

# Modeling Protein Structural Transitions as a Multiobjective Optimization Problem

Emmanuel Sapin<sup>1</sup>, Kenneth A De Jong<sup>1,2</sup>, and Amarda Shehu<sup>1,3,4\*</sup>

<sup>1</sup>Department of Computer Science, <sup>2</sup>Krasnow Institute of Neuroscience,

<sup>3</sup>Department of Bioengineering, <sup>4</sup>School of Systems Biology,  
George Mason University, Fairfax, VA, 22030, United States

esapin@gmu.edu, kdejong@gmu.edu, amarda@gmu.edu

\*Corresponding Author

**Abstract**—Proteins of importance to human biology can populate significantly different three-dimensional (3d) structures at equilibrium. By doing so, a protein is able to interface with different molecules in the cell and so modulate its function. A structure-by-structure characterization of a protein’s transition between two structures is central to elucidate the role of structural dynamics in regulating molecular interactions, understand the impact of sequence mutations on function, and design molecular therapeutics. Much wet- and dry-laboratory research is devoted to characterizing structural transitions. Computational approaches rely on constructing a full or partial, structured representation of the energy landscape that organizes structures by potential energy. The representation readily yields one or more paths that consist of series of structures connecting start and goal structures of interest. In this paper, we propose instead to cast the problem of computing transition paths as a multiobjective optimization one. We identify two desired characteristics of computed paths, energetic cost and structural resolution, and propose a novel evolutionary algorithm (EA) to compute low-cost and high-resolution paths. The EA evolves paths representing a specific structural excursion without *a priori* constructing the energy landscape. Preliminary applications suggest the EA is effective while operating under a reasonable computational budget.

## I. INTRODUCTION

While it has long been known that proteins are dynamic systems in perpetual motion [1], decades of experimental research in molecular biology have found that proteins undergo both fast vibrations and slower structural rearrangements. The latter, also referred to as structural transitions, allow a protein to access different 3d structures with which to interface with different molecular partners in the cell and so modulate its biological function [2]. Human biology is rich in such multi-state proteins that are able to populate different structural states at equilibrium [3], [4]; in fact, this ability is a direct result of evolution. Making use of the same protein in different chemical reactions is a unique characteristic of higher-order organisms [5].

A structure-by-structure characterization of a protein’s transition between two different, functional structures is central to molecular biology for several reasons. First, such a characterization elucidates the role of structural dynamics in regulating molecular interactions and thus protein function [6]. Second, many DNA mutations that result in amino-

acid variations in the amino-acid sequence of a protein can perturb these structural transitions, and this perturbation can result in dysfunction [5], [7]. Third, a detailed characterization of a structural transition of interest provides computational chemists with the structural information needed to design effective molecular interventions [8].

For these reasons, much wet- and dry-laboratory research in molecular biology is devoted to characterizing transitions of proteins between different functional structures. The last decade has seen rapid advancements in the wet laboratory, most notably by single-molecule techniques that can now report one or more intermediate structures in a transition of interest [9]–[11]. Advancements have also been made in the dry laboratory. At the moment, no single wet- or dry-laboratory technique can provide a detailed, structure-by-structure characterization of transitions independent of protein size and biological timescale [12]. The difficulty is related to the disparate spatio-temporal scales that may be involved in a transition; the structures of interest may differ by several angstroms (Å), and the transition connecting them may occur over several milliseconds.

Computational approaches that seek to characterize structural transitions utilize the concept of the energy landscape, which organizes structures by potential energies. Protein energy landscapes are rich in iso-energetic minima that are separated by energetic barriers. In the energy landscape, the long-lived, functional structural states that a protein employs to “stick to” other molecules in the cell correspond to basins (broad neighborhoods of local minima), and a structural transition can be represented as a series of points (structures) that allow the protein to hop from one basin to another.

In principle, the energy landscape contains all the necessary information to characterize and even simulate structural transitions [13], [14]. However, energy landscapes are vast, high-dimensional, and rich in local minima [15], [16]. These characteristics challenge conventional molecular dynamics (MD) approaches that rely on Newton’s second law of motion and numerical integration to follow motions of the atoms that comprise a protein [17]. MD approaches typically surpass only low energetic barriers and converge rapidly to local minima near the start structure of the sought transition. As a result, these approaches cannot be used to capture

structural transitions that occur beyond the nanosecond timescale. Monte Carlo (MC) approaches do away with Newton’s second law of motion and instead launch biased random walks in the energy landscape [18]. Many “enhanced sampling” techniques are regularly proposed to increase the exploration capability of MD- or MC-based modeling of protein structural transitions [12].

Improved exploration capability is also the focus of a complementary set of methods that exploit mechanistic analogies between molecular chains and robot kinematic linkages [19] and adapt techniques designed to plan robot motions to the problem of modeling protein structural transitions [20]. These methods essentially organize MC walks in trees or graphs (the latter are also referred to as roadmaps) that constitute structured representations of the energy landscape of a protein of interest. Such representations readily yield paths connecting given start and goal structures.

Tree-based methods build a partial representation of the energy landscape that corresponds to a highly local view of the landscape, whereas roadmap-based methods seek to obtain a more global view and encode that view in a graph that can then be queried to reveal multiple paths connecting any pair of a set of structures of interest. Tree-based methods bias the growth of the tree so as to propagate from the start to the goal structure with low-energy intermediate structures. The local view may result in missing the lowest-cost path representing a structural transition. On the other hand, roadmap-based methods have a higher likelihood of capturing such a path, but the nonlocal view comes at a higher computational cost. The bulk of the time is spent on generating many structures to provide the nonlocal view. Advances have been made recently in roadmap-based modeling of protein structural transitions, and the reader is referred to [20] for a review of the state of the art. We draw attention to state-of-the-art roadmap-based methods that can provide structure-by-structure characterizations of transitions between start and goal structures as far as 13Å away from each-other [21]–[25].

Current robotics-inspired methods are ineffective in considering two important characteristics of paths sought to represent a transition. The objective, based in thermodynamics, is that, if offering only one path, the lowest-cost path (often related to the path that requires the least amount of work) best represents a structural transition. Other similar-cost paths may exist, particularly considering the multidimensionality of the landscape; however, even finding the lowest-cost path representing a structural transition is challenging by itself. With this objective, roadmap-based methods compute many structures (via sampling), and then draw edges between a structure and its  $k$  nearest neighbors to obtain a graph that is then queried via path search algorithms for the lowest-cost path connecting a given start to a given goal structure. The resolution of the path (the maximal distance between two consecutive structures) is not directly

controlled; generally, these methods resort to sampling more structures in efforts to improve path resolution. On the other hand, tree-based methods can control resolution by generating a child structure that is both low in energy and close to its parent in the tree. However, the cost of the path connecting a start to a leaf that is deemed to be sufficiently close to the goal structure is not directly controlled.

To illustrate the relationship between path cost and path resolution, we direct the reader to Fig. 1. If the start and goal structures reside in two different basins separated by an energy barrier, a low-resolution path may erroneously report that the cost of the basin-basin transition is low, ignoring the separating barrier. In the scenario depicted in Fig. 1(a), the lowest-resolution path is the one with no intermediate structures; this path (and others with low resolution) effectively dig a tunnel through the energetic “mountain.” A similar scenario is depicted in Fig. 1(b), where the start and goal structures are on two different barriers. The lowest-resolution path with no intermediate structures effectively sets a bridge between the two structures, again ignoring the valley underneath. These two extreme examples illustrate the fact that a low-resolution path is easy to compute (saving time in the computation of intermediate structures) but associates an inaccurate, possibly low energetic cost with a transition. On the other hand, as one improves the resolution of a path by computing more intermediate structures, the path cost may indeed increase due to the path following the landscape more closely.

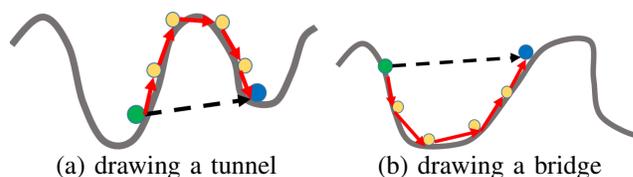


Figure 1: In (a), the lowest-resolution path that contains no intermediate structures and effectively tunnels through the basin-separating barrier. A higher-resolution path contains more intermediate structures (in yellow) and follows the landscape more faithfully in connecting the start (in green) and goal (in blue) structures. In (b), the lowest-resolution path builds a bridge to connect the two barriers, ignoring the basin underneath. In a higher-resolution path, the intermediate structures reside in the basin.

This suggests that the problem of computing paths representing structural transitions can be usefully cast as a multi-objective optimization problem, as both low cost and high resolution are desired characteristics of computed paths. Moreover, rather than one path, the view of a structural transition as an ensemble of paths makes this problem appropriate for a multi-objective, population-based EA. Unlike a robotics-inspired tree-based method, the EA proposed here computes many paths that connect a start to a goal structure.

Unlike a roadmap-based method, the EA directly controls both the cost and resolution of paths via the selection mechanism. Moreover, the proposed EA evolves paths without *a priori* constructing the energy landscape, thus resulting in lower computational demands over roadmap-based methods.

We first provide a short review of related work, where we summarize several building blocks employed in the proposed EA. Section II then relates details of the proposed EA. Section III showcases applications on healthy and pathogenic variants of a protein central to human health. The results are promising, and, as discussed in Section IV, warrant further investigation of evolutionary search techniques for modeling protein structural transitions under the umbrella of multi-objective optimization.

### A. Related Work

The EA proposed here leverages several algorithmic building blocks, most notably on effective, low-dimensional representations of structures of medium-size proteins. These building blocks have been developed and analyzed by us in prior work on memetic EAs for capturing full, structured representations of energy landscapes of multi-state proteins [26]–[28].

A key building block is the ability to exploit experimentally-known structures (detected in the wet laboratory) of a protein to define the variable space. In summary, the experimentally-known structures of a protein of interest (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 new axes of the variable space in which the EA operates. 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 as variables). The EA proposed here operates in the space of  $m$  PCs, extracted as summarized. A structure corresponds to a point in this space, and a series of structures (path) corresponds to a series of points. The reader is directed to Ref. [28] for further details.

## II. METHODS

The proposed EA evolves paths utilizing only two given (experimentally-known) structures. These structures initialize the start and end points for all sought paths in the variable space in which the EA operates. Specifically, a path individual is represented as an ordered (start-to-goal directed) list of points in the variable space. The fitness/cost of a path sums up the energy increase between structures corresponding to consecutive points. This way of tallying up the cost of a path effectively considers only the amount of

work that would be needed for a protein to “go up” energy barriers separating local minima in the landscape. This path cost formulation is employed in many methods that model transitions in landscapes [22], [29].

In addition to cost, resolution is also associated with a path. The resolution of a path is measured here as the maximum distance between two consecutive points in the path; distance is measured via the Euclidean distance function in the variable space. A complementary, valid consideration is to make use of the root-mean-root-mean-square-deviation (rmsd) [30]. However, rmsd is a computationally-costly measurement due to the need to align the two structures. In addition, the points in a path are not yet structures.

The initial population of  $N$  paths is constructed as follows. Initially,  $n\_points$  are obtained by linear interpolation between the given start and end points. Each obtained point undergoes a transformation, which converts a point to an all-atom protein structure that corresponds to a local minimum in the all-atom Rosetta energy landscape. The transformation utilizes stochastic optimization, so repeating it  $N$  times yields an initial population of  $N$  different paths. The transformation of a point in the variable space to an all-atom structure corresponding to a local minimum is also a building block developed and analyzed in detail in our prior work, and we direct the reader to Ref. [31] for details.

Successive generations evolve the population of paths in the following way. For every two consecutive points in a path, a variation operator yields a new mid-point, which is then converted to a (local minimum) all-atom structure utilizing the transformation summarized above. Note that this variation operator is not explicitly yielding a path offspring but instead providing additional points from a path individual. The burden of providing offspring paths is passed on to the selection operator.

The selection operator takes all points (existing ones in the paths in the population 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  $\epsilon$ ; the radius  $\epsilon$  is measured via the Euclidean distance function in the variable space.  $N$  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.

The value of  $\epsilon$  decreases or increases to apply pressure to find high-resolution paths. At the first generation (after the initial population),  $\epsilon$  is set to  $d/(x + i * 0.1)$ , where  $d$  is the Euclidean distance between the start and goal points corresponding to the given start and goal structures,  $x = 1$ , 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  $\epsilon$

takes on lower values). 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  $N^{\text{th}}$  path) is recorded and in the next generation  $\epsilon$  is initialized to a lower value by incrementing the last successful value of  $i$ . This process provides selection pressure for the EA to find both low-cost and high-resolution paths. We note that while in the initial population all paths have the same number of points ( $n_{\text{points}}$ , so same length), the paths obtained in subsequent generations have variable lengths.

To the best of our knowledge, the described selection operator is novel in several ways. The paths in a population are effectively rewired, and a graph structure is used to centralize the view of the landscape from the different paths. New points are added onto the graph to increase path diversity. Moreover, this selection mechanism circumvents the issue of comparing two paths to determine which one is better. The latter is not 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 mechanism is key to addressing both.

#### A. Implementation Details

The proposed 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 are employed to parallelize offspring improvements. There are  $n_{\text{points}} = 10$  points obtained by linear interpolation between the given start and end points, and the population size is  $N = 15$  paths. In the results related in Section III, the resolution of paths, however, is converted to rmsd [30] and reported in Å, as there is more 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 from points to structures. For the majority of the applications described in Section III, 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, with a significant portion of this time devoted to conduct  $N$  lowest-cost path searches on the induced ngraph, as described above.

### III. RESULTS

Three test cases are selected to showcase the performance of the proposed EA. They are all different (variant) human sequences of the H-Ras enzyme, which is a protein central to regulating cell growth. Mutations to H-Ras have been linked to several human cancers. Specifically, the first test case we consider here is the “healthy” variant, also referred to as the wildtype (WT). The other two variants are oncogenic forms of H-Ras. They are referred to as G12C and Q61L; in G12C, the glycine (G) amino acid at position 12 in the WT has been

mutated into a valine (V) in the oncogenic variant, and in Q61L, the glutamine (Q) amino acid at position 61 in the WT has been mutated into a leucine (L).

The EA is run to obtain paths connecting a given start structure found under entry id 1qra in the Protein Data Bank (PDB) [32] to a given end structure found under PDB entry id 4q21. These two structures correspond to two different functional states of H-Ras and are more than 2.5Å in all-atom rmsd from each-other. The start structure corresponds to the on or active state of H-Ras, and the end structure corresponds to the off or inactive state of H-Ras.

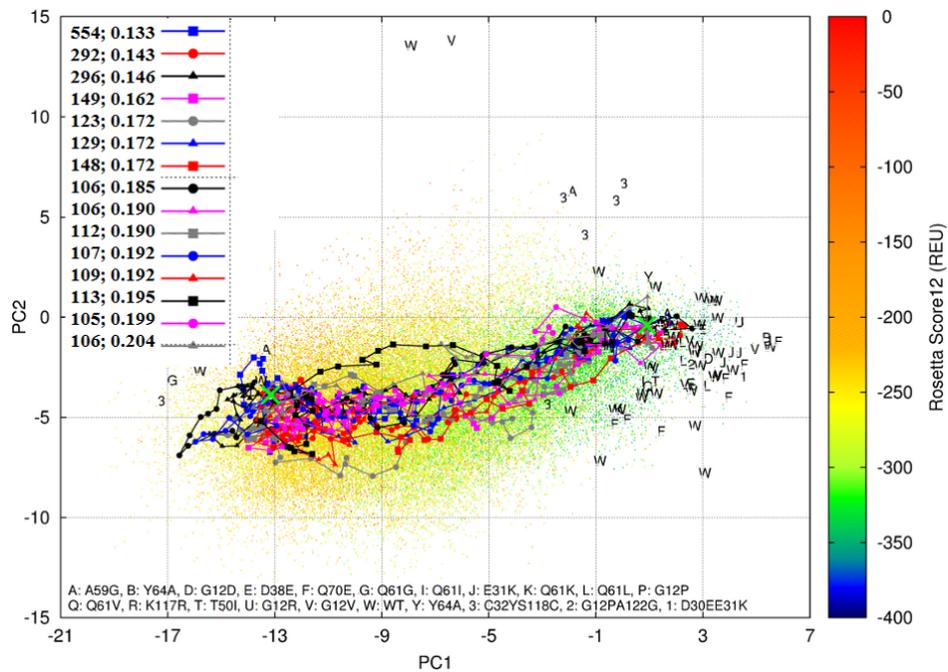
For most of the analysis on each test case, the EA is run until a computational budget of 100,000 fitness evaluations is exhausted. Three sets of results are related. First, we demonstrate the ability of the proposed EA to obtain low-cost yet high-resolution paths that are comparable in quality to a robotics-inspired, roadmap-based method that employs a computational budget of 1,000,000 fitness evaluations [28]. Second, we demonstrate that the proposed EA is able to further improve its solutions with a longer computational budget. These two sets of results focus on the H-Ras WT. In the third set of results, we show the performance of the proposed EA on the other two test cases, the G12C and the Q61L variants.

#### A. Analysis of the Performance of Path-evolving EA on H-Ras WT

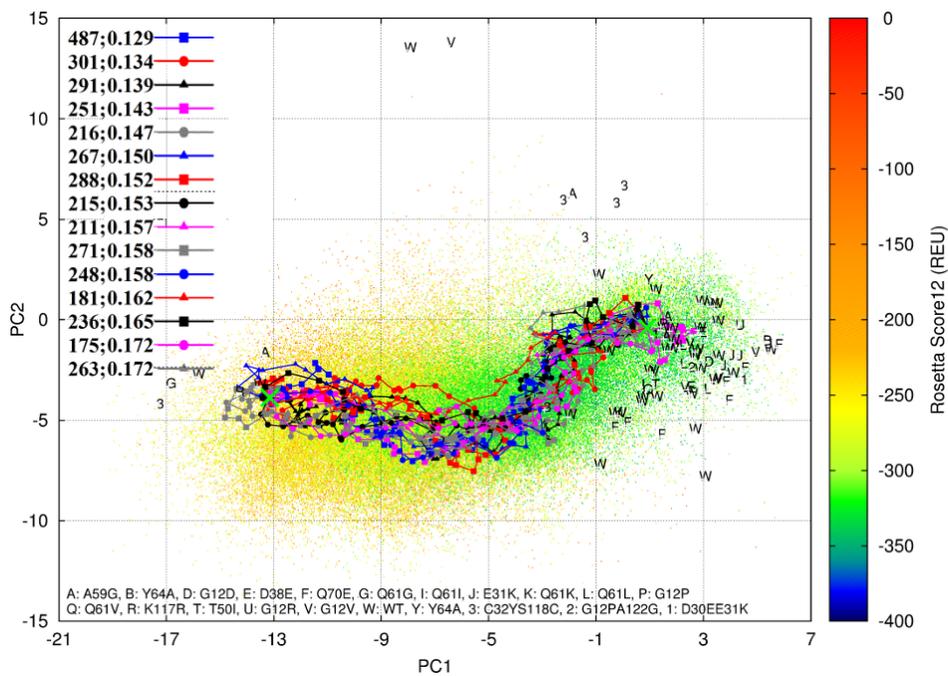
The 15 lowest-cost paths found by the proposed EA on the H-Ras WT are shown in Figure 2. Panel (a) shows these top paths when the EA is limited to a computational budget of 100,000 fitness evaluations, whereas panel (b) shows what happens with double the computational budget. The paths are drawn by connecting consecutive structures with edges. Structures are drawn as dots, projecting them onto the top two variables (the top two PCs) in the variable space. Structures generated by the algorithm but not utilized in the top 15 paths are also drawn and are color-coded in a red-to-blue color scheme that indicates high-to-low potential energy values; the latter are measured for each structure with the *score12* energy function in Rosetta [33], [34]. Since this function is not a physical-based one, the units are not physical units of kcal/mol but Rosetta energy units (REUs).

As shown in Figure 2(a), the proposed EA obtains very high-resolution (0.133Å and 0.129Å under the two budgets, respectively) paths with less computational budget (and so fewer generated structures) than the roadmap-based method in [28]; even with the modest budget of 100,000 fitness evaluations, the obtained path resolutions are better than those obtained with the roadmap-based method. Figure 2(b) also shows that, when the budget is doubled, the proposed EA is able to further improve both cost and resolution.

In Table I, we compare the quality of paths obtained by the proposed EA under each computational budget to the quality of paths obtained with the roadmap-based method



(a) Computational budget of 100,000 fitness evaluations



(b) Computational budget of 200,000 fitness evaluations

Figure 2: The EA is allowed a computational budget of (a) 100,000 or (b) 200,000 fitness evaluations on the H-Ras WT; the 15 lowest-cost paths are shown in each setting by drawing an edge between two consecutive structures in a path. Dots show PC1-PC2 projections of structures. Dots outside the drawn paths are color-coded and correspond to structures generated during the execution of the EA. The color-coding scheme on the right runs from low (blue) to high (red) energy values measured with the all-atom Rosetta energy function (*score12*). The text annotations indicate projections of experimentally-known structures detected in the wet laboratory for different variants of H-Ras). The legend in each plot lists the path costs in Rosetta Energy Units (REUs), as well as their resolutions in Å.

in [28]. Table I juxtaposes the costs and resolutions of the ten lowest-cost paths produced in each setting. The comparison is restricted to the 10 lowest-cost paths, as the resolution of the lowest-cost paths obtained by the proposed EA restricted to a computational budget of 100,000 fitness evaluations deteriorates afterwards. While all paths obtained by the roadmap-based method in [28] have the same resolution (as edges in the roadmap are only drawn when two structures are no further than a set distance), the ones obtained by the proposed EA have varying resolution, per the process described in the selection operator in Section II.

Table I: Top ten paths obtained by each algorithm.

Proposed EA 200K fitness evals		Proposed EA 100K fitness evals		Roadmap-based [28] 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 I orders paths from high to low cost and shows that the proposed 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 path resolution improves from 0.168Å (when the budget is 100,000 fitness evaluations) to 0.151Å when the budget is doubled. Visual comparison of panels (a) and (b) in Figure 2 also shows that the longer budget allows the paths to phenotypically converge to a narrower region of the fitness landscape. Their average cost, however, increases from 170.33REUs to 260.07REUs when the budget is doubled. The increase in cost is due to the *tunneling* phenomenon that we describe in Section I, where more faithfully following the landscape may result in higher, but more accurate energetic costs associated with a transition of interest. In summary, the proposed EA is able to find higher-resolution paths with even less computational budget and find that the demand on paths to follow the landscape reveals underlying energetic barriers that have to be crossed in the transition from the on to the off structure of H-Ras.

### B. Comparative Analysis on More Variants

Finally, we relate the performance of the proposed EA on the additional test cases when restricted to a budget of 100,000 fitness evaluations. Figure 3 shows the top 15 paths obtained on the G12C and the Q61L variants. The resolutions are generally worse in comparison to what the EA restricted to the same budget finds for the H-Ras WT, which helps reduce the total cost of a path, as described earlier. The resolutions obtained for these two variants

point to possibly more complex landscapes where more fitness evaluations (and more structures) may be needed. Visual comparison of the generated structures (color-coded projections) across the WT and these two variants suggests higher-energy regions separating the start and goal structures connected by the proposed EA.

## IV. CONCLUSION

We have presented a novel EA to model protein structural transitions. The latter is an important problem in computational structural biology that regularly motivates the design of novel algorithmic frameworks [31], [35]. Understanding and characterizing protein structural dynamics is central to understanding biological function in the healthy and diseased cell and devising novel therapeutics [6], [8], [12].

The EA proposed in this paper casts the problem as a multiobjective optimization one, as both energetic cost and resolution are key to the quality of a modeled structural transition. The EA makes use of important building blocks, most notably an exploitation of known structures of a protein, and relies on a novel selection operator to address both path cost and path resolution as competing objectives.

The evaluation of the proposed EA demonstrates its capability to improve over landscape-reconstructing EAs that are now considered state of the art [28] but frame the problem much differently and ignore the competing objectives. The analysis on th H-Ras WT also shows the ability of the proposed EA to further improve the quality of its solutions when allowed larger computational budgets.

The results related in this paper suggest that treating the problem of modeling protein structural transitions as a multiobjective one is indeed reasonable and can result in novel approaches likely to lead to further research. The emphasis in this paper on lower computational budgets is motivated by the potential of these frameworks to obtain and then compare the structural dynamics of various forms of a protein in a large-scale setting. The latter would allow understanding the how mutations alter structural dynamics to disrupt normal biological function.

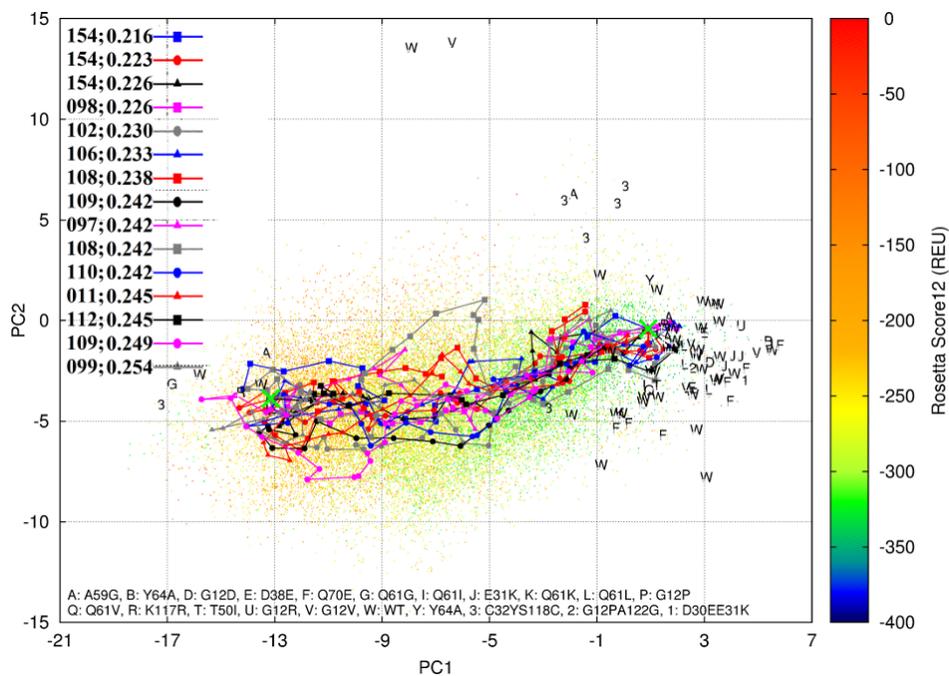
The techniques presented here are potentially useful for a broad range of problems on fitness landscapes. Evolving individuals with complex representations, such as paths, is of interest in evolutionary computation and is likely to spur further work by us and others on effective variation and selection operators.

## V. ACKNOWLEDGMENTS

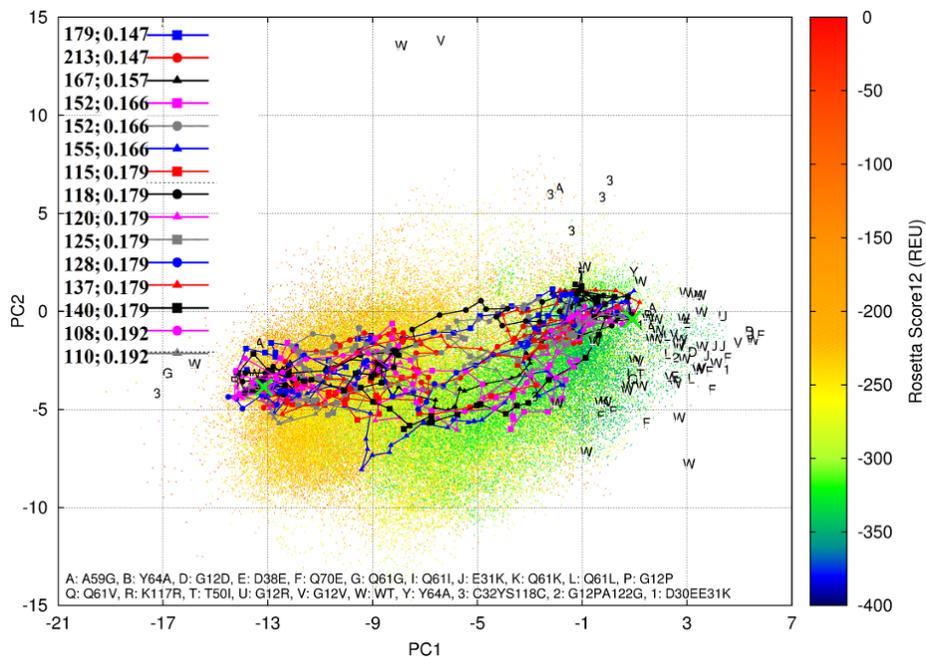
This work is supported in part by NSF CCF No. 1421001 and NSF IIS CAREER Award No. 1144106.

## REFERENCES

- [1] R. P. Feynman, "There's plenty of room at the bottom," *J Microelectromechanical Systems*, vol. 1, no. 1, Mar. 1992, transcript of a talk given by R. P. Feynman on December 26, 1959, at the Annual Meeting of the American Physical Society, held at the CA Institute of Technology.



(a) G12C



(b) Q61L

Figure 3: The top 15 paths obtained by the proposed EA restricted to a budget of 100,000 fitness evaluations are shown here for the (a) G12C and (b) Q61L variants. The same plotting style is followed as in Fig. 2.

- [2] K. Jenzler-Wildman and D. Kern, “Dynamic personalities of proteins,” *Nature*, vol. 450, no. 7172, pp. 964–972, 2007.
- [3] 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, “Intrinsic motions

- along an enzymatic reaction trajectory,” *Nature*, vol. 450, no. 7171, pp. 838–844, 2007.
- [4] O. Beckstein, E. J. Denning, J. R. Perilla, and T. B. Woolf, “Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open-closed transitions,” *J Mol*

- Biol.*, vol. 394, no. 1, pp. 160–176, 2009.
- [5] T. Sikosek and H. S. Chan, “Biophysics of protein evolution and evolutionary protein biophysics,” *J R Soc Interface*, vol. 6, no. 11, p. 20140419, 2014.
  - [6] C. Ó’Dúnláing, M. Sharir, and C. K. Yap, “Protein conformational dynamics,” in *Advances in Experimental Medicine and Biology*, K. Han, X. Zhang, and M. Yang, Eds. Switzerland: Springer International Publishing, 2014, p. 488.
  - [7] I. A. Prior, P. D. Lewis, and C. Mattos, “A comprehensive survey of Ras mutations in cancer,” *Cancer Res*, vol. 72, no. 10, pp. 2457–2467, 2012.
  - [8] F. Feixas, S. Lindert, W. Sinko, and J. A. McCammon, “Exploring the role of receptor flexibility in structure-based drug discovery,” *Biophys Chem*, vol. 186, pp. 31–45, 2014.
  - [9] W. J. Greenleaf, M. T. Woodside, and S. M. Block, “High-resolution, single-molecule measurements of biomolecular motion,” *Annu Rev Biophys Biomol Struct*, vol. 36, pp. 171–190, 2007.
  - [10] J. Hohlbein, T. D. Craggs, and T. Cordes, “Alternating-laser excitation: single-molecule FRET and beyond,” *Chem Soc Rev*, vol. 43, no. 4, pp. 1156–1171, 2014.
  - [11] A. Gall, C. Iliaia, T. P. Krüger, V. I. Novoderezhkin, B. Robert, and R. van Grondelle, “Conformational switching in a light-harvesting protein as followed by single-molecule spectroscopy,” *Biophys J*, vol. 108, no. 11, pp. 2713–2720, 2015.
  - [12] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu, “Principles and overview of sampling methods for modeling macromolecular structure and dynamics,” *PLoS Comput Biol*, vol. 12, no. 4, p. e1004619, 2016.
  - [13] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, “The energy landscapes and motion on proteins,” *Science*, vol. 254, no. 5038, pp. 1598–1603, 1991.
  - [14] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, “Funnels, pathways, and the energy landscape of protein folding: a synthesis,” *Proteins: Struct Funct Genet*, vol. 21, no. 3, pp. 167–195, 1995.
  - [15] H. Frauenfelder, P. W. Fenimore, and R. D. Young, “Protein dynamics and function: Insights from the energy landscape and solvent slaving,” *IUBMB Life*, vol. 59, no. 8-9, pp. 506–512, 2007.
  - [16] A. Shehu, “Probabilistic search and optimization for protein energy landscapes,” in *Handbook of Computational Molecular Biology*, S. Aluru and A. Singh, Eds. Chapman & Hall/CRC Computer & Information Science Series, 2013.
  - [17] M. Karplus and J. Kuriyan, “Molecular dynamics and protein function,” *Proc Nat Acad Sci*, vol. 102, no. 19, pp. 6679–6685, 2005.
  - [18] Y. Okamoto, “Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations,” *J Mol Graph Model*, vol. 22, no. 5, pp. 59–64, 2004.
  - [19] D. Manocha and Y. Zhu, “Kinematic manipulation of molecular chains subject to rigid constraints,” in *Intl Conf Intell Sys Mol Biol (ISMB)*, R. B. Altman, D. L. Brutlag, P. D. Karp, R. H. Lathrop, and D. B. Searls, Eds., vol. 2. Stanford, CA: AAAI, 1994, pp. 285–293.
  - [20] A. Shehu and E. Plaku, “A survey of computational treatments of biomolecules by robotics-inspired methods modeling equilibrium structure and dynamics,” *J of Artif Intell Res*, vol. 597, pp. 509–572, 2016.
  - [21] T. Maximova, E. Plaku, and A. Shehu, “Computing transition paths in multiple-basin proteins with a probabilistic roadmap algorithm guided by structure data,” in *Intl Conf on Bioinf and Biomed (BIBM)*. Washington, D.C.: IEEE, 2015, pp. 35–42.
  - [22] —, “Structure-guided protein transition modeling with a probabilistic roadmap algorithm,” *IEEE/ACM Trans Comput Biol & Bioinform*, vol. 13, no. 5, pp. 1–14, 2016.
  - [23] K. Molloy and A. Shehu, “Elucidating the ensemble of functionally-relevant transitions in protein systems with a robotics-inspired method,” *BMC Struct Biol*, vol. 13, no. Suppl 1, p. S8, 2013.
  - [24] —, “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. Norfolk, VA: Springer International Publishing, 2015, pp. 175–186.
  - [25] K. Molloy, R. Clausen, and A. Shehu, “A stochastic roadmap method to model protein structural transitions,” *Robotica*, vol. 34, no. 8, pp. 1705–1733, 2016.
  - [26] E. Sapin, K. A. De Jong, and A. Shehu, “From optimization to mapping: An evolutionary algorithm for protein energy landscapes,” *IEEE/ACM Trans Comput Biol & Bioinform*, 2017, in press.
  - [27] —, “A novel ea-based memetic approach for efficiently mapping complex fitness landscapes,” in *Conf on Genetic and Evolutionary Computation (GECCO)*. ACM, 2016, pp. 85–92.
  - [28] E. Sapin, D. B. Carr, K. A. De Jong, and A. Shehu, “Computing energy landscape maps and structural excursions of proteins,” *BMC Genomics*, vol. 17, no. Suppl 4, p. 456, 2016.
  - [29] L. Jaillet, J. Cortés, and T. Siméon, “Sampling-based path planning on configuration-cost costmaps,” *IEEE Trans Robot*, vol. 26, no. 4, pp. 635–646, 2010.
  - [30] A. D. McLachlan, “A mathematical procedure for superimposing atomic coordinates of proteins,” *Acta Crystallogr A*, vol. 26, no. 6, pp. 656–657, 1972.
  - [31] R. Clausen, B. Ma, R. Nussinov, and A. Shehu, “Mapping the conformation space of wildtype and mutant H-Ras with a memetic, cellular, and multiscale evolutionary algorithm,” *PLoS Comput Biol*, vol. 11, no. 9, p. e1004470, 2015.
  - [32] H. M. Berman, K. Henrick, and H. Nakamura, “Announcing the worldwide Protein Data Bank,” *Nat Struct Biol*, vol. 10, no. 12, pp. 980–980, 2003.
  - [33] K. W. Kaufmann, G. H. Lemmon, S. L. DeLuca, J. H. Sheehan, and J. Meiler, “Practically useful: What the rosetta protein modeling suite can do for you,” *Biochemistry*, vol. 49, no. 14, pp. 2987–2998, 2010, PMID: 20235548. [Online]. Available: <http://pubs.acs.org/doi/abs/10.1021/bi902153g>
  - [34] 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. Popovi, and et. al., “ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules,” *Methods Enzymol*, vol. 487, pp. 545–574, 2011.
  - [35] R. Clausen and A. Shehu, “A data-driven evolutionary algorithm for mapping multi-basin protein energy landscapes,” *J Comp Biol*, vol. 22, no. 9, pp. 844–860, 2015.