

# Multimeme Algorithms For Protein Structure Prediction

N. Krasnogor, B.P. Blackburne, E.K. Burke, and J.D. Hirst

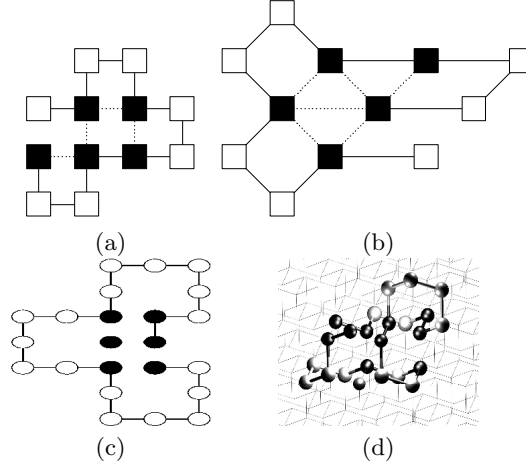
Automated Scheduling, Optimization and Planning Group  
and Computational Biophysics and Chemistry Group  
University of Nottingham, Nottingham, United Kingdom  
<http://dirac.chem.nott.ac.uk/~natk/Public/index.html>

**Abstract.** Despite intensive studies during the last 30 years researchers are yet far from the “holy grail” of blind structure prediction of the three dimensional native state of a protein from its sequence of amino acids. We introduce here a Multimeme Algorithm which is robust across a range of protein structure models and instances. New benchmark sequences for the triangular lattice in the HP model and Functional Model Proteins in two and three dimensions are included here with their known optima. As there is no favourite protein model nor exact energy potentials to describe proteins, robustness across a range of models must be present in any putative structure prediction algorithm. We demonstrate in this paper that while our algorithm presents this feature it remains, in terms of cost, competitive with other techniques.

## 1 Introduction

A protein’s structure determines its biological function. This is the reason why a central component in proteomics is the prediction of a protein’s *native* structure from its sequence. This task is called Protein Structure Prediction (PSP). “All-atom” simulations are extremely expensive so researchers often resort to simplified models of the *PSP*, but even the simplified problem still remains computationally intractable in the worst case[2].

The particular simplified models we are concerned with in this paper are the HP model[5] and Functional Model Proteins[10][3] in two and three dimensional lattices. The HP model (and its variants) abstracts the hydrophobic interaction process in protein folding by reducing a protein to a heteropolymer of non-polar or hydrophobic (H) and polar (P) or hydrophilic amino acids. A protein sequence  $s$  is represented by a string in a binary alphabet:  $s \in \{H, P\}$ . Simplified models restrict the space of conformations to self-avoiding paths on a lattice in which vertices are labeled by the amino acids. These lattices may be two-dimensional, e.g. square or triangular, or three dimensional, e.g. diamond. The energy potential in the HP model reflects the fact that hydrophobic amino acids have a propensity to form a hydrophobic core. To capture this feature of protein structures, the HP model adds a value  $\epsilon$  for every pair of hydrophobes that form a topological contact; a topological *contact* is formed by a pair of amino



**Fig. 1.** HP protein embedded in the square lattice (a) and triangular lattice (b). Functional Model protein embedded in the square lattice (c) and diamond (3D) lattice (d). In (c) and (d) native structures are not maximally compact as they must have a “binding pocket”.

acids that are adjacent on the lattice and not consecutive in the sequence. The value of  $\epsilon$  is typically taken to be  $-1$ . Figure 1 shows sequences embedded in the square and the triangular lattices, with hydrophobic-hydrophobic contacts (HH contacts) highlighted with dotted lines. The conformation in Figure 1 has an energy of  $-4$  in the square lattice embedding and  $-6$  in the triangular lattice embedding. A typical interaction matrix for the HP model is given in table 1(a). The energy interaction in Functional Model Proteins[10],[3] (which introduces

(a)	H	P
H	1	0
P	0	0

(b)	H	P
H	-2	1
P	1	1

**Table 1.** Interaction energy matrix for the standard HP model (a) and Interaction energy matrix for a shifted HP model (b).

repulsive forces) between residues that are in contact is given by table 1(b). Native protein structures in this model are required to have a binding pocket in their native structure (e.g. a hole in their conformation), an energy gap between the minimum energy conformation and the next excited state and to have a unique optimal conformation. Figure 1(c) shows a two dimensional embedding of a Functional Model Protein and 1(d) shows a diamond lattice embedding.

The construction of effective algorithms for solving structure prediction on simplified models (e.g. the HP model and Functional Model Proteins) is a key stepping-stone towards the structure prediction of real life proteins that cannot

be solved by homology or threading methods. Several successful methodologies from the last two Critical Assessment of Structure Prediction[21], CASP3 and CASP4, employed simplified models for sampling and optimising structures embedded in different lattices[20],[7],[12].

In this paper we will present a novel metaheuristic, called a Multimeme Algorithm, to PSP for four different models: HP model in the square lattice, HP model in the triangular lattice, Functional Model Proteins in the square lattice and the diamond lattice. To evaluate our algorithm in the first model we will use instances from the public domain that were used by other researchers to test their methods. In the case of the last three models new instances, with their respective optima, will be presented and used as test beds.

## 2 Evolutionary Algorithms Approaches to Protein Structure Prediction

Several evolutionary algorithms precede the application of Multimeme Algorithms in PSP. An early application of Genetic Algorithms (GAs) to PSP due to Unger and Moult [19] is a widely used benchmark. Patton et al. [6] described a standard GA employing, as Unger and Moult did, an internal coordinate representation. They used a penalty method to enforce the self-avoiding constraints. Khimasia and Coveney [11] considered the performance of Goldberg's Simple GA. The objective function was a hybrid between the Random Energy Model[4] and the HP model. Colosimo et al. [18] applied a standard GA to predict the minimum conformational energy of two small real proteins: crambin and ferredoxin. They used the HP model in various 3D cubic grids, where each one increased the spatial resolution. One of us [15][14] explored which kind of encodings, operators, constraint management and energy formulation is more suitable for an evolutionary algorithm designed to tackle minimalist models of PSP and Protein Structure Comparisons. Greenwood et al.[8] surveyed recent evolutionary approaches to the PSP. More recently Liang and Wong [17] published encouraging results on a hybrid between Monte Carlo optimization and GAs for the square HP model.

## 3 Multimeme Algorithms for Protein Structure Prediction

Memetic algorithms are evolutionary algorithms that include, as part of the "standard" evolutionary cycle of crossover-mutation-selection, a local search stage. They have been extensively studied and used on a wide range of problems. Multimeme evolutionary algorithms were introduced by Krasnogor and Smith [16] and applied to two bioinformatic problems[14]. The distinction between Memetic and Multimeme Algorithms is that the former uses only one (usually complex) local search while the later employs a *set* of local searchers. Multimeme algorithms self-adaptively select from this set which heuristic to use

for different instances, stages of the search or individuals in the population. This kind of algorithm exploits features from Evolutionary Algorithms and Variable Neighborhood Search (by virtue of its multi-operator local search).

In a Multimeme Algorithm an individual is composed of its genetic material and its memetic material. The mechanisms of genetic exchange and variation are the usual crossover and mutation operators but tailored to the specific problem one wants to solve. Memetic transmission is effected using the so called Simple Inheritance Mechanism (SIM)[16]. SIM can be formalized by:

$$L^{t,i} = \begin{cases} L^{t-1,j} & \text{if } \forall k, j \in Parents(i), k \neq j, L^{t-1,j} == L^{t-1,k} \\ L^{t-1,j} & \text{if } F(I_j^{t-1}) > F(I_k^{t-1}) \forall k, j \in Parents(i), k \neq j \\ L^{t-1,k} & \text{for any } k \in |Parents(i)| \quad \text{otherwise} \end{cases} \quad (1)$$

where a meme (local searcher)  $L$ , at time  $t - 1$  that is carried by parent  $j$  (or  $k$ ), will be transmitted to the offspring  $i$  if that meme is shared by all the parents. If they have different memes,  $L$  is associated to the fittest parent. Otherwise, when fitnesses( $F(\cdot)$ ) are comparable and memes different, a random selection is made. The rationale is to propagate local searchers (i.e. memes) that are associated with fit individuals, as those individuals were probably improved by their respective memes. During mutation, the meme of an individual can also be overridden and a local searcher assigned at random (uniformly from the set of all available local searchers) based on the value of the innovation rate parameter. This is done to introduce novelty in the local search phase of the MMA.

### 3.1 Tailoring the Multimeme Algorithm for Protein Structure Prediction

The basic evolutionary parameters and settings for the Multimeme Algorithm are now described. Tournament sizes of two and four, a crossover probability of 0.8 and a mutation probability of 0.3 were used. The runs were executed based on a (50,200), (100,400) and (500,1000) replacement strategies. Each generation of the Multimeme Algorithm consisted of a mating stage (two-point crossover with tournament selection), mutation (one and two-point mutation), local search and replacement. Every individual in the population went through an optimization period. The latter was governed by the meme held by the individual. Local search itself was restricted to three iterations in a randomized first improvement fashion, and consequently it was unconverged. For all the experiments reported in this paper the parameters were set according to the criteria described in [14] and [15] and the innovation rate was 0.2.

The memes available to the Multimeme Algorithm can be categorized as follows: pivot (rigid rotation) moves, stretching of a substructure (unfolding), random macro-mutation of a substructure, reflection of a substructure, non-local  $k$ -opt and local  $k$ -opt. These six local searchers types give rise to several different neighborhoods with which the Multimeme Algorithm will perform its search and were chosen based on previous analysis [14][15]. The Evolutionary Monte Carlo algorithm[17], which represents one of the state of the art systems for two dimensional HP lattice models, employs similar moves as mutation operators (except

for the stretch and  $k$ -opt operators). Space limitations preclude further description of the local searchers. More details are given elsewhere[14][15]. With the basic ingredients described above, the Multimeme Algorithm performed well on the standard HP model (in two and three dimensional lattices). However, it was not able to reach optimal configurations in the Functional Model Proteins. This was solved by the introduction of a *contact map memory* of current solutions in the mating strategy of the Multimeme Algorithm. With the new mating strategy we were able to solve to optimality instances of both the HP and Functional Model Proteins in two and three dimensional lattices.

### 3.2 A New Mating Strategy

As mentioned above, a contact map memory was included into the Multimeme Algorithm. During the reproduction phase of the algorithm, each generated offspring was evaluated for compatibility with the contact map memory. An offspring was *compatible* with the memory if at least  $\phi_1$  of the contacts defined by its structure were themselves compatible. In turn, a contact was *compatible* if not more than  $\phi_2$  of the individuals already in the population shared that contact. This method involves the determination of the fractions  $\phi_1$  and  $\phi_2$  which was done by *ad hoc* experimentation. In this paper  $\phi_1 = 25\%$ ,  $\phi_2 = 66\%$ . The inclusion of a memory of the contact maps of already visited solutions has as an advantage (over simply storing fitness evaluations or having an archive of genotypes, i.e. solutions ) that the contact maps abstract the geometric details of the structures and keep only the essential topological features of a two dimensional or three dimensional shape. Rotations and symmetries are filtered out and need not be explicitly considered. Given that a contact map can be realized by several different structures, the additional requisite of only accepting offspring that are compatible with the contact map memory pushes the search toward a more exploratory regime, thereby increasing diversity in the population. By holding the information of just a few contact maps in the memory the new mating strategy is actually storing information of a wide area of the whole search space. With this simple strategy we were able to improve on results previously obtained with Multimeme Algorithms[14] on the standard HP model, but more importantly, we were able to solve to optimality instances of the Functional Model Proteins that our previous algorithms were not capable of solving.

## 4 Results

In this section we will present results obtained with the Multimeme Algorithm using the new mating strategy based on the contact map memory. Functional Model Proteins were introduced in [10]. The optima for the sequences of the Functional Model Protein were obtained by an exhaustive parallel enumeration algorithm. The diamond Functional Model Protein instances and their optima are first published here. Functional Model Proteins are a challenging set of instances, as each one has a unique native state (this is not the case for other well

known minimalistic models) which is surrounded by several first excited states. Moreover, there is an energetic barrier of at least two bonds between the first excited state and the native structure. The Functional Model Proteins presented here are a subset of the available instances with known optima. We computed the native state and first excited states for all of the  $2^{23}$  sequences for the square lattice and the diamond lattice in this model. These can be obtained from [13]. The optima for the triangular lattice instances were obtained by construction in the design process of the sequences. The standard HP lattice sequences were taken from [19],[17],[9] and other references. In all experiments five runs were executed per instance. If the optimum value was not achieved by any of the five runs then we report the best sub-optimum found in bold face. The sequences and results for the *Square Lattice in the Standard HP Model* are shown in table 2. *Two Dimensional Triangular Lattice in the Standard HP Model* instances and

#	Sequence	Size	Opt.	MMA
1	HPHPPHHPHPHPHHPHPH	20	-9	-9
2	PPPHHPHPPPPPPHHHHHHHPHPPPPHHPHP	36	-14	-14
3	$H^2(PH)^4H^3PHP^3HP^3HP^4HP^3HP^3HPH^4(PH)^4H$	50	-21	-21
4	$H^{12}(PH)^2(P^2H^2)^2(PPH)^2(HP^2H)^2(P^2H^2)^2P^2(HP)^2H^{12}$	<b>64</b>	<b>-42</b>	<b>-39</b>
5	HPHPPHHPHPHPHHPHPH	20	-9	-9
6	PPHPPHHPPPPHHPPPHHPPPHH	25	-8	-8
7	$(P^2H)^2HP^2H^2P^5H^{10}P^6(H^2P^2)^2HP^2H^5$	48	-22	-22
8	PHPPHPHHHPHHPH <sup>5</sup>	18	-9	-9
9	HPHPHHHPPPHHHHPHH	18	-8	-8
10	HHPPPPPHHPPPHPPHP	18	-4	-4
11	HHHPPHHPHPPHHPHPPH	20	-10	-10

**Table 2.** Two dimensional square lattice Standard HP instances

results are displayed in table 3 below. The sequences and results for the *Square Lattice in the Functional Model Proteins* can be found in table 4. The number of energy evaluations required to achieve those optimal values is reported. The sequences and results for the *Diamond Lattice in the Functional Model Proteins* can be found in table 4(indexed with letters). To the best of our knowledge, the best algorithm for reduced models is PERM[1]. The results presented here use some of the instances and models tested in [1] and for these cases our results, the positive and the negative ones (e.g. failure to solve instance 4 of the standard, square 2D, HP model), are equivalent. Another point to note is that PERM makes assumptions about “compactness” of the native structure of a protein, which clearly do not apply to the Functional Model Proteins. Indeed, for some instances the optimal structure has one or more binding pockets. Hence, although they do not use the mentioned model, we speculate that their algorithm will not be robust in this domain. Furthermore, it is not possible to compare our algorithm directly with PERM, as our method, like the Evolutionary Monte

#	Sequence	Size	Opt.	MMA
1	HHPPHPPHPPHPPHPPH	14	-11	-11
2	HHPPHPPHPPHPPHPPH	14	-11	-11
3	HHPPHPPHPPHPPHPPH	16	-11	-11
4	HHPPHPPHPPHPPHPPH	16	-11	-11
5	HHPPHPPHPPHPPHPPH	17	-11	-11
6	HHPPHPPHPPHPPHPPH	17	-17	-17
7	HHPPHPPHPPHPPHPPH	20	-17	-17
8	HHPPHPPHPPHPPHPPH	20	-17	-17
9	HHPPHPPHPPHPPHPPH	21	-17	-17
10	HHPPHPPHPPHPPHPPH	21	-17	-17
11	HHPPHPPHPPHPPHPPH	21	-17	-17
12	HHPPHPPHPPHPPHPPH	22	-17	-17
13	HHPPHPPHPPHPPHPPH	23	-25	-25
14	HHPPHPPHPPHPPHPPH	<b>24</b>	<b>-17</b>	<b>-16</b>
15	HHPPHPPHPPHPPHPPH	24	-25	-25
16	HHPPHPPHPPHPPHPPH	24	-25	-25
17	HHPPHPPHPPHPPHPPH	<b>30</b>	<b>-25</b>	<b>-24</b>
18	HHPPHPPHPPHPPHPPH	<b>30</b>	<b>-25</b>	<b>-24</b>
19	HHPPHPPHPPHPPHPPH	<b>37</b>	<b>-29</b>	<b>-26</b>

**Table 3.** Two dimensional triangular lattice Standard HP instances

Carlo method[17], performs a blind search, whereas PERM utilizes information of the sequence being folded. Consequently, we compare our results with other blind methods and assess the robustness of the Multimeme Algorithms across a range of models and instances. Table 5 shows a representative example of the increased robustness of a Multimeme Algorithm when it is compared against a GA and a Memetic Algorithm (that uses only one type of local searcher). In the table, results for instance 15 of table 3 are displayed. The Multimeme approach achieves the optimum solution more frequently than the other approaches and also faster (the GA was given an equivalent number of energy evaluations).

Table 6 shows a comparison of the number of energy evaluations employed by our algorithm and other two well known methods (the GA and Monte Carlo reported in [19]) to solve the square lattice HP instances. Although the Evolutionary Monte Carlo method[17] finds optimum solutions for very challenging instances of the square lattice in the HP model a direct comparison with our algorithm is not possible. Liang and Wong report the number of *feasible* conformations scanned before reaching an optimal structure. However, their algorithm generates thousands of non-feasible structures during the run and this number is not provided in their paper<sup>1</sup>. Nevertheless, their algorithm does solve to optimality all the instances in table 6 and a few longer ones.

Across all the models investigated, our algorithm identifies optimal structures, regardless of considerations of compactness or the size of the protein in-

<sup>1</sup> The authors confirmed this with us in a private communication

#	Sequence	Opt.	MMA	#Evaluations
1	RHPHPHPHHHHHPHPHPHPHH	-20	-20	15170
2	RHPHPHPHHHHHPHPHPHPH	-17	-17	61940
3	HPHPHPHHHPHPHPHPHH	-16	-16	132898
4	HHHPHHHPHPHPHPHHHH	-20	-20	66774
5	RHPHPHPHPHPHPHHHPH	-17	-17	53600
6	HHHPHPHPHPHPHPHPHH	-13	-13	32619
7	RHPHPHHHHHPHPHHHHHH	-26	-26	114930
8	HPHPHPHHHPHPHPHPHH	-16	-16	28425
9	RHPHPHPHPHPHPHPHPH	-15	-15	25545
10	HPHPHPHPHPHPHPHPHH	-14	-14	111046
11	RHPHPHPHPHPHPHPHPH	-15	-15	52005
A	RHPHPHPHPHPHPHPHPH	-11	-11	123979
B	RHPHPHPHPHPHPHPHPH	-11	-11	301205
<b>C</b>	RHPHPHPHHHHHPHPHPH	<b>-14</b>	<b>-12</b>	12618
D	RHPHPHPHHHHHPHPHPH	-14	-14	1334661
E	RHPHPHPHPHPHPHHHH	-14	-14	482259
F	HPHPHPHPHPHPHPHHHH	-15	-15	332842
<b>G</b>	RHPHPHPHHHHHPHHHHHH	<b>-16</b>	<b>-14</b>	11132
H	HPHPHPHPHPHPHPHHHH	-18	-18	261027
I	HPHPHPHPHPHPHHHPH	-18	-18	550121

**Table 4.** Two dimensional square lattice Functional Model instances (indexed by numbers) and Three dimensional diamond lattice Functional Model instances (indexed by letters) and the number of energy evaluations required by the best run to achieved the optimum or a sub-optimum (in bold face).

volved. There are few cases of mis-folding, that is, only a local optimum was found. Unfortunately we were not able to detect any pattern of failure so an improvement cannot be suggested at this time. When comparing the number of energy evaluations of the Monte Carlo and our algorithm we can clearly see the benefits of the Multimeme Algorithm. If we turn to the GA then we find that for one protein (instance 7 in table 6) our approach needed considerably more evaluations.

## 5 Conclusions and Future Work

The main feature of our algorithm is that it is robust finding optimal structures, across a range of models and difficulty. This is an essential feature needed of any search method for PSP, as the precise energy formulation that must be optimized is not known. Moreover, the development of energy potentials is a very active area of research and one should expect the frequent publication of new models. A robust search mechanism, such as our Multimeme Algorithm, allows one to change the energy potential without altering too much the algorithmic infrastructure and to investigate folding prediction under the new model. The robustness of our algorithm arises from the evolutionary, population oriented,



Static Memes	Num Optima / Num Runs	Mean First Hitting Time
GA (no memes)	0/10	-
MA with Macro Mutation (r=4)	2/10	27.5
MA with Macro Mutation (r=8)	3/10	53.3
MA with Macro Mutation (r=16)	2/10	43.0
MA with Reflect (r=4)	3/10	20.6
MA with Reflect (r=8)	1/10	79.0
MA with Reflect (r=16)	1/10	45.0
MA with Stretch (r=4)	0/10	-
MA with Stretch (r=8)	0/10	-
MA with Stretch (r=16)	0/10	-
MA with Pivot	5/10	27.0
MultiMeme (all local searchers)	8/10	16.87

**Table 5.** Number of times and mean first hitting time (in generations) to achieve an optimal solution to instance 15 in table 3. Different algorithms are compared based on 10 independent runs.

#	Sequence	GA	MC	MMA
1	HPHPPHHPHPHPHPHPHPH	30492	292443	14621
2	$P(PPHH)^2PPPPH^7PPHHP^4HHPPHPP$	301339	6557189 (-13)	208233
3	$H^2(PH)^4H^3PHP^3HP^3HP^4HP^3HP^3HPH^4(PH)^4H$	592887	15151203	336763
6	PPHPPHHPPPPHHPPPPHHPPPPHH	20400	2694572	18736
7	$(P^2H)^2HP^2H^2P^5H^{10}P^6(H^2P^2)^2HP^2H^5$	126547	9201755 (-20)	1155656

**Table 6.** Energy evaluations used by the Genetic Algorithm and a Monte Carlo approach as quoted from [19] and our MMA for sequences in the HP square lattice.

nature of the search it performs and the amalgamation of several neighborhoods to further improve solutions kept in the population. All of these must be complemented with a suitable “evolutionary memory”, that in the work reported here takes the form of a contact map memory.

More experimentation will be undertaken to try to determine the reasons behind the failure to obtain optimum structures in certain sequences. Other models of the PSP will be investigated and the behaviour of our algorithm assessed there.

## References

1. U. Bastolla, H. Frauenkron, E. Gerstner, P. Grassberger, and W. Nadler. Testing a new monte carlo algorithm for protein folding. *Proteins: Structure, Function and Genetics*, 32,52, 1998.
2. B. Berger and T. Leight. Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete. In *Proceedings of The Second Annual International Conference on Computational Molecular Biology, RECOMB 98*, 1998.

3. B.P. Blackburne and J.D. Hirst. Evolution of functional model proteins. *Journal of Chemical Physics*, 115(4):1935–1942, 2001.
4. B. Derrida. Random energy model: Limit of a family of disordered models. *Physical Review Letters*, 45, 1980.
5. Ken A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24:1501, 1985.
6. A.L. Patton et al. A standard ga approach to native protein conformation prediction. In *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 574–581. Morgan Kauffman, 1995.
7. M. Feig and C.L. Brooks III. Multiscale modeling protocol for ab initio structure prediction. *In press*, 2000.
8. G.W. Greenwood, B. Lee, J. Shin, and G.B. Fogel. A survey of recent work on evolutionary approaches to the protein folding problem. In *Proceedings of the Congress of Evolutionary Computation (CEC)*, pages 488–495. IEEE, 1999.
9. W.E. Hart. Hp instances. In [http://www.cs.sandia.gov/tech\\_reports/compbio/tortilla-hp-benchmarks.html](http://www.cs.sandia.gov/tech_reports/compbio/tortilla-hp-benchmarks.html).
10. J.D. Hirst. The evolutionary landscape of functional model proteins. *Protein Engineering*, 12:721–726, 1999.
11. M. Khimasia and P. Coveney. Protein structure prediction as a hard optimization problem: The genetic algorithm approach. In *Molecular Simulation*, volume 19, pages 205–226, 1997.
12. A. Kolinski, M.R. Betancourt, D. Kihara, P. Rotkiewicz, and J. Skolnick. Generalized comparative modeling (genecomp): A combination of sequence comparison, threading, and lattice modeling for protein structure prediction and refinement. *PROTEINS: Structure, Function, and Genetics*, 44:133–149, 2001.
13. N. Krasnogor. Standard hp and functional model proteins instances. <http://dirac.chem.nott.ac.uk/~natk/Public/index.html>, 2001.
14. N. Krasnogor. *Studies on the Theory and Design Space of Memetic Algorithms*. Ph.D. Thesis, Faculty of Computing, Mathematics and Engineering, University of the West of England, Bristol, United Kingdom, 2002.
15. N. Krasnogor, W.E. Hart, J. Smith, and D.A. Pelta. Protein structure prediction with evolutionary algorithms. In W. Banzhaf, J. Daida, A.E. Eiben, M.H. Garzon, V. Honavar, M. Jakaiela, and R.E. Smith, editors, *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*. Morgan Kaufman, 1999.
16. N. Krasnogor and J.E. Smith. Emergence of profitable search strategies based on a simple inheritance mechanism. In *Proceedings of the 2001 Genetic and Evolutionary Computation Conference*. Morgan Kaufmann, 2001.
17. F. Liang and W.H. Wong. Evolutionary monte carlo for protein folding simulations. *Journal of Chemical Physics*, 115(7):3374–3380, 2001.
18. P. Montanari, A. Colosimo, and P. Sirabella. The application of a genetic algorithm to the protein folding problem. In *Proceedings of Engineering of Intelligent Systems (EIS 98)*, 1998.
19. R. Unger and J. Moult. Genetic algorithms for protein folding simulations. *Journal of Molecular Biology*, 231(1):75–81, 1993.
20. Y. Xia, E.S. Huang, M. Levitt, and R. Samudrala. Ab initio construction of protein tertiary structures using a hierarchical approach. *Journal of Molecular Biology*, 300:171–185, 2000.
21. A. Zemla, C. Venclovas, J. Moult, and K. Fidelis. Processing and analysis of casp3 protein structure predictions. *PROTEINS: Structure, Function, and Genetics Suppl*, 3:22–29, 1999.