Int. J. Advance Soft Compu. Appl, Vol. 18, No. 2, July 2026
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Protein Tertiary Structure Prediction using a β-Hill Climbing Optimizer

Mohammed Abual-Rub

Department of Management Information Systems, School of Business, King Faisal University, Al Ahsa, Saudi Arabia
e-mail: mabualrub@kfu.edu.sa

Abstract

Predicting the protein tertiary structure from its amino acids is a big challenge in current scientific research. There are huge numbers of protein sequences with unknown structures. Many computational methods, however, tried to solve this problem, but they achieved good results only for small size proteins. Most of these methods were not powerful- especially when applied to huge conformational. Many studies have attempted to solve this problem using either local search algorithms or global search algorithms. Other studies, on the other hand, have used hybrid algorithms that adapted local search-based algorithms as well as population-based meta-heuristic algorithms. This study introduces β- hill climbing algorithm as a robust algorithm to solve the ab initio protein tertiary structure prediction problem. The β- hill climbing Algorithm is used to find the local optimal solution within the search space by its Iterated Local Search (ILS). Furthermore, the proposed algorithm predicts the tertiary structure of a protein based on its sequence- and without any prior knowledge. The proposed algorithm has been evaluated using two protein sequences; Met-enkephalin (1PLW) and Plant Seed Protein (1CRN). The results show that the proposed algorithm provides more efficient solutions- compared to other previous studies.

Keywords: *Ab initio, Hill Climbing, Protein Structure Prediction, Protein Folding, Optimization.*

1 Introduction

Predicting protein's tertiary structure prediction from its amino acids sequence is a current big challenge in computational biology. There have been many attempts to overcome this challenge, but the methods used were not efficient in resolving that challenge. There are two main approaches to protein structure prediction, experimental and computational. X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM) are some of the primary experimental approaches for protein structure prediction. Although these methods are effective and powerful, they are expensive and time-consuming [1]. Computational methods , however, become more accurate useful in predicting the structure of the protein [2]. The computational methods for protein structure prediction can be classified into three categories:

- Homology modeling –also known as comparative modeling.
- Fold recognition –also known as protein threading.

Received 19 February 2026; Accepted 10 April 2026

- Ab initio modeling.

The applicability of each method is determined by the degree of similarity of protein to other sequences in the database. For example, fold recognition relies on the observation that any newly predicted structure will be almost folded to an existing structure in the database. This is because the number of newly folded protein structures is not rising quickly- compared to the number of newly discovered protein sequences; the accuracy of homology modeling, on the other hand, is determined by the degree in which the targeted sequence matches other sequences in the protein database. Ab initio approach relies on the thermodynamic concept that protein's tertiary structure is the lowest free energy conformation [3]. The difficulty of this approach stems from the fact that it is possible to deduce the structure of a protein from its amino acid sequence alone- without resorting to prior knowledge of similar folds. This technique is crucial because there are many proteins that don't share any similarities with other structured proteins. Nevertheless, some highly identical proteins have distinct tertiary structures [4]. An efficient search strategy that can locate the least energy of a protein with a given energy function is necessary for a successful ab initio approach to protein structure prediction. There are many methods to explore protein conformational search space, namely: Molecular Dynamics (MD), Monte Carlo (MC), Genetics Algorithm (GA), and harmony search algorithm (HSA).

β-hill climbing Algorithm- a recent algorithm that is an extension of the hill climbing algorithm- shows good results in enhancing the original algorithm [5]. The main advantage of this algorithm is that it controls the balance between exploration and exploitation during the search process.

The primary goal of this research is to provide an initial investigation of adapting β -hill climbing algorithm for ab initio Protein Tertiary Structure Prediction (PPSP). As compared to other approaches, β-hill climbing algorithm provides a higher quality solution and more accurate outcomes- when using a well-studied benchmark set for PPSP.

2 Related Work

There are two main approaches to protein structure prediction, experimental and computational. This research reviews the main computational methods that attempted to solve the protein tertiary structure prediction. The methods applied to the protein structure prediction can be classified into three categories: homology, protein folding, and ab initio. The homology method identifies an initial three-dimensional structure by comparing the linear amino acid sequence of the target protein to proteins that have determined structures. This process typically involves multiple alignments between the target and the candidate proteins.[6] Homology models , however, cannot fully resolve the issue since they require an amino acid sequence that closely resembles the target. When a homologous protein is unavailable, the fold recognition method can be employed. The fold recognition method can be applied if a homologous protein cannot be discovered. This approach utilizes templates of well-known structures that have previously been resolved and published in databases- like the Proteins Database (PDB). Additionally, homology and threading require prior knowledge of other proteins that have already been solved; the thing that does not guarantee a complete resolution of the issue. Unlike homology and protein folding methods, the ab initio strategy is not constrained by templates. It relies solely on the amino acid sequence to predict the tertiary structure. Additionally, ab initio poses a significant challenge for the scientific community due to its NP-hard computational complexity. The ab initio method has been used to apply a wide variety of

computational algorithms to solve the PTSP. PTSP is an optimization problem as we try to find the lowest energy of the protein, many optimization algorithms can contribute to solving such a problem. However, optimization algorithms can be classified into two categories: physical based algorithms and evolutionary based algorithms. Many physical based algorithms have achieved good results in optimization including [7] and [8]. On the other hand, there are many studies that have applied evolutionary based algorithms to solve PTSP: Harmony search algorithm [9], genetic Algorithms [10] Bees Algorithm [11], and Simulated Annealing [12] are among the most successful algorithms that solve the problem. Recently, an extension of the hill climbing algorithm has indicated good results in enhancing the original algorithm [5]. The main advantage of this algorithm is that it controls the balance between exploration and exploitation during the search process. Many optimization and search algorithms suffer from getting stuck in local optima, so, hybrid algorithms were one of the recommendations. Another advantage of β-hill climbing algorithm is that it controls the balance between exploration and exploitation during the search process. Moreover, it enhanced avoiding being stuck in local optima.

3 Protein Tertiary Structure Prediction (PTSP)

The thermodynamic hypothesis states that the tertiary structure of the protein is the conformation of the lowest free energy; this is the basis for *ab initio* protein tertiary structure prediction. [3]. The ab initio protein tertiary structure prediction method, which predicts the tertiary structure of a protein from its amino acid sequence alone, allows formulation of the PSPP as an optimization problem that attempts to locate the protein conformation with the lowest energy. Selecting an appropriate representation of the conformation and an appropriate energy function is the most significant task in the process of solving the PSPP using an optimization algorithm. In the following two sections, problem modeling is presented in accordance with these two aspects of the situation.

3.1 Mathematical Formulation

3.1.1 Problem representation

There are numerous popular representations of polypeptide chains, some of which are listed in [13]:

Cα coordinates, all-heavy-atom coordinates, all-atom three-dimensional coordinates, backbone atom coordinates + side-chain centroid, and backbone and side-chain torsion angles. Some of those representations provide a thorough portrayal, while others provide a simpler one.

All of the protein's atoms and their interactions lead to the most accurate representation; nevertheless, this representation is computationally expensive and unnecessary for the search process [14]. For that, many researchers used a simplified representation to reduce the computational time and space [15], [16], [17],[18] and [19]. This research, like many others, employs the backbone and side-chain torsion angles representation. This is a streamlined representation of the polypeptide chain which is based on the fact that each type of residue needs a specific number of torsion angles to determine the three-dimensional coordinates of all atoms.

The protein is represented in this research as a vector of amino acids, $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_M)$, where $\boldsymbol{x}_1 = (\Phi_1(1), \Psi_1(1), \omega_1(1), x_1(1), x_1(2), ..., x_1(N))$, $\boldsymbol{x}_2 = (\Phi_2(1), \Psi_2(1), \omega_2(1), x_2(1), x_2(2), ..., x_2(N))$, ..., $\boldsymbol{x}_M = (\Phi_M(1), \Psi_M(1), \omega_M(1), x_M(1), x_M(2), ..., x_M(N))$.

The technique that has been suggested, on the other hand, treats this vector as if it were a vector of torsion angles, denoted by $\mathbf{x}_i = (x(1), x(2), ..., x(N))$, where N is the total number of torsion angles present in the protein, and where each angle ($x(i)$) in this vector can be assigned a value that falls anywhere between $[-\pi, \pi]$. As a result, the number of torsion angles in the protein has an equivalent relationship to the length of the solution. The amino acid consists of two components: (i) main chain angles (Φ, Ψ, ω) and (ii) side chain angles $x_i=(x_i(1), x_i(2), ..., x_M(N))$.

The main chain and the appropriate amount of side chain angles (this varies from amino acid to amino acid) must be present for each amino acid. Therefore, this study focuses on the vector $\mathbf{X} = (x(1), x(2), ..., x(\mathrm{N}))$ of amino acids in the side chain, where each amino acid is assigned a torsion angle $x_i(j)$) in the range $[-\pi, \pi]$.

3.1.2 Energy function

CHARMM by [20], AMBER by [21], and OPLS by [22] are only a few examples of well-known physics-based force fields. Unfortunately, customers have a difficulty in modifying these force field packages to work with their own algorithms [23]. Instead, many researchers have sought an energy function that may be tailored to their own research objectives -while still providing adequate performance in assessment studies. Therefore, in this study, an energy function based on Simple Molecular Mechanics for Proteins (SMMP) is employed. An up-to-date software suite for modeling proteins. This set of force fields was initially developed by [24] and updated by [23]. The updated version is used in this study.

Using SMMP has many advantages, first: the software just needs a single computer to run at maximum speed and be fully exploited. Second, it is a free and available source. Third, it is easy for users to make changes to the code and tailor it to their own requirements. Fourth, there are no hardware specific procedures in the code. It has been tried and examined in numerous tiny protein simulations, and it has been proven to be a useful tool for numerical examinations of proteins [15]. Each protein molecule in SMMP is represented by a set of internal coordinates, and the dihedral angles (Φ, Ψ, ω); this indicates rotations around the chemical bonds in the backbone of the amino acids. In addition, the dihedral angles X_i in the side chains are flexible. These dihedral angles describe rotations around the chemical bonds in the side chains. When calculating the internal energy, a series of energy minimization methods are utilized. These routines are based on the ECEPP force field and they use two separate parameter sets (ECEPP/2 potential and ECEPP/3 potential) respectively. SMMP utilized the following energy function in its calculations [24]:

$$Min\ f(x) = ELJ + Eel + Ehb + Etors \quad (1)$$

$$E_{LJ} = \sum_{j>i}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) \quad (2)$$

$$E_{el} = 332 \sum_{j>i}\left(\frac{q_i}{\varepsilon} \times \frac{q_j}{r_{ij}}\right) \quad (3)$$

$$E_{hb} = \sum_{j>i}\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) \quad (4)$$

$$E_{tors} = \sum_n U_n(1 \pm cos(k_n \times \varphi_n)) \quad (5)$$

r_{ij} refers to the distance in Å between atoms i and j whereas Aij, Bij, Cij, Dij are the empirical potential parameters. qi and qj are abbreviations for the partial charges that are carried by atoms i and j, respectively. ε stands for the dielectric constant of the environment, and it is advised that ε = 2 be used. The value 332, which can be found in Equation 3, is the factor that is used to define the energy in kcal/mol. The energetic torsion barrier of rotation about the bound n is denoted by the symbol U_n, the multiplicity of the torsion angle φn is denoted by the symbol k_n.

3.2 *β*-Hill Climbing Optimizer for PTSP: Proposed Method

Recently, a new version of the simple local search is introduced by Al-Betar [5] for continuous optimization benchmark functions. Its results were successful on IEEE-CEC2005 datasets [25]. The main advantage of the *β*-hill climbing is its ability to explore the wide search space regions trying to find the absolute optima. *β*-operator controlled by *β* parameter is used to explore the promising search space regions. Making use of the tricky behavior of the *β*-operator functions, the existing solution's decision variable values are examined to determine if they should be replaced with random values drawn from the possible ranges for those variables.

Technically speaking, *β*-hill climbing algorithm begins with a tentative solution, $\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N})$, This tentative solution is transformed into a new nearby solution, $\mathbf{X}r = (\boldsymbol{x}^{\mathbf{r}}, \boldsymbol{x}^{\mathbf{r}}, \ldots, \boldsymbol{x}^{\mathbf{r}})$ at each iteration by employing two operators: N -operator and *β*-operator. While N-operator performs exploitation through the neighborhood search of a random walk, β-operator performs exploration in a manner that is analogous to that of the uniform mutation operator. Iteration after iteration, up until the stop condition is satisfied, a more efficient result could be reached by using either the N-operator stage or the β -operator stage (or both).

To make it clear, assume that the optimization problem is represented as below:

$$\text{Min } f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbf{X},$$

f (x) is the objective function which evaluates the solution $X = (x_1, x_2, \ldots, x_N)$ which involves a set of decision variables. Each decision variable $xi \in Xi$ is a possible value range for each decision variable, where $Xi \in [LBi, UBi]$, and LBi is the lower bound, and UBi is the upper bound of the decision variable xi while N is the total number of decision variables.

An initial solution is chosen at random, and then it is subjected to an evaluation with the help of the objective function f(x). Using an N-operator, the solution will be improved iteration by iteration, the function improve(N (x)) will be used in conjunction with an acceptance rule known as "random walk" [5] such that a random neighboring solution of the current solution x is visited as follows:

$$x' = xi \pm U(0, 1) \times N \exists i \in [1, N] \quad (6)$$

where i is a number within the feasible range [1, 2, … , N], and i is selected at random. The distance N between the present solution and the next-door solution is the bandwidth parameter. β-operator: The present solution is the N -operator neighboring solution. New solution variables are: either assigned values that are mathematically connected to the present solution's values , or are selected at random from the β parameter's range as follows:

$$x_i'' \leftarrow \begin{cases} x_k & ra < \beta \\ x_i' & otherwise \end{cases}$$

where $\beta \in [0, 1]$, a uniform random integer between 0 and 1 is produced by *ra*, and the value of the decision variable x^r is chosen at random from the interval xk Xi. The -hill climbing pseudocode is shown in Algorithm 1.

ra generates a uniform random number between 0 and 1, and $x_k \in \boldsymbol{X}_i$ is randomly selected from the possible range for the decision variable x^r . Algorithm 1 shows the pseudocode of the β-hill climbing.

Algorithm 1 β–hill climbing pseudocode

1: $x_i = LB_i + (UB_i - LB_i) \times U(0, 1), \forall i = 1, 2, \ldots, N$

2: Calculate($f(\boldsymbol{x})$)

3: $t = 0$

4: **while** ($t \leq$ Max-t) **do**

5: $\boldsymbol{x}' = improve(\mathrm{N}(\boldsymbol{x}))$

6: $\boldsymbol{x}'' = \boldsymbol{x}'$

7: **for** $i = 1, \cdots, N$ **do**

8: **if** ($ra \leq \beta$) **then**

9: $x'' = LB_i + (UB_i - LB_i) \times U(0, 1)$

10: **end if**

11: **end for** 12: **if** ($f(\boldsymbol{x}'') \leq f(\boldsymbol{x})$ **then** 13: $\boldsymbol{x} = \boldsymbol{x}''$

14: $f(\boldsymbol{x}) = f(\boldsymbol{x}'')$

15: **end if**

16: $t = t + 1$

17: **end while**

4 Experimental Results

4.1 Dataset

The algorithm has been tested on two protein sequences from the Protein Data Bank (1PLW, or "Met-enkephalin") and (1CRN," or "Plant Seed Protein"), respectively. Many approaches, including the methodology provided in this study, are evaluated using these two proteins. Although it was initially isolated from the brain's enkephalin mixture, the short residue sequence of 5 amino acids that make "Met-enkephalin" enables it to be a popular model. The 24 backbone and side chain angles in its five amino acids explain the orientations of 75 atoms in the molecule. Met-enkephalin, although being a relatively small protein, requires a highly complicated conformational space with more than 10^{11} local minima [26]. Met-enkephalin has been utilized extensively for testing simulation methods due to the complexity of its configuration space. In addition, its relatively short sequence

facilitates extensive computational research. Met-enkephalin is a standard which is used as measurement by many computational models [27].

The second protein sequence, '1CRN', has 46 amino acids consisting of 327 atoms described by 238 independent backbone and side chain angles [9].

4.2 Experimental design

A series of experiments have been conducted with different values of the main parameters to measure the effect of different values on the performance of the proposed β-hill climbing algorithm. As indicated in Table 1, nine distinct parameter configurations have been selected for this study. Each case has been examined a total of thirty times for both benchmark protein sequences. The β value is selected in different cases with large (i.e., 0.10), medium (i.e., 0.05), and small (i.e., 0.01) values. Recall, β is used to explore the promising search space regions. When the value of β increases, there is a corresponding increase in exploration, but there is a corresponding drop in exploitation. When β has a low value, the search algorithm prioritizes efficient use of resources by prioritizing exploitation over exploration in the early stages of the search process. On the other hand, having a big value for β has resulted in improved exploitation during the first stage of the search.

Table 1: Cases used to evaluate the β-Hill Climbing convergence ability.

Cases	β	N
case1	0.01	0.10
case2	0.01	0.50
case3	0.01	0.90
case4	0.05	0.10
case5	0.05	0.50
case6	0.05	0.90
case7	0.10	0.10
case8	0.10	0.50
case9	0.10	0.90

The N (bandwidth) is the rate at which the angles are chosen, either from the best vector in memory with the lowest energy value, or at random. Low (i.e. 0.10), medium (i.e. 0.50), and high amounts (i.e. 0.90) of N have all been examined. If N is set to a value of 0.10, then the angle's value will be chosen at random from all solutions in memory with a 10% probability, and from the best vector with a 90% probability. The number of improvised iterations (NI) is set to NI=100,000 across all 9 cases.

4.3 Results

The energy values obtained for both 'Met-enkephalin' and '1CRN' for each of the 9 cases after 30 runs are presented in Table 2 -along with the best, average, and worst results as well as the standard deviation for each case. The best individual outcomes are shown in bold, and the best averages are denoted in italics. As indicated in Table 2, case2 and case3 have obtained the lowest energy and best results; that is because β value in both cases is equal to 0.01, which is low, with N value equals to 0.50 and 0.90 respectively. As previously explained, when β has a low value, the search algorithm prioritizes exploitation

over exploration in the early stages of the search process, so case2 and case3 with N = 0.90 means that the angle's value will be chosen from the best vector with a 90% probability, which causes the best and lowest energy value.

The proposed algorithm has determined the lowest energy of Met-enkephalin from the two versions of the ECEPP force field, ECEPP/2 and ECEPP/3 with ω relaxed. However, it outperformed the lowest energy obtained by both ECEPP/2 and ECEPP/3 with ω fixed at 180. The new energy minimum for E/3π = -10.90 and for E/2π = -11.26 while the previous best energies were: E/2π = -10.72 and E/3π = -10.90 by Zhan et al. [27].

Table 2: Results of *β*-Hill Climbing for 30 Runs of the 9 Cases.

	Met-enkephalin				**1CRN**			
Case	best	avg	worst	std.dev	best	avg	worst	std.dev
Case1	-9.98	-8.77	-6.73	1.00	-150.01	-125.52	-1.68	28.81
Case2	**-12.43**	-10.87	-10.11	0.44	**-204.81**	*-156.83*	-141.42	16.85
Case3	**-12.43**	-9.05	-7.17	1.37	-155.56	-115.76	89.87	56.89
Case4	-11.52	-4.52	-2.57	0.11	-184.85	-130.25	-52.63	31.94
Case5	-3.39	-1.4	0.11	0.95	-149.95	578.61	20246.3	3715.17
Case6	-12.41	*-11.55*	-10.75	0.51	-201.65	-140.15	-103.25	16.70
Case7	-11.17	-7.68	-6.16	1.06	-134.01	-54.50	518.31	129.06
Case8	-4.13	-2.32	-1.3	0.72	-134.47	281.35	8829.89	1645.14
Case9	-12.39	-11.15	-10.59	0.50	-165.41	-154.31	-108.15	10.98

4.4 Comparative Analysis

Table 3 shows the results of β-Hill Climbing algorithm in comparison to some previous studies. The proposed algorithm has obtained the best results recorded until now for the Met-enkephalin protein- which are **-12.43** for ECEPP/3 force field by Zhan et al. [27]. and -12.91 for ECEPP/2 by Zhan et al. [27], Meirovitch et al. [28], and Li and Scheraga [26]. Moreover, the algorithm has outperformed some other previous results by Androulakis et al. [29] and Eisenmenger et al. [23].

Table 3. The lowest energies (in kcal/mol) obtained in previous studies of Met-enkephalin and in this study based on ECEPP/3 and ECEPP/2 force fields with ω relaxed or fixed at 180˚.

Source	**Force field**	**Energy**
Eisenmenger & Hansmann (1997)	ECEPP/3 with ω = 180	-10.85
Androulakis et al (1997)	ECEPP/3	-11.70
Lixin Zhan (2006)	ECEPP/3	-12.43
Lixin Zhan (2006)	ECEPP/3 with ω = 180	-10.90
Classical HSA	ECEPP/3	-11.95
β-Hill Climbing	**ECEPP/3**	**-12.43**
β-Hill Climbing	**ECEPP/3 with ω = 180**	**-11.26**
Eisenmenger&Hansmann (1997)	ECEPP/2 ω = 180	-10.72
Meirrovitch et. al (1994)	ECEPP/2	-12.91
Lixin Zhan (2006)	ECEPP/2	-12.91
Lixin Zhan (2006)	ECEPP/2 ω = 180	-10.72
Eisenmenger&Hansmann (1997)	ECEPP/2 ω = 180	-10.72

Meirrovitch et. al (1994)	ECEPP/2	-12.91
β-Hill Climbing	**ECEPP/2**	**-12.91**
β-Hill Climbing	**ECEPP/2 ω = 180**	**-11.21**

Table 4. The internal coordinates of the lowest energy for Met-enkephalin

		β-Hill Climbing		Zhan		Androulaksi	Eisenmenger
	Torsion	**E/3**	**E/3 π**	**E/3 (a)**	**E/3 (a) π**	**E/3 (b)**	**E/3 (b) π**
Tyr1	x1	**-173.20**	**180.00**	-173.20	59.90	-173.20	-174.20
	x2	**79.30**	**92.77**	-100.70	94.10	-100.50	-85.20
	x6	**-166.30**	**180.00**	13.70	-21.30	13.60	2.80
	φ	**-83.10**	**-162.92**	-83.10	168.10	-83.50	-162.70
	ψ	**155.80**	**-40.62**	155.80	0.90	155.80	-41.70
	ω	**-177.10**	**180.00**	-177.10	180.00	177.20	180.00
Gly2	φ	**-154.20**	**65.62**	-154.20	126.80	-154.30	65.80
	ψ	**85.80**	**-89.02**	85.50	-21.20	86.00	-87.00
	ω	**168.50**	**180.00**	168.50	180.00	168.50	180.00
Gly3	φ	**83.00**	**-153.34**	83.00	83.70	83.00	-157.30
	ψ	**-75.00**	**34.24**	-75.00	-61.60	-75.10	34.90
	ω	**-170.00**	**180.00**	-170.00	180.00	-169.90	180.00
Phe4	x1	**58.90**	**52.21**	58.90	58.60	58.80	52.40
	x2	**94.50**	**-97.06**	-85.50	92.90	-85.50	-96.00
	φ	**-136.80**	**-155.95**	-136.80	-128.20	-136.90	-158.80
	ψ	**19.10**	**160.10**	19.10	18.80	19.10	159.50
	ω	**-174.10**	**180.00**	-174.10	180.00	-174.10	180.00
Met5	x1	**52.90**	**-66.82**	52.90	55.70	52.90	-66.10
	x2	**175.30**	**-180.00**	175.30	-178.60	175.30	-179.60
	x3	**180.10**	**180.00**	-179.90	177.00	-179.90	-179.90
	x4	**-58.60**	**180.00**	-178.60	-179.30	-178.60	60.10
	φ	**-163.40**	**-79.33**	-163.40	-162.10	-163.40	-82.40
	ψ	**160.80**	**129.64**	160.80	7.50	160.80	134.10
	ω	**180.00**	**180.00**	-179.80	180.00	-179.80	180.00
Energy (kcal/mol)		**-12.43**	**-11.26**	-12.43	-10.90	-11.71	-10.85

The internal coordinates of the lowest energy for the Met-enkephalin protein found in this study (columns 3&4), Zhan study (columns 5&6), Androulaksi study (column 7), and the & Hansmann study (column 8) are shown in Table 4. The label E/3 denotes the configurations obtained using ECEPP/3 force field. As indicated in Tables 3 and 4, the lowest energies that have been obtained in this paper match the previous best results

exactly for both ECEPP/3 and ECEPP/2 force fields while it outperformed the lowest energy using the ECEPP/3 and ECEPP/2 force fields with ω fixed at 180˚.

5 Conclusion

In this paper, the problem of predicting the protein tertiary structure from its sequence is considered. The Iterated Local Search (ILS) of the β- hill climbing Algorithm is used to find the local optimal solution within the search space after modelling protein tertiary structure prediction problem- as an optimization of an objective function. The proposed algorithm has solved this problem successfully. A series of experiments has been conducted to measure the influence of different parameters and operators on the convergence of β- hill climbing Algorithm.

The experimental results of two different proteins have showed that β- hill climbing Algorithm is able to obtain good results and can determine the lowest energy of Met-enkephalin from the ECEPP/3 force field. With the peptide dihedral angles ω relaxed, while it outperforms the lowest energy recorded for the ECEPP/3 force field with ω fixed at 180˚ the new energy recorded is -11.26 while the previous lowest energy was -10.90. One avenue for future research is to hybridize our proposed algorithm with other population-based metaheuristic algorithms, and another avenue is to try other beta operators such as self-adaptive mutation. Moreover, more benchmarks can be investigated to further test the performance of the proposed algorithm.

ACKNOWLEDGEMENTS

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU250412].

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] Abual-Rub, M. S., & Abdullah, R. (2008). A survey of protein fold recognition algorithms. *Journal of Computer Science*, *4*(9), 768.

[2] Liu, J., Wu, T., Guo, Z., Hou, J., & Cheng, J. Improving protein tertiary structure prediction by deep learning and distance prediction in CASP14. bioRxiv 2021. *CrossRef][PubMed]*.

[3] Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. Science, 181(4096), 223-230.

[4] Helles, G. (2008). A comparative study of the reported performance of ab initio protein structure prediction algorithms. *Journal of the royal society interface*, *5*(21), 387-396.

[5] Al-Betar, M. A. (2017). β-hill climbing: an exploratory local search. *Neural Computing and Applications*, *28*(Suppl 1), 153-168.

[6] Frausto-Solís, J., Sánchez-Hernández, J. P., Maldonado-Nava, F. G., & González-Barbosa, J. J. (2019). GRSA Enhanced for Protein Folding Problem in the Case of Peptides. *Axioms*, *8*(4), 136.

[7] Shehadeh, H.A. (2023) Chernobyl disaster optimizer (CDO): a novel meta-heuristic method for global optimization. *Neural Computing and Applications*, **35**(15): p. 10733-10749.

[8] Shehadeh, H.A. (2025). Bermuda Triangle Optimizer (BTO): A Novel Metaheuristic Method for Global Optimization. I*nternational Journal of Advances in Soft Computing & Its Applications*, **17**(2).

[9] Abual-Rub, M. S., Al-Betar, M. A., Abdullah, R., & Khader, A. T. (2012). A hybrid harmony search algorithm for ab initio protein tertiary structure prediction. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *1*(3), 69-85.

[10] Sindhu, G., & Sudha, S. (2012). Prediction of protein tertiary structure using genetic algorithm. In *Soft Computing Techniques in Vision Science* (pp. 15-22). Berlin, Heidelberg: Springer Berlin Heidelberg.

[11] Bahamish, H. A. A., Abdullah, R., & Salam, R. A. (2008, May). Protein conformational search using bees algorithm. In *2008 Second Asia International Conference on Modelling & Simulation (AMS)* (pp. 911-916). IEEE.

[12] Lee, J., & Scheraga, H. A. (1999). Conformational space annealing by parallel computations: extensive conformational search of Met-enkephalin and of the 20-residue membrane-bound portion of melittin. *International Journal of Quantum Chemistry*, *75*(3), 255-265.

[13] Cutello, V., Narzisi, G., & Nicosia, G. (2006). A multi-objective evolutionary approach to the protein structure prediction problem. *Journal of The Royal Society Interface*, *3*(6), 139-151.

[14] Chivian, D., Robertson, T., Bonneau, R., & Baker, D. (2003). Ab initio methods. *Structural bioinformatics*, *44*, 547-557.

[15] Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding. *Journal of molecular biology*, *104*(1), 59-107.

[16] Abagyan, R. A., & Maiorov, V. N. (1988). A simple qualitative representation of polypeptide chain folds: Comparison of protein tertiary structures. *Journal of Biomolecular Structure and Dynamics*, *5*(6), 1267-1279.

[17] Hinds, D. A., & Levitt, M. (1992). A lattice model for protein structure prediction at low resolution. *Proceedings of the National Academy of Sciences*, *89*(7), 2536-2540.

[18] Baker, D. (2000). A surprising simplicity to protein folding. *Nature*, *405*(6782), 39-42.

[19] Dudek, M. J. (2007). igor, a Simple Integrable Model of a Polypeptide Chain III. Model Parameterization.

[20] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), 187-217.

[21] Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., ... & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, *106*(3), 765-784.

[22] Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, *110*(6), 1657-1666.

[23] Eisenmenger, F., Hansmann, U. H., Hayryan, S., & Hu, C. K. (2006). An enhanced version of SMMP—open-source software package for simulation of proteins. *Computer physics communications*, *174*(5), 422-429.

[24] Eisenmenger, F., Hansmann, U. H., Hayryan, S., & Hu, C. K. (2001). [SMMP] A modern package for simulation of proteins. *Computer physics communications*, *138*(2), 192-212.

[25] Suganthan, P. N., Hansen, N., Liang, J. J., Deb, K., Chen, Y. P., Auger, A., & Tiwari, S. (2005). Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization. *KanGAL report*, *2005005*(2005), 2005.

[26] Li, Z., & Scheraga, H. A. (1988). Structure and free energy of complex thermodynamic systems. *Journal of Molecular Structure: THEOCHEM*, *179*(1), 333-352.

[27] Zhan, L., Chen, J. Z., & Liu, W. K. (2006). Conformational study of met-enkephalin based on the ECEPP force fields. *Biophysical journal*, *91*(7), 2399-2404.

[28] Meirovitch, H., Meirovitch, E., Michel, A. G., & Vasquez, M. (1994). A simple and effective procedure for conformational search of macromolecules: application to met-and leu-enkephalin. *The Journal of Physical Chemistry*, *98*(25), 6241-6243.

[29] Androulakis, I. P., Maranas, C. D., & Floudas, C. A. (1997). Prediction of oligopeptide conformations via deterministic global optimization. *Journal of Global Optimization*, *11*(1), 1-34.