2011 IEEE International Conference on Bioinformatics and Biomedicine

Populating Local Minima in the Protein Conformational Space

Brian Olson[1] and Amarda Shehu[1,2*]
[1]*Department of Computer Science*
[2]*Department of Bioinformatics and Computational Biology*
George Mason University
Fairfax, VA, 22030, United States
bolson3@gmu.edu, amarda@gmu.edu
∗Corresponding Author

***Abstract*—Protein Modeling conceptualizes the protein energy landscape as a funnel with the native structure at the low-energy minimum. Current protein structure prediction algorithms seek the global minimum by searching for low-energy conformations in the hope that some of these reside in local minima near the native structure. The search techniques employed, however, fail to explicitly model these local minima. This work proposes a memetic algorithm which combines methods from evolutionary computation with cutting-edge structure prediction protocols. The Protein Local Optima Walk (PLOW) algorithm proposed here explores the space of local minima by explicitly projecting each move in the conformation space to a nearby local minimum. This allows PLOW to jump over local energy barriers and more effectively sample near-native conformations. Analysis across a broad range of proteins shows that PLOW outperforms an MMC-based method and compares favorably against other published *ab-inito* structure prediction algorithms.**

***Keywords*-protein native structure; near-native conformations; local minimum; iterated local search; fragment-based assembly.**

I. Introduction

The determination of a protein's three-dimensional structure from sequence alone remains a central challenge in computational structural biology. The Anfinsen experiments showed that this "native" structure is encoded in the amino-acid sequence [1]. Development of a computational approach to discover a protein's native structure will not only elucidate the function of existing proteins, but will also aid the development of synthetically engineered proteins, improve our models of protein ligand docking for drug development, and assist the prediction of protein-protein interactions in supramolecular assemblies [2], [3].

A protein conformation can be represented by the dihedral bond angles which determine the spatial arrangement of its atoms. This leads to a vast high-dimensional search space with a funnel-like energy surface with the native structure at the low-energy basin [4]. The energy functions available to traverse this landscape are semi-empirical, and their inaccuracies result in a rugged energy surface. An emerging template in protein structure prediction employs a two-stage process to tackle this vast and rugged energy surface [2], [5]–[7]. Stage one is a coarse-grained search for a diverse set of local minima. Stage two is the refinement of these local minima at all-atom detail.

This work focuses on coarse-grained representations of the conformational space. These simplified representations are attractive as they can be further refined at all-atom detail in later stages. Existing coarse-grained methods, however, fail to explicitly sample local minima. A common approach is to launch many Metropolis Monte Carlo (MMC) or molecular dynamics trajectories to obtain a large number of low-energy conformations. A post-processing analysis of these conformations often groups them by geometric similarity in order to locate local minima though clustering. Furthermore, analysis shows that promising conformations are frequently discarded during the clustering phase. FeLTr, a new search framework recently introduced by our lab, incorporates this geometric analysis into the coarse-grained search process itself, but also fails to explicitly model local minima [8], [9].

This paper introduces a novel algorithm, Protein Local Optima Walk (PLOW), to explicitly populate local minima in the energy surface. Our algorithm, inspired by the Iterated Local Search (ILS) framework in evolutionary computation [10], combines a global search method with an exploitative local search. The global search allows the algorithm to explore the breadth of the energy surface, biasing towards lower-energy regions, while the local search steers each exploration at the global level to the closest low-energy local minimum. PLOW essentially projects the protein conformational space onto the space of local minima. The result is more effective at sampling a wide range of near-native conformations. Section III benchmarks PLOW on 13 diverse proteins, comparing it both to our previously published FeLTr framework, as well as published results from the Sosnick and Baker research groups [6], [11].

Our previous work on the FeLTr framework seeks to ensure a more geometrically-diverse conformational sampling by employing a geometric projection layer [8], [9]. The algorithm grows a search tree in the conformational space by expanding selected conformations with short fixed-

978-0-7695-4574-5/11 $26.00 © 2011 IEEE
DOI 10.1109/BIBM.2011.22

length MMC trajectories and maintains a representative ensemble of previously visited conformations in memory. Selection from this ensemble is biased towards low-energy conformations and regions in under explored areas of the conformational space. In this way, FeLTr is able to dynamically redirect computational resources at the global level to ensure a degree of geometric diversity in its conformational sampling. Results show that the FeLTr framework samples near-native conformations more effectively than MMC-based methods [8], [9], [12], [13]. However, like other coarse-grained sampling methods, FeLTr does not explicitly sample local minima, but rather relies on clustering analysis to filter its results down to a subset of conformations which will hopefully correspond to local minima. The results presented here show that the approach taken by PLOW outperforms that of FeLTr.

II. METHODS

A. Representation and Energy Function

PLOW uses a coarse-grained representation, modeling each conformation as a vector of $2n$ dihedral bond angles, where n is the number of amino acids in the protein. Modification of a conformation to obtain a new one is performed using fragment-based assembly. Details of the representation, fragment library, and energy function can be found in previous work [8], [9].

B. Canonical Iterated Local Search (ILS)

PLOW employs a two layer search process to explore the space of local minima. The outer layer (see Algorithm 1) simulates an MMC search at the global level, while the inner layer performs a greedy local search to project each point found in the outer layer onto a nearby local minimum.

The PERTURBATION function jumps PLOW out of its current local minimum, H, to a nearby region of space, H_{new}, allowing it to easily overcome local energy barriers (Algo. 1, line 5). H_{new} is then projected onto its nearest local minimum via the LOCALSEARCH function (Algo. 1, line 6). Finally, the algorithm uses the ACCEPTANCECRITERION function to decide whether to keep its home base at H or move it to H_{new} (Algo. 1, line 7).

The starting point of the search is determined by the INITIALSELECTION function (Algo. 1, line 3). The process of local minima hopping is repeated until a specified number of energy function evaluations has occurred (Algo. 1, line 1). Since, each instance of LOCALSEARCH runs for a variable length, the number of evaluations, $Eval_{count}$, is incremented within LOCALSEARCH (Algo. 1, line 6).

PLOW adapts this general framework into an algorithm suitable for navigation of the protein conformational landscape. INITIALSELECTION initializes H as a fully extended conformation projected onto a nearby local minimum using the LOCALSEARCH function. ACCEPTANCECRITERION uses the Metropolis Criterion to decide whether or not to move its home base to H_{new} or remain at the current value of H. LOCALSEARCH and PERTURBATION are defined in sections II-C and II-D, respectively.

Algo. 1 The canonical Iterated Local Search (ILS) framework is shown. This work defines domain-specific implementations of INITIALSELECTION, LOCALSEARCH, PERTURBATION, and ACCEPTANCECRITERION.

Input: Max number of energy function evaluations

1: $\text{Eval}_{max} \leftarrow$ *(UserDefined)*
2: $\text{Eval}_{count} \leftarrow 0$
3: H $\leftarrow$ INITIALSELECTION()
4: **while** $\text{Eval}_{count} < \text{Eval}_{max}$ **do**
5: $\quad \text{H}_{new} \leftarrow$ PERTURBATION(H)
6: $\quad \text{H}_{new}$, $\text{Eval}_{count} \leftarrow$ LOCALSEARCH(H_{new}, Eval_{count})
7: $\quad$ H $\leftarrow$ ACCEPTANCECRITERION(H,H_{new})

C. Local Search

The LOCALSEARCH function in Algorithm 1 projects a conformation onto a nearby local minimum using a greedy local search incorporating fragment-based assembly and the coarse-grained energy function referenced in section II-A. At each iteration, the local search generates a child conformation by performing a single configuration replacement. If the energy of the child conformation is lower than that of its parent, then the child conformation replaces its parent, otherwise the child is discarded. This process is repeated until k children in a row have been discarded, indicating the presence of a local minimum. When this occurs, LOCALSEARCH stops and returns its current conformation, signaling that a local minimum has been fully explored. The value of k is set to the length of the target protein.

D. Perturbation

PERTURBATION modifies a conformation just enough to get out of its current local minimum, such that the LOCALSEARCH function is unlikely to return it to the same local minimum. However, if PERTURBATION makes too drastic a change, then the search is unable to benefit from knowledge of the previous local minimum. Our experiments find that a single random trimer fragment replacement as the PERTURBATION function accomplishes this goal.

Low-energy conformations tend to be compact and leave little room for movement in their backbone chain without raising their energy. Even a single fragment replacement to a structure that is already at a local minimum is likely to cause significant disruption in a conformation and thus greatly increase its energy. The perturbed conformation shares nearly all of its local structural features with its parent, but the new conformation has a much higher energy and may have a significantly altered overall global structure. Given a high energy, the LOCALSEARCH function will be able to easily optimize the perturbed conformation to one of

many distinct local minima, leaving little chance that it will return to its previous local minimum.

III. Results

We apply the PLOW algorithm described in section II to the 13 target proteins listed in Table I. These proteins range in size from 61 to 123 amino acids in length and represent a diverse set of α and β fold topologies. Section III-A briefly outlines our experimental procedure. In section III-B, results obtained from PLOW are compared with results from our previously published FeLTr framework. Section III-C modifies the local search depth of our FeLTr framework in order to provide a more fair comparison between PLOW and FeLTr. Finally, section III-D compares the results from PLOW to those published by the Sosnick [6] and Baker [11] research groups.

A. Experiments and Measurements

All experiments are run with a fixed budget of N energy function evaluations. Over 90% of the CPU time for both PLOW and FeLTr is spent computing potential energies. The cost of an energy function evaluation is directly related to protein length. Therefore, holding the number of evaluations constant (rather than CPU time) ensures a fair comparison between both methods across a range of protein lengths. N is set to 10,000,000 in order to give both methods an exhaustive run, requiring two to four days of CPU time on a 2.66 GHz Opteron processor with 8 GB of memory.

We compare conformations generated by each search method to the native structure obtained from the Protein Data Bank (PDB). Results are compared using least Root Mean Square Deviation (lRMSD), a measure of the average distance between the C_α atoms of two aligned conformations. In PLOW, the lRMSD from the native structure is measured for each local minimum discovered during the search. For FeLTr, the lRMSD is measured for each vertex created in its search tree.

B. Comparison to Previous Work

We compare the results obtained by PLOW to the results obtained by FeLTr [8]. Table I shows that PLOW outperforms FeLTr by more than 0.5Å lRMSD in all but four cases: 1aoy, 1c8cA, 1fwp, and 1hz6A. PLOW significantly outperforms FeLTr by at least 1.5Å in key cases, including longer proteins (2ezk, 2h5nd, and 3gwl). On 1ail PLOW finds structures below 3Å lRMSD to native. This suggests that PLOW's locating of local minima is able to more effectively locate near-native conformations. Furthermore, if used as the first stage in a two stage structure prediction framework, PLOW results in several fold fewer conformations which must be further refined at the all-atom level of detail.

C. Local Search Depth

The FeLTr framework uses an approach similar to PLOW in that it performs a global search on top of many short local searches. In FeLTr, branches of the search tree are fixed-length MMC trajectories. In PLOW, the lengths of the inner local searches are dynamically selected and tend to be several times longer, on average, than the fixed length MMC trajectories used in FeLTr. To rule out the possibility that PLOW is merely benefiting from longer local searches, we conducted an additional experiment to fairly compare FeLTr and PLOW. FeLTr-Ext runs FeLTr with the length of the inner MMC trajectory extended to the average PLOW search length (Table I, column 4). Table I shows that FeLTr-Ext performs slightly better, on average, than standard FeLTr and is comparable to PLOW in a few cases: 1c8cA, 1dtdB, 1hz6A, 1isuA, and 1wapA. However, on average, PLOW still outperforms FeLTr-Ext, especially in the case of the three longer proteins. This suggests that there is a distinct advantage to the local optimization approach employed in PLOW. While the average length of the local search portion is the same between both methods, PLOW is able to vary this length as necessary to fully explore its current local minimum while not wasting resources once a minimum has been reached.

D. Comparison to other state-of-the-art methods

We compare the results of 11 of the 13 target proteins against published results from the Sosnick and Baker research groups [6], [11]. PLOW outperforms the other groups by more than 0.5Å in five cases (1ail, 1cc5, 1fwp, 1wapa, and 2ezk), which include all three fold topologies and the longest of the 11 proteins. Of the remaining six proteins, PLOW only performs worse in four of the cases: 1c8cA, 1dtdB, 1hz6A, and 1sap. This result is not unexpected, as the different energy functions used by both groups, as well as a very different sampling technique employed by the Sosnick group, can easily account for the performance difference on specific proteins.

IV. Conclusion

The Protein Local Optima Walk (PLOW) algorithm proposed here is a novel *ab-initio* structure prediction algorithm for effectively sampling local minima in the protein energy landscape. The algorithm works by effectively projecting the search space onto the sub-space of local energy minima. By traversing only these local minima, PLOW is able to more effectively sample near-native conformations which are candidates for all-atom refinement in further studies. PLOW outperforms our previous work [8] on a diverse set of target proteins and performs favorably when compared to published results from two other research groups [6], [11].

Many studies have demonstrated the effectiveness of an evolutionary framework to optimize intermediate conformations [14], [15]. However, these studies use overly simplified

Table I

THE LOWEST LRMSD TO THE NATIVE STRUCTURE IS SHOWN FOR PLOW AND OUR PREVIOUSLY DEVELOPED FELTR FRAMEWORK AS WELL AS PUBLISHED RESULTS FROM THE SOSNICK [6] AND BAKER [11] RESEARCH GROUPS. THE LRMSDS SHOWN ARE AVERAGED OVER THREE RUNS, WITH THE MINIMUM OF THE THREE RUNS SHOWN IN PARENTHESES. COLUMN 4 SHOWS THE AVERAGE NUMBER OF ITERATIONS IN EACH PLOW LOCALSEARCH. COLUMN 6 (FELTR-EXT) REPRESENTS FELTR USING THE VALUE FROM COLUMN 4 AS ITS MMC SEARCH LENGTH.

				avg local	avg (min) lowest lRMSD to native in Å				
	PDB ID	length	fold	search length	PLOW	FeLTr-Ext	FeLTr	Sosnick	Baker
1	1ail	70	α/β	237	2.7(2.3)	4.0(3.4)	4.7(4.7)	5.4	6.0
2	1aoy	78	α/β	258	5.4(5.2)	5.9(5.2)	5.1(4.6)	5.7	5.7
3	1c8cA	64	α/β	199	7.0(6.8)	7.1(5.8)	6.8(6.0)	3.7	5.0
4	1cc5	76	α	274	5.5(5.1)	6.0(4.9)	6.7(6.7)	6.5	6.2
5	1dtdB	61	α/β	160	7.1(6.9)	7.5(7.0)	7.7(7.6)	6.5	5.7
6	1fwp	69	α/β	210	6.5(6.3)	7.2(6.8)	6.8(6.4)	8.1	7.3
7	1hz6A	67	α/β	182	6.4(6.3)	6.5(6.1)	6.7(6.6)	3.8	3.4
8	1isuA	62	α/β	173	6.3(6.0)	6.4(5.7)	6.8(6.7)	6.5	6.9
9	1sap	66	α/β	211	6.5(6.0)	7.2(6.8)	7.1(6.5)	4.6	6.6
10	1wapA	68	β	199	7.2(6.7)	7.4(6.5)	7.8(7.3)	8.0	7.7
11	2ezk	93	α	293	4.6(4.2)	5.9(4.7)	6.4(6.0)	5.5	6.6
12	2h5nD	123	α	482	7.0(6.1)	8.8(8.3)	9.0(8.5)	NA	NA
13	3gwl	106	α	375	5.1(4.7)	6.7(6.4)	6.7(6.0)	NA	NA

models, focus solely on optimization of an objective function, and fail to compare results with experimentally determined structures. Here we combine cutting-edge stochastic optimization strategies from evolutionary computation with established procedures for assembly of coarse-grained structures and analysis of results. Our results show this approach offers improved sampling at the coarse-grained level, the results of which may be further refined by additional studies.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation Grant No. 1016995.

REFERENCES

[1] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, no. 4096, pp. 223–230, 1973.

[2] P. Bradley, K. M. S. Misura, and D. Baker, "Toward high-resolution de novo structure prediction for small proteins," *Science*, vol. 309, no. 5742, pp. 1868–1871, 2005.

[3] T. Kortemme and D. Baker, "Computational design of protein-protein interactions," *Curr. Opinion Struct. Biol.*, vol. 8, no. 1, pp. 91–97, 2004.

[4] K. A. Dill and H. S. Chan, "From levinthal to pathways to funnels," *Nat. Struct. Biol.*, vol. 4, no. 1, pp. 10–19, 1997.

[5] A. Shehu, L. E. Kavraki, and C. Clementi, "Multiscale characterization of protein conformational ensembles," *Proteins: Struct. Funct. Bioinf.*, vol. 76, no. 4, pp. 837–851, 2009.

[6] J. DeBartolo, A. Colubri, A. K. Jha, J. E. Fitzgerald, K. F. Freed, and T. R. Sosnick, "Mimicking the folding pathway to improve homology-free protein structure prediction," *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 10, pp. 3734–3739, 2009.

[7] A. Shehu, L. E. Kavraki, and C. Clementi, "Unfolding the fold of cyclic cysteine-rich peptides," *Protein Sci.*, vol. 17, no. 3, pp. 482–493, 2008.

[8] B. Olson, K. Molloy, and A. Shehu, "In search of the protein native state with a probabilistic sampling approach," *J. Bioinf. and Comp. Biol.*, vol. 9, no. 3, pp. 383–398, 2011.

[9] A. Shehu and B. Olson, "Guiding the search for native-like protein conformations with an ab-initio tree-based exploration," *Int. J. Robot. Res.*, vol. 29, no. 8, pp. 1106–11 227, 2010.

[10] S. Luke, *Essentials of Metaheuristics*. Lulu, 2009, available for free at http://cs.gmu.edu/∼sean/book/metaheuristics/.

[11] J. Meiler and D. Baker, "Coupled prediction of protein secondary and tertiary structure," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 12 105–12 110, 2003.

[12] B. Olson, K. Molloy, and A. Shehu, "Enhancing sampling of the conformational space near the protein native state," in *BIONETICS: Intl. Conf. on Bio-inspired Models of Network, Information, and Computing Systems*, Boston, MA, December 2010.

[13] A. Shehu, "An ab-initio tree-based exploration to enhance sampling of low-energy protein conformations," Seattle, WA, USA, 2009.

[14] M. Islam and M. Chetty, *Novel Memetic Algorithm for Protein Structure Prediction*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009, vol. 5866, pp. 412–421.

[15] J. Keum, E.S., K. Kim, and E. Santos, "Local minima-based exploration for off-lattice protein folding," in *Bioinformatics Conference, 2003. CSB 2003. Proceedings of the 2003 IEEE*, aug. 2003, pp. 615 – 616.