Evolving Conformation Paths to Model Protein Structural Transitions

Emmanuel Sapin
Dept of Computer Science
George Mason University
4400 University Dr.
Fairfax, Virginia 22030
esapin@gmu.edu

Kenneth De Jong
Dept of Computer Science
George Mason University
4400 University Dr.
Fairfax, Virginia 22030
kdejong@gmu.edu

Amarda Shehu*
Dept of Computer Science
George Mason University
4400 University Dr.
Fairfax, Virginia 22030
amarda@gmu.edu

ABSTRACT

Proteins are dynamic biomolecules. A structure-by-structure characterization of a protein's transition between two different functional structures is central to elucidating the role of dynamics in modulating protein function and designing therapeutic drugs. Characterizing transitions challenges both dry and wet laboratories. Some computational methods compute discrete representations of the energy landscape that organizes structures of a protein by their potential energies. The representations support queries for paths (series of structures) connecting start and goal structures of interest. In this paper address the problem of modeling protein structural transitions under the umbrella of stochastic optimization and propose a novel evolutionary algorithm (EA). The EA evolves paths without reconstructing the energy landscape, addressing two competing optimization objectives, energetic cost and structural resolution. Rather than seek one path, the EA yields an ensemble of paths to represent a transition. Preliminary applications suggest the EA is effective while operating under a reasonable computational budget.

CCS CONCEPTS

•Applied computing → Molecular structural biology; Bioinformatics;

KEYWORDS

protein modeling, structural transition, energy landscape, conformation path, stochastic optimization, evolutionary algorithm

ACM Reference format:
Emmanuel Sapin, Kenneth De Jong, and Amarda Shehu. 2017. Evolving Conformation Paths to Model Protein Structural Transitions. In *Proceedings of BCB '17 Companion, Boston, Massachusetts, August 20-23, 2017,* 6 pages.
DOI: http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx

*Corresponding Author

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
BCB '17 Companion, Boston, Massachusetts
© 2017 Copyright held by the owner/author(s). yyy-y-yyyy-yyyy-y/17/08…$15.00
DOI: http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx

1 INTRODUCTION

We have now known for decades that proteins undergo fast vibrations and slower structural rearrangements/transitions [8]. Structural transitions allow a protein to access different three-dimensional structures with which to interact with different molecular partners, thus modulating its biological function [11]. A detailed structure-by-structure characterization of a protein's transition between two different functional structures is central to elucidating the role of dynamics and mutation-altered dynamics in (dys)function and the design of novel therapeutic drugs [20].

Rapid advances have occurred in the wet laboratory, most notably by single-molecule methods that can resolve one or more intermediate structures in a transition of interest [6, 7, 9]. Advances have also been made in the dry laboratory, but at the moment no single wet- or dry-laboratory method can provide a detailed, structure-by-structure characterization of transitions on any protein of interest [13]. Challenges are due to the disparate spatio-temporal scales involved; the start and goal structures may differ by several angstroms (Å), and the transition may occur over several milliseconds.

An important concept regarding protein dynamics is the energy landscape, which organizes structures of a protein by their potential energies. A protein energy landscape is rich in iso-energetic minima that are separated by energetic barriers. Different functional structures that allow a protein to interact with different molecular partners correspond to basins (broad neighborhoods of local minima), and a structural transition can be represented as a series of structures that allow the protein to hop from one basin to another [2, 5].

Two canonical approaches to modeling protein structural transitions are Molecular Dynamics (MD) and the Monte Carlo (MC). In MD, repeated applications of Newton's second law of motion and numerical integration allow following small-timescale motions of the atoms constituting a protein of interest. Due to the dependency of numerical integration on small time steps, an MD simulation is likely to converge rapidly to a local minimum near the start structure of the sought transition. The MC approach does away with integration and employs instead biased random walks in the energy landscape. This results in higher exploration capability, though the walks also converge to nearby local minima. Many "enhanced sampling" techniques are proposed to increase the exploration capability of MD or MC approaches [13].

A complementary group of methods aim to increase exploration capability by constructing discrete representations (a tree or a graph) of the protein energy landscape that support queries for

series of intermediate structures that detail a start-to-goal transition of a protein of interest. Some of the earliest work originated in a subcommunity of scientists that took inspiration from mechanistic analogies between modeling protein motions and planning robot motions [24]. Often, such methods are referred to as robotics-inspired.

Robotics-inspired methods rely on sampling conformations to generate new structures (a conformation is a chosen representation of a molecular structure). Tree-based methods generate such samples as part of the process of growing a tree (vertices are the samples). The growth of the tree is biased so as to propagate from the start to the goal structure with low-energy intermediate structures. The result is a local (discrete) representation of the energy landscape (the tree), that may not correspond to the lowest-cost path that connects the given start and goal structures. On the other hand, graph-/roadmap-based methods first generate samples and then embed them in a nearest-neighbor graph where each sample is connected to its nearest neighbors. These methods are more likely to capture the lowest-cost path (as well as support multiple path queries between any structures of interest), but the nonlocal view comes at a higher computational cost (many samples/structures). Advances have been made in roadmap-based modeling of protein structural transitions, and the reader is referred to [24] for a review of the state of the art. For instance, some roadmap-based methods can model transitions between start and goal structures as far as 13Å away from each-other [14, 15, 17–19].

Tree- and roadmap-based methods seek the lowest-cost path to represent a transition; the notion of cost captures that of the work that has be to done for a protein to "go over" the energetic barriers. Similar-cost paths may exist, considering the multidimensionality of the landscape; however, finding the lowest-cost path is challenging by itself. In roadmap-based methods, path resolution (the maximal distance between two consecutive structures) is not an explicit objective alongside cost, and sampling more structures is seen as the only way to improve path resolution. Tree-based methods can directly control resolution by generating a child structure that is low in energy and close to its parent in the tree. The cost of the path connecting the root (the start structure) to a leaf deemed close to the goal structure is not directly controlled.

In this paper we treat both path cost and path resolution as objectives when seeking paths to represent a structural transition. Specifically, we cast the problem of modeling protein structural transitions as a stochastic optimization problem; to the best of our knowledge, this is a novel approach. We propose an evolutionary algorithm (EA) that evolve paths that are both low in cost and high in resoluton. The EA directly controls both path cost and resolution and yields an ensemble of paths to better represent a structural transition in the multidimensional protein energy landscape. We now relate details of the proposed EA.

2 METHODS

Several building blocks employed in the proposed EA have been developed and analyzed by us in prior work on memetic EAs for capturing structured, sample-based representations of protein energy landscapes [21–23]. A key building block is the exploitation of experimentally-known structures of a protein to select a representation of a structure (a conformation). Briefly, the experimentally-known structures of a protein (of the wildtype form and mutated versions/variants) are subjected to principal component analysis (PCA). PCA yields a new basis whose axes are ordered by the amount of variance they capture in the structure data. The axes are also referred to as principal components (PCs). The top m PCs that collectively capture a threshold variance are selected as variables with which to represent a conformation. Typically, a threshold of 80−90% results in $m \in [10, 25]$, which is over a hundred-fold reduction over the number of Cartesian coordinates (were one to use such coordinates of all the atoms in a protein). Further details can be found in Ref. [4].

Path Representation. The proposed EA evolves path individuals utilizing only two given (experimentally-known) structures. These structures are projected onto the start and goal conformations. A path is represented as an ordered (start-to-goal directed) list of conformations.

Path Cost The fitness/cost of a path sums up the energy increase between structures corresponding to consecutive conformations in the path. This path cost formulation considers only the amount of work needed for a protein to "climb"' energy barriers separating local minima in the landscape (and has been employed in robotics-inspired methods [10, 15]).

Path Resolution The resolution of a path is measured here as the maximum distance (via the Euclidean distance function) between two consecutive conformations. A complementary, valid consideration is to make use of the root-mean-root-mean-square-deviation (rmsd) [16]. However, rmsd is a computationally-costly measurement due to the need to align the two structures. In addition, the conformations in a path are not yet structures (but rather points in the m−dimensional space of PCs extracted from experimentally-known structures).

Cost versus Resolution. Fig. 1 illustrates the relationship between path cost and resolution. A low-resolution path may erroneously report that the cost of a basin-basin transition is low, ignoring the separating barrier. Fig. 1(a) juxtaposes the lowest-resolution path (no intermediate structures) in such a scenario to a path that follows the landscape more closely and as a result has higher cost. Fig. 1(b) illustrates another scenario, where the start and goal structures reside on nearby barriers separated by a basin. These two examples illustrate that a low-resolution path is easy to compute (fewer intermediate structures) but associates an inaccurate, possibly low energetic cost with a transition. On the other hand, as the resolution of a path improves due to more intermediate structures, the path cost may increase due to the path following the landscape more closely.

Initial Population. n_points are obtained by linear interpolation between the given start and end conformations. Each resulting conformation undergoes a transformation, which converts a conformation to an all-atom protein structure that corresponds to a local minimum in the all-atom Rosetta energy landscape [12]. The transformation relies on the Rosetta *relax* function, which is stochastic; repeating it N times allows us to obtain an initial population of N different paths. The transformation of a conformation to an all-atom structure corresponding to a local minimum is also a building block developed and analyzed in prior work [3].

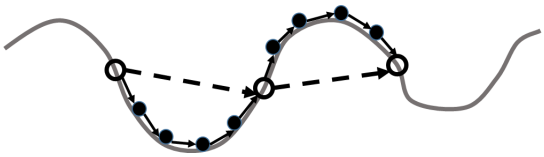


Figure 1: The first pair of structures (empty circles) can be connected via a path with no intermediate structures. A higher-resolution path with intermediate structures (black dots) would correctly take into account the basin in the cost of the transition. The second pair of structures (empty circles) can be connected via a path with no intermediate structures, missing the barrier. A higher-resolution path would contain intermediate structures in the barrier and consider that in the cost of the transition.

Variation Operator. Successive generations evolve the population of paths as follows. For every two consecutive conformations in a path, a variation operator yields a new mid-point/conformation, which is converted to a (local minimum) all-atom structure utilizing the transformation summarized above. The variation operator is not explicitly yielding a path offspring but rather additional conformations from a path individual. The burden of providing offspring paths is passed on to the selection operator.

Selection Operator. The selection operator takes all conformations (existing ones in the paths and new ones obtained from the variation operator) and inserts them into a nearest-neighbor graph (nngraph). The nngraph connects a point to others within a ball of radius ϵ; ϵ is measured via the Euclidean distance function. N consecutive applications of Dijkstra's path search algorithm on the nngraph yield the N paths that initialize the next generation. After each application, the points utilized in the found path are removed from the graph, and the next application of Dijkstra's is carried out in the induced subgraph. This mechanism yields N low-cost non-redundant paths.

Evolutionary Pressure. ϵ decreases or increases to apply pressure to find high-resolution paths. At the first generation (after the initial population), ϵ is set to $d/(x+i*0.1)$, where d is the Euclidean distance between the start and goal conformations, $x = 1$, and $i = 1$. If a path is found, i is incremented, and this process continues until a path cannot be found (we adapt Dijkstra's to consider dynamic neighbors that disappear as ϵ decreases). If a path is not found, the path that is reported is the one found at the prior attempt, and this is the value of i that is employed in the next application of Dijkstra's in the induced subgraph to find another path. The last successful value of i (yielding the N^{th} path) is recorded and in the next generation ϵ is initialized to a lower value by incrementing the last successful value of i. This process provides pressure for the EA to find both low-cost and high-resolution paths. Note that while in the initial population all paths have the same length in terms of the number of conformations, the paths obtained in subsequent generations have variable lengths. Also note that the so-described selection operator circumvents the issue of comparing two paths to determine which one is better. The latter is non-trivial due to the complex relationship between cost and resolution. Cost and resolution are treated as different optimization objectives by the EA, and the selection operator is key to addressing both.

Implementation Details. The EA is implemented in C++ and run on a 16 core Red Hat Linux box with 3.2 GhZ HT Xeon CPU and 8GB RAM. The cores allow to parallelize the transformation of conformations into structures. $m = 10$, $N = 15$, and $n\_points = 10$. In Section 3, the resolution of paths is converted to rmsd [16] and reported in Ådue to domain-specific insight on what rmsds correspond to low or high values. The EA operates under a fixed computational budget, tallying up the number of energy evaluations in the transformations of conformations to structures. For most of the applications described in Section 3, the budget is 100,000 fitness evaluations. On a protein of 166 amino acids, the total running time of the EA is about 38 CPU hours; a significant portion of this time is devoted to conduct N lowest-cost path searches in each generation.

3 RESULTS

We analyze the performance of the proposed EA on four variant sequences of the H-Ras enzyme, which is a central enzyme to cell growth regulation. Specifically, we apply the EA to the WT sequence, two oncogenic variants, G12C and Q61L, and a syndrome-causing variant Y32CC118S. The naming convention `[AACode1][Position][AACode2]` for a single-mutant variant indicates that the amino acid `AACode1` (using one-letter codes) at position `Position` in the WT is replaced with the amino acid `Code2` in the particular variant. Additional mutations are joined in order of positions, as in the Y32CC118S variant we consider here.

The EA is run to obtain paths connecting two different biologically-relevant structures of H-Ras that correspond to the on/active state and the off/inactive states. The start structure used to represent the on state is obtained from the Protein Data Bank (PDB) [1] under PDB entry id `1qra`. The goal structure selected to represent the off state is obtained from the PDB entry with id `4q21`.

For each of the test cases, the computational budget of the EA is fixed to 100,000 fitness evaluations, which is 10 times less than that used in the roadmap-based method in [21] to which we compare the quality of the computed paths. Specifically, the 15 lowest-cost paths obtained by the proposed EA with this budget are compared to the top paths found by the roadmap-based method in [21].

We relate the following results. First, we compare the top paths found by the EA to those found by the roadmap-based method in [21] on H-Ras WT in terms of both cost and resolution. We then draw these paths and further demonstrate that the proposed EA has not reached saturation; when provided a higher computational budget, it is able to further improve the quality of paths. Finally, we show the performance of the EA on the other three H-Ras variants.

3.1 Comparative Analysis on H-Ras WT

Table 1 compares the cost and resolution of ten top paths found by the proposed EA (Columns 1-4) to those of the top ten paths found by the roadmap-based method in [21] (Columns 5-6). Paths are ordered from high to low costs. The proposed EA is run under two regimes, under a budget of 100,000 fitness evaluations (Columns 3-4) and then a budget of 200,000 fitness evaluations (Columns 1-2). These budgets are 1/10 and 1/5, respectively, the budget employed

by the roadmap-based method in [21]. All paths obtained by the roadmap-based method have the same resolution, as this is imposed in the construction of the nngraph prior to answering path queries. As described in Section 2, the paths obtained by the proposed EA have varying resolution.

The comparison in Table 1 shows that the proposed EA obtains very high-resolution (0.133Å and 0.129Å under the two budgets) paths with much less computational budget (and so fewer generated structures) than the roadmap-based method. The resolutions even with the more modest budget are better than the roadmap-based method. As described in Section 2, path costs at higher resolutions typically increase due to the high ruggedness of protein energy landscapes. Taken altogether, these results demonstrate that the proposed EA is able to find paths with better resolution than the roadmap-based method while operating under a smaller computational budget.

Table 1: Top ten paths obtained by each algorithm.

Proposed EA 200K fitness evals		Proposed EA 100K fitness evals		Roadmap-based [21] 1M fitness evals	
Cost	Res	Cost	Res	Cost	Res
487	0.129	554	0.133	588	0.145
301	0.134	296	0.146	546	0.145
291	0.139	292	0.143	504	0.145
288	0.152	149	0.162	470	0.145
271	0.157	148	0.172	408	0.145
267	0.150	129	0.172	395	0.145
263	0.172	123	0.172	376	0.145
251	0.143	113	0.195	324	0.145
248	0.158	112	0.190	306	0.145
236	0.172	109	0.192	266	0.145

Table 1 also shows that the EA is able to further improve the quality of paths when afforded more fitness evaluations (and so allowed to evolve paths over more generations). Specifically, the average value over the resolutions of of the paths improves from 0.168Å (when the budget is 100, 000 fitness evaluations) to 0.151Å when the budget is doubled. Comparing Columns 1 to 3 makes it clear that the higher resolutions invariably increase the costs of paths. Specifically, the average path cost increases from 170.33REUs to 260.07REUS when the budget is doubled. This supports the relationship between path cost and resolution that is illustrated in Section 2 and motivates the recasting of the path computation problems as a multi-objective optimization problem. With paths following the rugged landscape more closely, path costs are expected to increase.

The actual paths found by the proposed EA under the two computational budgets are drawn in Figure 2. Panel (a) draws the paths in the final generation of the proposed EA restricted to a budget of 100, 000 fitness evaluations, and panel (b) does so when the budget is doubled. The paths are drawn by connecting consecutive structures with edges, with the start structure on the right of each panel and the goal structure on the left. Conformations are drawn as dots, using their coordinates along the top two PCs. Conformations generated by the EA but not utilized in the top paths are also drawn and color-coded by the Rosetta all-atom energy values of their corresponding structures. Comparison of panels (a) and (b) in Figure 2 clearly reveals that under the higher computational budget, the top paths phenotypically converge to a narrower region of the fitness landscape and thus capture the sought structural transition more accurately.

3.2 Comparative Analysis on Pathogenic Variants

Figure 3 shows the paths in the last generation of the proposed EA for each of the three other test cases that are oncogenic variants of H-Ras. The resolutions are generally worse in comparison to what the EA can find for the WT sequence, which helps reduce the total cost of a path, as described earlier (the resolutions are better than those found by the roadmap-based EA on the WT sequence). The lower resolutions for these three variants in comparison to what the proposed EA finds on the WT sequence point to possibly more complex landscapes where more applications of the variation operator may be needed to uncover local minima in higher-energy regions. Visual comparison of the generated conformations (color-coded projections) across the WT and these three variants suggests higher-energy regions separating the start and goal structures sought to be connected by the path-evolving EA for G12C and Y32CC118S, but a possibly higher-energy, off basin (left of the panels in Fig. 3.

4 CONCLUSION

We have presented a novel algorithm that, to the best of our knowledge, is the first to cast and address the problem of modeling protein structural transitions as a multi-objective optimization problem. Specifically, this paper makes the case that both energetic cost and resolution, which are ineffectively treated at the moment by other methods, are key to the quality of a modeled structural transition.

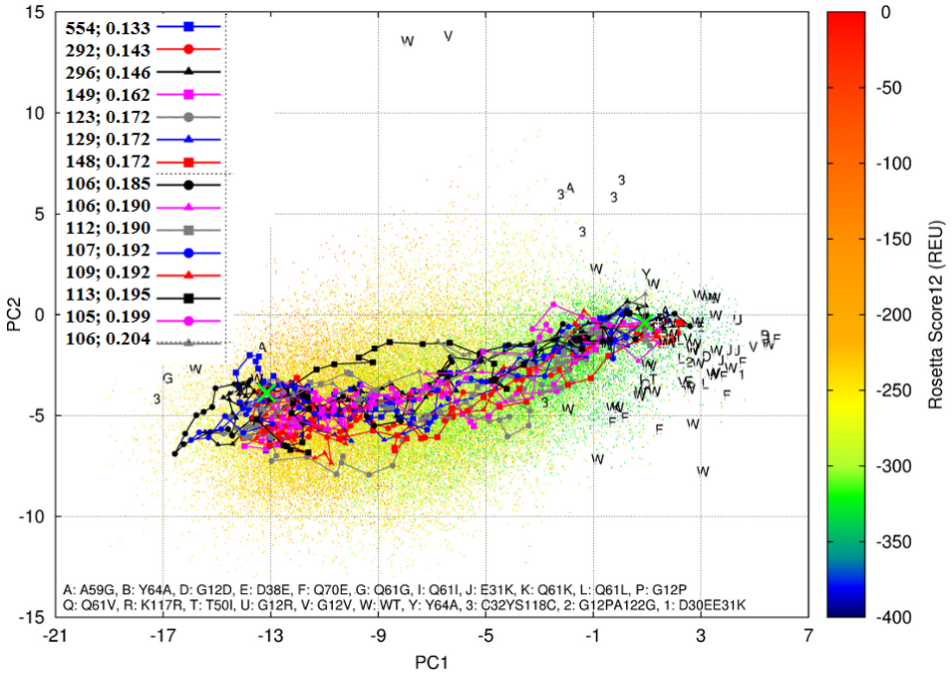
The proposed algorithm is an EA, as EAs are effective at addressing hard optimization problems. Moreover, unlike previous adaptations of EAs for structure-related problems in computational structural biology, the proposed EA evolves paths rather than molecular structures. The EA also relies on a novel selection operator to address both path cost and path resolution as competing objectives.

The evaluation demonstrates the capability of the proposed EA to improve over state-of-the-art, landscape-reconstructing, roadmap-based methods [21]. The analysis on th H-Ras WT also shows the ability of the proposed EA to further improve the quality of its solutions with larger computational budgets. The emphasis on lower computational budgets is due to the potential of path-evolving EAs to obtain and then compare the structural dynamics of various forms of a protein in a large-scale setting. This would allow elucidating mutation-altered dynamics and its impact in dysfunction.

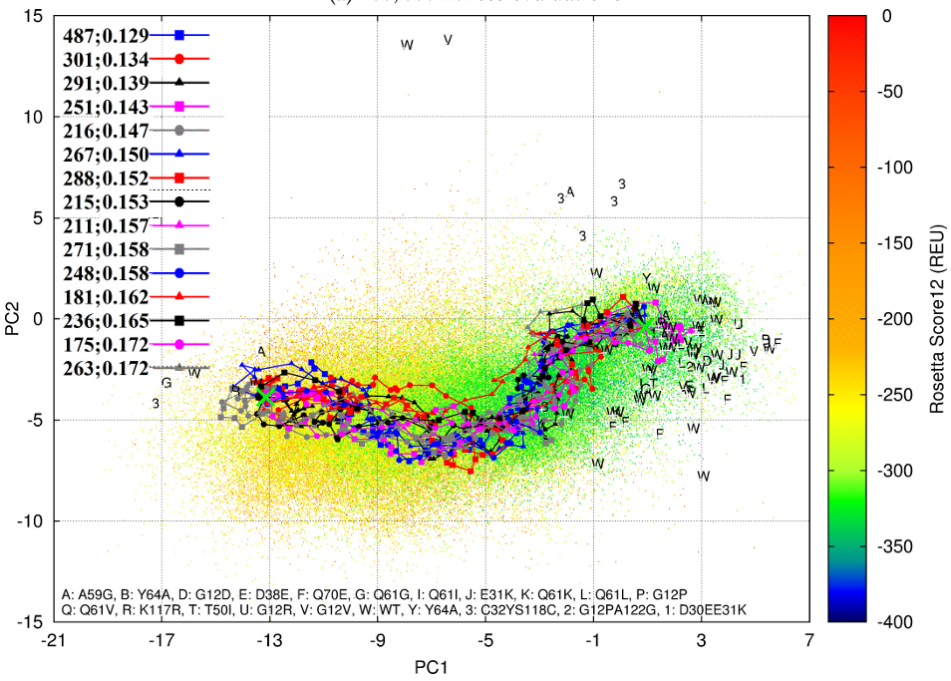
The results suggest that treating the problem of modeling protein structural transitions as a multi-objective one is reasonable and can result in novel research directions. Evolving individuals with complex representations, such as paths, is of interest in evolutionary computation and is likely to spur further research on effective variation and selection operators.

5 ACKNOWLEDGMENTS

This work is supported in part by NSF CCF No. 1421001 and NSF IIS CAREER Award No. 1144106.

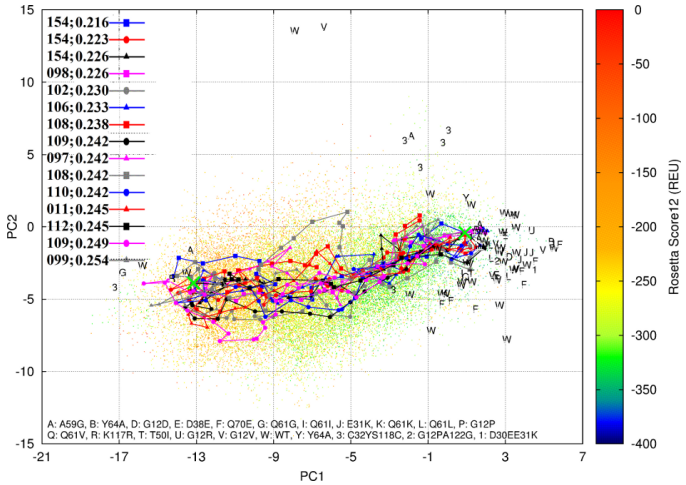


(a) 100,000 fitness evaluations

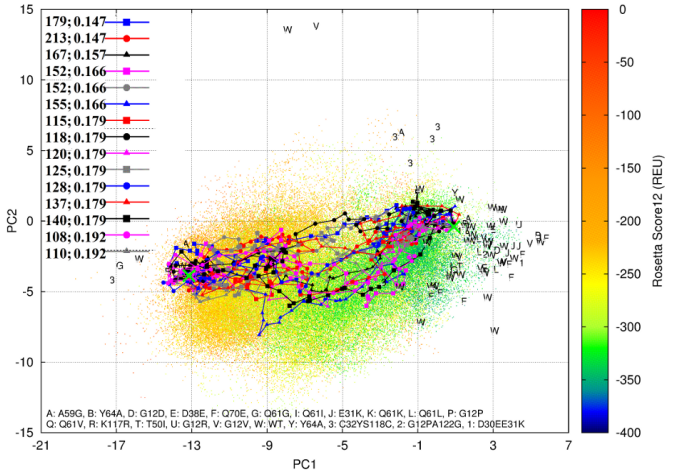


(b) 200,000 fitness evaluations

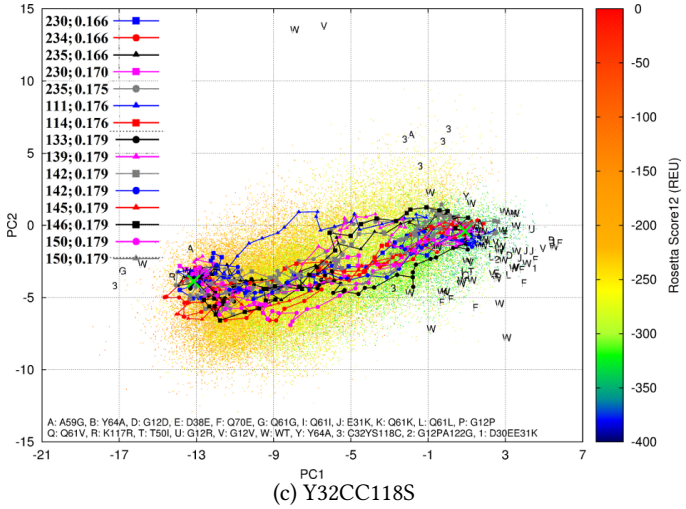
Figure 2: The EA is allowed a computational budget of (a) $100,000$ or (b) $200,000$ fitness evaluations; the paths in the last generation are shown in each setting, drawing an edge between two consecutive structures. Dots show conformations. Dots outside the drawn paths are color-coded and note conformations generated during the execution of the algorithm but not selected by the selection operator. The blue-to-red color-coding scheme encodes low-to-high energy values measured with the all-atom Rosetta energy function (`score12`). The text annotations indicate projections of experimentally-known structures. The legend in each plot lists the path costs in Rosetta energy units (REUs) and their resolutions in angstroms (Å).



(a) G12C



(b) Q61L



(c) Y32CC118S

Figure 3: The paths obtained by the proposed EA in the last generation (with a budget of 100,000 fitness evaluations) are shown here for three more test cases. The same plotting style is followed as in Fig. 2.

REFERENCES

[1] H. M. Berman, K. Henrick, and H. Nakamura. 2003. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 10, 12 (2003), 980–980.

[2] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct Funct Genet* 21, 3 (1995), 167–195.

[3] R. Clausen, B. Ma, R. Nussinov, and A. Shehu. 2015. Mapping the Conformation Space of Wildtype and Mutant H-Ras with a Memetic, Cellular, and Multiscale Evolutionary Algorithm. *PLoS Comput Biol* 11, 9 (2015), e1004470.

[4] R. Clausen and A. Shehu. 2015. A Data-driven Evolutionary Algorithm for Mapping Multi-basin Protein Energy Landscapes. *J Comp Biol* 22, 9 (2015), 844–860.

[5] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motion on proteins. *Science* 254, 5038 (1991), 1598–1603.

[6] A. Gall, C. Ilioaia, T. P. Krüger, V. I. Novoderezhkin, B. Robert, and R. van Grondelle. 2015. Conformational Switching in a Light-Harvesting Protein as Followed by Single-Molecule Spectroscopy. *Biophys J* 108, 11 (2015), 2713–2720.

[7] W. J. Greenleaf, M. T. Woodside, and S. M. Block. 2007. High-Resolution, Single-Molecule Measurements of Biomolecular Motion. *Annu Rev Biophys Biomol Struct* 36 (2007), 171–190.

[8] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hübner, and D. Kern. 2007. Intrinsic motions along an enzymatic reaction trajectory. *Nature* 450, 7171 (2007), 838–844.

[9] J. Hohlbein, T. D. Craggs, and T. Cordes. 2014. Alternating-laser excitation: single-molecule FRET and beyond. *Chem Soc Rev* 43, 4 (2014), 1156–1171.

[10] L. Jaillet, J. Cortés, and T. Siméon. 2010. Sampling-based path planning on configuration-cost costmaps. *IEEE Trans Robot* 26, 4 (2010), 635–646.

[11] K. Jenzler-Wildman and D. Kern. 2007. Dynamic personalities of proteins. *Nature* 450, 7172 (2007), 964–972.

[12] A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovifj, and et. al. 2011. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* 487 (2011), 545–574.

[13] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu. 2016. Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLoS Comput Biol* 12, 4 (2016), e1004619.

[14] T. Maximova, E. Plaku, and A. Shehu. 2015. Computing Transition Paths in Multiple-Basin Proteins with a Probabilistic Roadmap Algorithm Guided by Structure Data. In *Intl Conf on Bioinf and Biomed (BIBM)*. IEEE, Washington, D.C., 35–42.

[15] T. Maximova, E. Plaku, and A. Shehu. 2016. Structure-guided Protein Transition Modeling with a Probabilistic Roadmap Algorithm. *IEEE/ACM Trans Comput Biol & Bioinform* 13, 5 (2016), 1–14.

[16] A. D. McLachlan. 1972. A mathematical procedure for superimposing atomic coordinates of proteins. *Acta Crystallogr A* 26, 6 (1972), 656–657.

[17] K. Molloy, R. Clausen, and A. Shehu. 2016. A Stochastic Roadmap Method to Model Protein Structural Transitions. *Robotica* 34, 8 (2016), 1705–1733.

[18] K. Molloy and A. Shehu. 2013. Elucidating the Ensemble of Functionally-relevant Transitions in Protein Systems with a Robotics-inspired Method. *BMC Struct Biol* 13, Suppl 1 (2013), S8.

[19] K. Molloy and A. Shehu. 2015. Interleaving Global and Local Search for Protein Motion Computation. In *LNCS: Bioinformatics Research and Applications*, R. Harrison, Y. Li, and I. Mandoiu (Eds.), Vol. 9096. Springer International Publishing, Norfolk, VA, 175–186.

[20] Colm Ó'Dúnlaing, Micha Sharir, and Chee K. Yap. 2014. Protein Conformational Dynamics. In *Advances in Experimental Medicine and Biology*, K.-L. Han, X. Zhang, and M.-J. Yang (Eds.). Springer International Publishing, Switzerland, 488.

[21] E. Sapin, D. B. Carr, K. A. De Jong, and A. Shehu. 2016. Computing energy landscape maps and structural excursions of proteins. *BMC Genomics* 17, Suppl 4 (2016), 456.

[22] E. Sapin, K. A. De Jong, and A. Shehu. 2016. A Novel EA-based Memetic Approach for Efficiently Mapping Complex Fitness Landscapes. In *Conf on Genetic and Evolutionary Computation (GECCO)*. ACM, 85–92.

[23] E. Sapin, K. A. De Jong, and A. Shehu. 2017. From Optimization to Mapping: An Evolutionary Algorithm for Protein Energy Landscapes. *IEEE/ACM Trans Comput Biol & Bioinform* (2017). in press.

[24] A. Shehu and E. Plaku. 2016. A Survey of omputational Treatments of Biomolecules by Robotics-inspired Methods Modeling Equilibrium Structure and Dynamics. *J of Artif Intell Res* 597 (2016), 509–572.