Spring*. https://doi.org/10.1007/978-3-319-42007-3_38
23. Xu J, Mcpartlon M, Li J (2021) Improved protein structure prediction by deep learning irrespective of co-evolution information. *Nat Mach Intell*. 3(7):1–9. <https://doi.org/10.1038/s42256-021-00348-5>
24. Correa LD, Lima Dorn M (2018) A knowledge-based artificial bee colony algorithm for the 3-D protein structure prediction problem. 2018 IEEE Congress on Evolutionary Computation (CEC). IEEE, New Jersey, USA, pp 1–8
25. Correa L, Borguesan B, Farfán C, Inostroza-Ponta M, Márcio Dorn (2016) A memetic algorithm for 3D protein structure prediction problem. *IEEE/ACM Trans Comput Biol Bioinform*. 15(3):690–704. <https://doi.org/10.1109/TCBB.2016.2635143>
26. Yousef M, Abdelkader T, Khaled, ElBahassy (2017) A hybrid model to predict proteins tertiary structure. In: 2017 12th International Conference on Computer Engineering and Systems (ICCES). IEEE. 85–91. <https://doi.org/10.1109/ICCES.2017.8275282>
27. Iraj MS, Ameri H (2016) RMSD protein Tertiary structure prediction with Soft Computing. *IJ Math Sci Comput*. 2:24–33. <https://doi.org/10.5815/ijmsc.2016.02.03>
28. Bao W, Chen Y, Chen Y (2013) Multiple Feature Fusion Protein Tertiary Structure Prediction. In: 2013 International Conference on Information Science and Cloud Computing Companion. IEEE. 751–756. <https://doi.org/10.1109/ISCC-C.2013.88>
29. Protein data bank.National Science Foundation. <https://www.rcsb.org/>
30. Kaushik AC, Sahi S (2017) Biological complexity: ant colony meta-heuristic optimization algorithm for protein folding. *Neural Comput Appl*. 28:3385–3391. <https://doi.org/10.1007/s00521-016-2252-5>
31. Resende M, Ribeiro C, Glover F, Mart R (2010) Scatter search and path-relinking: Fundamentals, advances, and applications. In: Gendreau M, Potvin J-Y (eds) *Handbook of Metaheuristics*. Springer, New York, pp 87–107
32. Mandale AK, Jain P, Shrivastava SK (2012) Protein structure prediction using support vector machine. *Int J Soft Comput* 3(1):67
33. Hou J, Wu T, Cao R, Cheng J (2019) Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. *Prot Struct Func Bioinform*. 87(12):1165–1178
34. Kousalya K, Balasubramanie P (2009) To improve ant algorithm's grid scheduling using local search. *Int. J. Comput. Cogn*. 7(4):47–57
35. Glover F (1997) Tabu search and adaptive memory programming—advances, applications and challenges. *Interf Comput Sci Operat Res Spring*. https://doi.org/10.1007/978-1-4615-4102-8_1
36. Chaudhury S, Lyskov S, Gray JJ (2010) Pyrosetta: a script-based interface for implementing molecular modelling algorithms using rosetta. *Bioinformatic*. 26(5):689–691. <https://doi.org/10.1093/bioinformatics/btq007>
37. Jain S (2018) Classification of protein kinase b using discrete wavelet transform. *Int J Inform Technol* 10(2):211–216. <https://doi.org/10.1007/s41870-018-0090-7>
38. Sofi MA, Wani MA (2022) Protein secondary structure prediction using data-partitioning combined with stacked convolutional neural networks and bidirectional gated recurrent units. *Int J Inform Technol*. 14(5):1–11. <https://doi.org/10.1007/s41870-022-00978-x>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.