Efficient Basin Hopping in the Protein Energy Surface

Brian Olson[1] and Amarda Shehu[1,2,3*]
[1]*Department of Computer Science*
[2]*Department of Bioinformatics and Computational Biology*
[3]*Department of Bioengineering*
George Mason University
Fairfax, VA, 22030, United States
bolson3@gmu.edu, amarda@gmu.edu
**Corresponding Author*

***Abstract*—The vast and rugged protein energy surface can be effectively represented in terms of local minima. The basin-hopping framework, where a structural perturbation is followed by an energy minimization, is particularly suited to obtaining this coarse-grained representation. Basin hopping is effective for small systems both in locating lower-energy minima and obtaining conformations near the native structure. The efficiency decreases for large systems. Our recent work improves efficiency on large systems through molecular fragment replacement. In this paper, we conduct a detailed investigation of two components in basin hopping, perturbation and minimization, and how they work in concert to affect the sampling of near-native local minima. We show that controlling the magnitude of perturbation jumps is related to the ability to effectively steer the exploration towards conformations near the protein native state. In minimization, we show that a simple greedy search is just as effective as Metropolis Monte Carlo-based minimization. Finally, we show that an evolutionary-inspired approach based on the Pareto front is particularly effective in reducing the ensemble of sampled local minima and obtains a simpler representation of the probed energy surface.**

***Keywords*-basin hopping; local minima; greedy local search; fragment-based assembly; protein native state.**

I. Introduction

A multi-dimensional funnel-like energy surface underlies the protein conformational space [1]. The size and ruggedness of this surface are the primary reasons why obtaining structural information on the biologically-active (native) state of a protein based on the amino-acid sequence alone is an outstanding challenge in computational structural biology [2]. Despite these challenges, computational research is needed to close the growing gap between the wealth of protein sequence data and the scarce information on native structures. Obtaining structural information de novo promises to elucidate the structure-function relationship and advance structure-driven studies and applications [3]–[5].

Methods based on Molecular Dynamics (MD) that simulate the folding kinetics demand significant computational resources. Sacrificing information on folding kinetics and conducting instead global (energy) optimization is useful, because the protein native state is related to the lowest energies of the potential energy surface [1]. Optimization-based approaches can obtain native conformations orders of magnitude faster than approaches that simulate folding pathways [6]. The complexity of the protein energy surface still presents a significant challenge, especially on medium-size proteins [2]. For this reason, research into development and analysis of stochastic optimization algorithms for conformational search abounds [7].

A unifying strategy among many stochastic optimization techniques for de novo protein structure prediction is the sampling of a large number of low-energy conformations. Predominantly, the conformations are end points of many independent MD or Monte Carlo (MC) trajectories locally optimizing some chosen coarse-grained energy function. In complete studies, the obtained conformations are grouped by structural similarity to reveal local minima that are worth optimizing further through some finer-grained energy function in order to improve proximity to the native state [3], [8], [9], [9]–[12]. Alternatively, the local (trajectory) searches can be integrated in a tree to better control the exploration and even use online analysis to bias the tree away from high-energy oversampled regions [13], [14].

While successful on many small-to-medium proteins, current approaches are bound by the accuracy of the employed energy function. Many studies analyze the inherent errors due to approximations in state-of-the-art energy functions [15]. These errors are responsible for deviations between the reported global minimum of an energy function and the experimentally-determined structure. Some studies report the deviations can vary between 2-4Å [6]. In this context, approaches that aim to obtain a broad view of different low-energy local minima are more appropriate, particularly if they are to be followed by detailed heavy-duty optimization techniques on select minima.

Basin hopping (BH) is a suitable framework to sample relevant local minima in the protein energy surface. It was originally introduced to obtain the Lennard-Jones minima of small atomic clusters [16]. Procedurally, the framework consists of repeated applications of a structural perturbation

978-1-4673-2560-8/12/$31.00 ©2012 IEEE

followed by an energy minimization. A Metropolis-like criterion is often employed to bias the sampling of local minima towards lower energy ones. The result is a trajectory of consecutively-sampled local minima in the energy surface. The appeal of the BH framework is that it transforms the energy surface into a collection of interpenetrating staircases. A succinct and (discrete) coarse-grained representation is obtained for the energy surface in terms of local minima. BH algorithms essentially differ in how they implement perturbation and minimization. Perturbation predominantly modifies atomic coordinates, and minimization is either a gradient descent or a Metropolis MC at low temperature.

Recently, the BH framework has gained new attention for protein structure prediction [17]–[19]. In [17], the perturbation changes cartesian coordinates by values sampled uniformly at random over a small range. The minimization is implemented through a gradient descent of a selected coarse-grained energy function. The resulting BH algorithm succeeds in locating both lower-energy mimina and conformations closer to the experimentally-determined native structure than MD with Simulated Annealing on small proteins. On sequences longer than 75 amino acids, the efficiency decreases [17].

Recent work in [18], [19] addresses this issue and extends the applicability of the BH framework to longer protein sequences by incorporating fragment-based assembly. Both the perturbation and minimization employ configuration replacements of trimers selected at random over the protein sequence. Application of the resulting BH algorithm shows that the obtained proximity to the known native structure is similar to that reported by many state-of-the-art structure prediction protocols. It is worth noting that the BH algorithm in [19] employs a coarse-grained energy function and is intended to be the first step in a structure prediction protocol that then further refines select minima.

Given the newly-gained attention and promise of the BH framework for structure prediction, it is important to obtain more understanding and assess the components and efficiency of this framework. We conduct a detailed investigation of the two main components, perturbation and minimization and analyze how they work in concert in the BH framework. We show that controlling the magnitude of jumps in conformational space due to perturbation allows directly controlling the distance between consecutively-sampled local minima. We show in turn that this distance is related to the ability to effectively steer the exploration towards near-native conformations. We also show that a greedy search in minimization is just as effective as Metropolis MC-based minimization. We demonstrate that an evolutionary-inspired approach based on the Pareto front is particularly effective in reducing the ensemble of sampled local minima and so obtaining a simpler representation of the probed energy surface.

II. Methods

The BH framework is summarized first, followed by descriptions of different implementations for the perturbation and minimization components. The Pareto-front filtering of sampled local minima is described last.

A. Basin Hopping Framework

This paper builds on the BH algorithm proposed in [19]. The algorithm hops between consecutive minima C_i and C_{i+1} through an intermediate $C_{\mathrm{perturb,i}}$ conformation. A perturbation modifies C_i to obtain a higher-energy conformation $C_{\mathrm{perturb,i}}$ that allows escaping the current minimum. A minimization conducts a series of modifications starting from $C_{\mathrm{perturb,i}}$ to reach a new minimum C_{i+1}. C_{i+1} is added as a new minimum in the trajectory according to the Metropolis criterion based on the energetic difference between C_i and C_{i+1}.

Perturbation and minimization modify conformations through fragment replacement. Three consecutive amino acids are selected at random from a given conformation. A new configuration for this trimer is then obtained at random from a pre-compiled library of fragments extracted from known native structures [19]. While perturbation replaces one trimer, minimization consists of repeated replacements until k consecutive attempts fail to lower energy.

Energy is measured through the Associate Memory hamiltonian with Water (AMW) which sums 6 terms: $Energy_{AMW} = E_{Lennard-Jones} + E_{H-Bond} + E_{contact} + E_{burial} + E_{water} + E_{Rg}$. AMW is inspired by physical models of the protein folding process, namely the principle of minimal frustration, and uses nonadditive water-mediated interactions [20].

B. Perturbation

The magnitude of the jump provided by the perturbation needs to be large enough to escape the current minimum (so the following minimization does not bring the trajectory back to it), but also not so large that consecutive minima are unrelated (in terms of proximity in the conformational space). If the magnitude is too small, the BH search is inefficient. If the magnitude is too large, the search effectively resorts to minimizations of conformations sampled at random and the Metropolis criterion does not provide the intended energy bias. In this paper, we quantify how the perturbation magnitude controls the distance between consecutive minima, and whether this control has any bearing on the BH sampling of near-native conformations.

The following technique controls the magnitude of each perturbation jump to a configured distance D, measured as the least Root Mean Square Deviation (lRMSD) between C_i and $C_{\mathrm{perturb,i}}$. A target distance d is sampled from a Gaussian distribution centered at D with a standard deviation of 1. A new perturbed conformation C_{perturb} is sampled using a single trimer configuration replacement. C_{perturb} is

accepted if the lRMSD between C_i and C_{perturb} is within a tolerance, t, of the target distance d. The process is repeated for a maximum n number of attempts or until a C_{perturb} that satisfies the lRSMD criterion is obtained. If not, the ensuing minimization uses as $C_{\mathrm{perturb,i}}$ the C_{perturb} conformation with the lRMSD from C_i closest to d over all n ones obtained in this process. The value of n is set to 20, which is large enough to find an accepted C_{perturb} within a tolerance $t = 0.5$Å in most cases. Since candidates for $C_{\mathrm{perturb,i}}$ are not evaluated for energy, this process adds insignificant additional computation to the overall BH search.

C. Minimization

The two main alternatives we study for the minimization component are the greedy search implemented originally in [19] and Metropolis MC (MMC) trajectories of different effective temperatures.

Greedy search insists on lowering the energy after every modification. An MMC search instead can cross over energetic barriers whose height is controlled through the effective temperature (T) in the Metropolis criterion. Employing a small non-zero T allows MMC to jump over low barriers and possibly probe lower-energy levels than a strictly downhill greedy search. The MMC trajectory continues until k consecutive trimer fragment replacements have been rejected (k is the number of amino acids in the sequence).

Finding true local minima in the energy surface is computationally intensive and previous work shows that the native structure lies somewhere above the true global minimum [19]. This working definition of a local minimum in terms of k is sufficient to reach near-native conformations.

Controlling the effective temperature controls the height of the barriers crossed by the MMC search. The greedy search, employed in previous work, is a special case where the effective temperature is set to 0. In section III, three different effective temperatures are studied for the MMC search. A very low one, T_0, corresponds to accepting a 1.4 kcal/mol energy increase with probability 0.1, and two slightly higher ones, T_1 and T_2, respectively, accept energy increases of 1.7 and 2.6 kcal/mol with probability 0.1.

D. Pareto Front Ensemble Filtering

The Ω conformational ensemble of sampled local minima can be large; Ω needs to be reduced in order to provide a few relevant minima for further refinement in the context of a complete structure prediction protocol. The reduction necessitates a trade-off between selecting a small number of conformations and selecting a sample diverse enough so that near-native conformations are not discarded.

Selecting all conformations below an energy threshold is problematic. First, there is no consistent technique for selecting an appropriate energy threshold. Second, it is likely that a threshold will either include a large portion of the ensemble, making fine-grained refinement prohibitive, or exclude many near-native conformations (recall that the native structure may deviate from the global energy minimum).

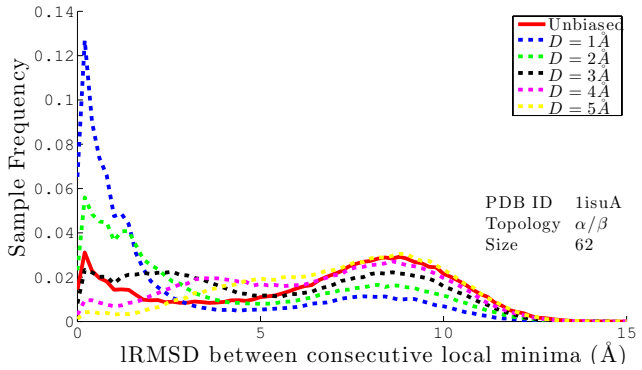
The error resulting from the weighted linear combination of energy terms in current energy functions [21] can be avoided by conducting a more nuanced, multi-objective, energetic comparison that considers energy terms individually. A conformation C_i is said to dominate a conformation C_j when every energy term in C_i is lower than the corresponding term in C_j. If there is no conformation in Ω that dominates C_j, then C_j is said to be non-dominated. Conformations in the non-dominated ensemble, referred to as the Pareto front, are considered equivalent with respect to a multi-objective analysis. This work filters the Ω ensemble of conformations representing local minima by retaining only those corresponding to the Pareto front as Ω_P.

III. Results

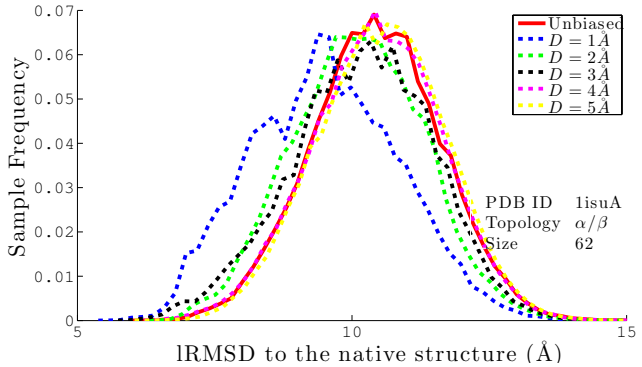
Analysis is conducted over 15 target protein systems listed in Table I which range from 61-123 amino acids in length and cover α, β, and α/β folds. Experiments are run for a fixed budget of $10,000,000$ energy function evaluations. Since over 90% of CPU time is spent on such evaluations, the limit ensures a fair comparison between different parameter selections on diverse proteins. In practice this takes 1-4 days of CPU time on a 2.4Ghz Core i7 processor, depending on protein length. The perturbation and minimization components are analyzed first, in sections III-A and III-B, respectively. The multi-objective analysis to filter the Ω ensemble of sampled minima is detailed last in section III-C.

A. Perturbation Distance

Our previous work shows a direct correlation between the mean lRMSD between consecutive local minima (referred to from now on as $\mu_{|MM|}$) and the ability of the BH framework to sample near-native conformations [19]. Analysis shows that $\mu_{|MM|}$ can be effectively controlled by biasing the magnitude of the perturbation jump through a target perturbation distance D; as D is increased, there is a corresponding increase in $\mu_{|MM|}$ (data not shown here). Tuning D does not have any significant effect on the single lowest lRMSD obtained (lRMSD is computed over the heavy backbone atoms and measures the proximity of a conformation to the experimental native structure). However, D affects the frequency with which near-native conformations are obtained (that is, the distribution of sampled minima) in cases where unbiased perturbation results in large $\mu_{|MM|}$ values. Figure 1 illustrates this by plotting, for different values of D, the distribution of $\mu_{|MM|}$ values and the resulting distribution of lRMSD values. These results show that there is a distinct advantage to biasing the perturbation distance to $D = 1$Å or $D = 2$Å. Figure 1a shows that the frequency of small $\mu_{|MM|}$ is larger when $D \in \{1, 2\}$ Å vs. an unbiased perturbation. Figure 1b then shows a corresponding increase in low-lRMSD conformations generated with $D \in \{1, 2\}$ Å.



(a) 1isuA



(b) 1isuA

Figure 1: The frequencies of $\mu_{|MM|}$ sampled during the search and lRMSDs to the native structure for the protein with native structure PDB ID 1isuA are shown in (a) and (b), respectively. The solid red line represents BH employing the unbiased perturbation method. The dashed lines represent BH with median perturbation distances $D = 1$Å to $D = 5$Å.

The effect of controlling D shown in Figure 1 is strongest on more heavily β-sheet proteins (native PDB IDs 1dtdB, 1isuA, 1wapA, and 1hhp). On these proteins, an unbiased perturbation results in few small consecutive local minima distances. More near-native conformations are also obtained (though to a lesser extent) when $D \in \{1, 2\}$ for some proteins where unbiased perturbation results in larger numbers of small consecutive local minima distances (native PDB IDs 1ail, 1sap, and 2h5nD). These proteins still benefit from enhanced sampling of neighboring local minima.

B. MMC vs. Greedy Search in Minimization

Table I compares the greedy search ($T = 0$) to MMC searches with T_0, T_1, and T_2. Columns 7-10 show the lowest energy achieved under each setting. Three observations can be made: (i) Lower energies are obtained by MMC than the greedy search. (ii) Overall, on proteins with less than 80 amino acids, the lowest energy is achieved by MMC with T_0. (iii) On longer proteins, the slightly higher T_1 achieves lower energies, possibly because in more complex rugged surfaces, small uphill moves allow reaching deeper minima.

Table II: Column 3 shows the reduction in the size of Ω_P as a percentage of the size of the original ensemble Ω. Columns 4-6 show the lowest lRMSD to the native structure of any conformation in Ω, Ω_E and Ω_P, respectively.

	Native PDB ID	Reduction (%) Ω_P	min lRMSD (Å) Ω	Ω_E	Ω_P
1	1dtdB	97	7.2	7.9	7.7
2	1isuA	94	6.0	6.3	6.5
3	1c8cA	97	7.4	7.5	7.5
4	1sap	98	6.5	7.6	7.5
5	1hz6A	98	5.9	6.7	6.5
6	1wapA	98	7.7	8.7	8.7
7	1fwp	94	6.4	8.1	7.4
8	1ail	98	3.4	6.8	5.9
9	1aoy	96	5.7	6.9	6.6
10	1cc5	94	5.6	8.6	7.0
11	2ezk	97	4.4	8.0	7.3
12	1hhp	99	10.7	12.0	12.0
13	2hg6	95	8.6	11.5	10.5
14	3gwl	96	4.2	4.7	5.4
15	2h5nD	94	7.9	10.7	10.0

Columns 11-14 in Table I show, for each value of T, the lowest lRMSD to the native structure over Ω. Comparable lowest lRMSDs are obtained whether greedy or MMC search is employed in the minimization. Probing deeper into minima in the MMC-based minimization does not necessarily bring the BH search closer to the native structure.

MMC-based minimization is costly, resulting in longer minimizations and fewer sampled minima (total number of energy evaluations is fixed). Employing MMC over greedy search thus shortens the BH trajectories by 50 to 70% in terms of the number of sampled minima. Columns 11-14 in Table I show that a lower number of sampled minima does not necessarily correlate with worse proximity to the native state. Even at lower energy levels, the many sampled local minima can represent noise. Focusing on a smaller ensemble of "interesting" local minima allows more computationally intensive refinement steps to focus resources more effectively. The next section outlines a method for explicitly filtering local minima to reduce the size of the ensemble Ω.

C. Ensemble Filtering

Table II compares the ensemble reduced through an energy threshold, referred to as Ω_E, to the ensemble reduced through multi-objective analysis, referred to as Ω_P (P for Pareto front). Column 3 shows the reduction in the size of Ω_P as a percentage of the size of the original ensemble Ω. Multi-objective analysis is highly effective at reducing the size of Ω, achieving at least a 94% reduction in all cases.

Columns 4-6 show the lowest lRMSD to the native structure in Ω, Ω_E, and Ω_P, respectively. Ω_E is defined as the n lowest-energy conformations, where n is the size of Ω_P. This removes the size of Ω_E as a factor in the lowest lRMSD values retained. For six of the target proteins, Ω_P retains lower lRMSD structures than the energetic reduction

Table I: Columns $2-4$ show the native PDB ID, size and fold topology for each of the 15 target protein systems. Columns 5 and 6 break the fold topology down as the percentage of amino acids which are part of α-helices and β-sheets. Columns $7-10$ report the minimum energy achieved for each temperature T of the minimization component of the BH framework. Columns $11-14$ then report the corresponding lowest lRMSD to the native structure achieved for each T.

	Native					Lowest Energy (kcal/mol)				Lowest lRMSD (Å)			
	PDB ID	Size	fold	% α	% β	$T=0$	T_0	T_1	T_2	$T=0$	T_0	T_1	T_2
1	1dtdB	61	α/β	15	46	-128.2	-132.1	-131.6	-127.9	6.9	6.6	6.9	7.0
2	1isuA	62	α/β	15	19	-127.8	-130.3	-130.7	-130.2	6.3	6.0	6.4	6.0
3	1c8cA	64	α/β	22	48	-133.5	-134.8	-130.8	-129.6	6.5	6.6	7.4	7.3
4	1sap	66	α/β	30	44	-132.8	-132.3	-133.6	-127.3	6.5	6.0	6.8	6.9
5	1hz6A	67	α/β	31	42	-143.5	-144.7	-142.1	-138.9	5.7	5.9	6.0	6.0
6	1wapA	68	β	0	62	-118.4	-127.2	-133.9	-127.9	7.4	7.6	7.4	7.5
7	1fwp	69	α/β	30	26	-152.8	-152.0	-143.5	-143.2	6.3	6.7	6.5	6.1
8	1ail	70	α	84	0	-170.6	-171.0	-167.3	-168.4	3.2	3.2	3.4	3.3
9	1aoy	78	α/β	41	10	-183.9	-181.2	-180.8	-184.1	5.7	6.4	6.0	6.4
10	1cc5	83	α	47	4	-170.9	-171.5	-179.1	-173.8	5.8	5.7	5.8	5.8
11	2ezk	93	α	68	0	-217.3	-218.6	-224.4	-216.0	4.3	4.6	4.2	4.4
12	1hhp	99	β	7	48	-168.7	-175.4	-179.0	-175.9	10.4	10.4	10.0	10.5
13	2hg6	106	α/β	34	21	-233.6	-236.8	-239.5	-235.1	8.8	9.0	8.8	9.2
14	3gwl	106	α	70	0	-264.6	-270.4	-273.9	-267.3	4.9	4.9	4.4	5.2
15	2h5nD	123	α	71	2	-307.8	-313.0	-316.5	-313.2	7.5	7.9	7.4	8.1

in Ω_E. In only the single case with native PDB ID 3gwl does Ω_E contain a structure more than 0.5Å lower than Ω_P.

The reduced Ω_P ensembles are shown for three representative proteins in Figure 2, which draws the energy vs. lRMSD to the native structure for each conformation in the ensemble. Data from Ω_P are drawn in dark blue circles over the data from the original ensemble Ω, which are drawn in light red x's. For reference, the solid green line in Figure 2 represents the cutoff point for Ω_E in each plot, where only conformations below the line are maintained in Ω_E. The purple dashed line marks the energy of the experimental native structure. Figure 2 shows that not only is the area covered by the blue circles smaller, but the density of circles is also much lower than that of the x's. For all three of the proteins shown, there are a significant number of conformations with low lRMSDs to the native structure above the green line, many of which are captured by Ω_P.

Figures 2a and 2b illustrate that the Pareto front captures the left side of the distribution of energy vs. lRMSD; that is, many conformations with both low energy and low lRMSD are retained. For most proteins (represented by Figures 2a and 2b), the difference in the lowest lRMSD to the native structure between Ω and Ω_P is primarily due to a few high-energy outliers. In just two cases (proteins with native PDB IDs 1ail and 2ezk), Ω_P fails to effectively represent Ω, with larger differences in lRMSD of 2Å to 3Å not caused by outliers. Figure 2c shows a cluster of lower-energy conformations on the left side of the ensemble which is very close to the lowest lRMSD structure; however, the Pareto front ignores it entirely.

Taken together, these results show that employing multi-objective analysis to filter the output ensemble provides a distinct advantage over a single energy criterion. The ensemble size reduction is dramatic, yet non-outlier low lRMSD local minima are maintained.

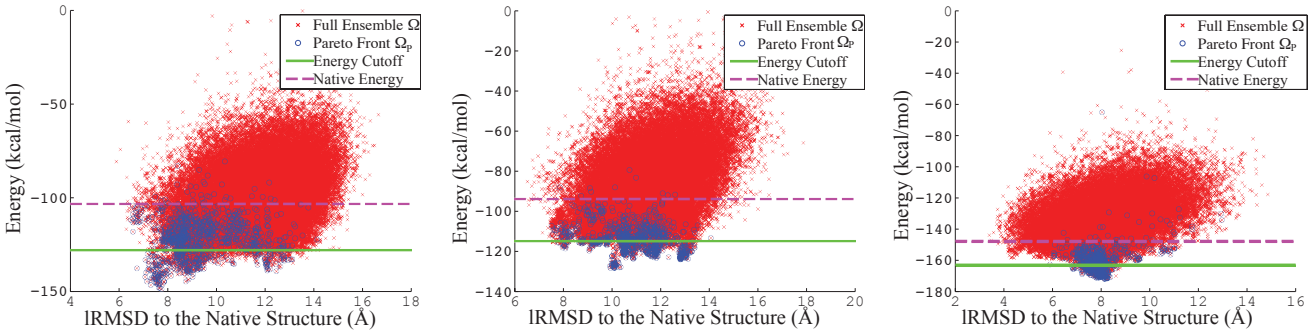
IV. Conclusion

This work presents a detailed investigation into two key components of BH, perturbation and minimization. We show that biasing the perturbation distance directly affects the distance between consecutively-sampled local minima. By tuning this distance, we then enhance near-native sampling in the BH framework. For minimization we show that a simple greedy search is just as effective at sampling conformations near the native state as a more computationally intensive MMC trajectory.

Employing short greedy searches for minimization allows sampling a significantly larger number of local minima than longer MMC trajectories. This larger ensemble provides a broad view of low-energy local minima in the coarse-grained energy surface that may deviate from the true energy surface. To deal with this larger ensemble of local minima, we present a filtering method based on multi-objective analysis. This filtering approach is highly effective for most proteins at reducing the ensemble size while still maintaining non-outlier near-native conformations. Further analysis is needed to determine why the filtering approach fails in a minority of cases. This multi-objective analysis can also serve as a platform to compare different energy functions, not only in their ability to sample local minima, but also in their ability to effectively recognize near-native conformations.

Acknowledgments

This work is supported in part by NSF CCF No. 1016995 and NSF IIS CAREER Award No. 1144106.



(a) 1hz6A (b) 1sap (c) 1ail

Figure 2: Energy vs. lRMSD to the native structure is shown for each conformation in the full ensemble Ω and Pareto front ensemble Ω_P for three representative proteins. The Ω ensemble is plotted as red x's, whereas Ω_P is plotted as blue circles.

REFERENCES

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

[2] J. Moult, K. Fidelis, A. Kryshtafovych, and A. Tramontano, "Critical assessment of methods of protein structure prediction (CASP) round IX," *Proteins: Struct. Funct. Bioinf.*, vol. Suppl, no. 10, pp. 1–5, 2011.

[3] 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.

[4] S. Yin, F. Ding, and N. V. Dokholyan, "Eris: an automated estimator of protein stability," *Nat Methods*, vol. 4, no. 6, pp. 466–467, 2007.

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

[6] A. Verma, A. Schug, K. H. Lee, and W. Wenzel, "Basin hopping simulations for all-atom protein folding," *J. Chem. Phys.*, vol. 124, no. 4, p. 044515, 2006.

[7] A. Shehu, "Conformational search for the protein native state," in *Protein Structure Prediction: Method and Algorithms*, H. Rangwala and G. Karypis, Eds. Fairfax, VA: Wiley Book Series on Bioinformatics, 2010, ch. 21.

[8] 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.

[9] R. Bonneau and D. Baker, "De novo prediction of three-dimensional structures for major protein families," *J. Mol. Biol.*, vol. 322, no. 1, pp. 65–78, 2002.

[10] T. J. Brunette and O. Brock, "Guiding conformation space search with an all-atom energy potential," *Proteins: Struct. Funct. Bioinf.*, vol. 73, no. 4, pp. 958–972, 2009.

[11] 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.

[12] J. DeBartolo, G. Hocky, M. Wilde, J. Xu, K. F. Freed, and T. R. Sosnick, "Protein structure prediction enhanced with evolutionary diversity: SPEED," *Protein Sci.*, vol. 19, no. 3, pp. 520–534, 2010.

[13] 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.

[14] 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.

[15] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, "Comparison of multiple amber force fields and development of improved protein backbone parameters," *Proteins: Struct. Funct. Bioinf.*, vol. 65, no. 3, pp. 712–725, 2006.

[16] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," *J. Phys. Chem. A*, vol. 101, no. 28, pp. 5111–5116, 1997.

[17] M. C. Prentiss, D. J. Wales, and P. G. Wolynes, "Protein structure prediction using basin-hopping." *The Journal of Chemical Physics*, vol. 128, no. 22, pp. 225 106–225 106, Jun. 2008.

[18] B. Olson and A. Shehu, "Populating local minima in the protein conformational space," in *IEEE Intl Conf on Bioinf and Biomed (BIBM)*, 2011, pp. 114–117.

[19] ——, "Evolutionary-inspired probabilistic search for enhancing sampling of local minima in the protein energy surface," *Proteome Sci*, 2012, in press.

[20] G. A. Papoian, J. Ulander, M. P. Eastwood, Z. Luthey-Schulten, and P. G. Wolynes, "Water in protein structure prediction," *Proc. Natl. Acad. Sci. USA*, vol. 101, no. 10, pp. 3352–3357, 2004.

[21] P. Widera, J. M. Garibaldi, and N. Krasnogor, "Evolutionary design of the energy function for protein structure prediction," in *IEEE Congress on Evol Comput (CEC)*, no. 5, 2009, pp. 1067–1077.