An Evolutionary Algorithm to Model Structural Excursions of a Protein

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

Excursions of a protein between different structures at equilibrium are key to its ability to modulate its biological function. The energy landscape, which organizes structures available to a protein by their energetics, contains all the information needed to characterize and simulate structural excursions. Computational research aims to uncover such excursions to complement wet-laboratory studies in characterizing protein equilibrium dynamics. Popular strategies adapt the robot motion planning framework and construct full or partial, structured representations of the energy landscape. In this paper, we present a novel, complementary approach based on evolutionary computation. We propose an evolutionary algorithm that evolves path representations of a specific structural excursion without *a priori* construction of the energy landscape. Preliminary applications on healthy and pathogenic variants of a protein central to human health are promising and warranting further investigation of evolutionary search techniques for modeling protein structural excursions.

CCS CONCEPTS

•Applied computing → Molecular structural biology; Bioinformatics;

KEYWORDS

computational structural biology, protein modeling, structural dynamics, energy landscape, stochastic optimization, evolutionary algorithm

ACM Reference format:
Emmanuel Sapin, Kenneth De Jong, and Amarda Shehu. 2017. An Evolutionary Algorithm to Model Structural Excursions of a Protein. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 5 pages.
DOI: http://dx.doi.org/10.1145/3067695.3082544

*Corresponding Author

GECCO '17 Companion, Berlin, Germany

DOI: http://dx.doi.org/10.1145/3067695.3082544

1 INTRODUCTION

Proteins are inherently dynamic systems that harness their equilibrium structural dynamics into productive events in the cell [11]. Improvements in both dry- and wet-laboratory technologies have yielded evidence of proteins using often markedly different three-dimensional structures to stick with different molecular partners, thus modulating their biological function [1, 6, 10]. Obtaining a detailed view of the structural excursions that proteins employ to modulate their function is the subject of much research in molecular biology, as it promises to improve our understanding of molecular mechanisms in living and healthy cell and even suggest directions for design of molecular therapeutics [7].

While great strides are being made by wet-laboratory single-molecule techniques, no single wet- or dry-laboratory technique can currently provide a detailed, structure-by-structure characterization of excursions independent of protein size and biological timescale [15]. The challenge lies in the disparate spatial and temporal scales involved; the start and end structures of an excursion of interest may differ by several angstroms (Å), and the excursion may take several micro- or even milli-seconds. Yet, these challenges motivate protein modeling research to complement wet-laboratory studies in uncovering functionally-relevant structural excursions that a protein employs to tune its biological function in the living cell [15].

In principle, the energy landscape, which organizes the structures available to a protein by their energetics, contains all the information needed to characterize and simulate structural excursions [3, 9]. Yet, energy landscapes are vast, high-dimensional, and rich in local minima [8, 21]. The latter provide pitfalls for computational approaches that rely on numerical integration and Newton's second law of motion to follow motions of the atoms that comprise a molecular system. Computational strategies using this approach, also known as Molecular Dynamics (MD), are regularly proposed [15]. Other similar efforts are undertaken in the complementary, Monte Carlo (MC) approach, which explores the energy landscape of a protein one biased random walk at a time [17].

One category of efforts, referred to as robotics-inspired, essentially organize MC walks in trees or graphs to enhance the exploration capability. These methods were originally inspired by mechanistic analogies between molecular motions and motions of articulated kinematic linkages [14]. A more informative summary of these methods is that they aim to construct a structured representation of an energy landscape.

Robotics-inspired methods can be classified as tree- or roadmap-based. Tree-based methods build partial representations to model

an excursion of interest. The representation provides a highly local view of the energy landscape insofar as it helps provide a constraint-satisfying series of structures (*a path*) that connect a given start to a given goal structure; the constraints can be energetic and/or distance-based. Roadmap-based methods build more comprehensive representations that provide a non-local view of the landscape. In principle, these methods can model many excursions among more than two structures of interest. The non-local view, however, comes at a higher computational cost.

A recent roadmap-based approach, which is the subject of our comparative analysis in Section 3 first reconstructs the energy landscape of a protein with a powerful memetic EA (making use of several building blocks developed over the years [4, 5, 19, 20]) and then exploits a graph-based representation of the landscape to answer path queries corresponding to structural excursions of interest [18]. This direction has revealed key insights on many proteins but has a large computational footprint due to the need to construct comprehensive and detailed representations of energy landscapes [18]. The reader is referred to Ref. [22] for a review of robotics-inspired methods for biomolecular dynamics.

In this paper we present a novel, complementary method that approaches the problem of modeling a structural excursion of a protein of interest under the umbrella of evolutionary computation. The method is an evolutionary algorithm (EA), and it falls between tree-based and roadmap-based methods. Unlike a robotics-inspired tree-based method, the EA proposed in this paper computes many paths that connect a given start to goal structure approaching the problem of modeling structural excursions as a path optimization problem. Unlike a robotics-inspired roadmap-based method, the EA does not construct a protein's energy landscape *a priori*, thus reducing the computational burden of modeling a specific excursion of interest. The EA leverages several algorithmic innovations in prior work by us on effective, low-dimensional representations of structures of medium-size proteins, but employs here novel evolutionary operators. Preliminary applications on healthy and pathogenic variants of a protein central to human health are promising and warranting further investigation of evolutionary search techniques for modeling protein structural excursions.

2 METHODS

The EA presented here evolves a population of paths directly, exploits experimentally-known structures of a protein in its initialization, and makes use of novel selection and crossover operators. Key building blocks in the proposed path-evolving EA have been developed and analyzed in prior work [18–20]. They include exploiting known structures of a protein (of healthy and diseased sequence variants) to extract a lower-dimensional variable space for exploration. While details can be found in prior work, in summary, the experimentally-known structures of a protein of interest (its sequence and similar sequence variants) are subjected to a principal component analysis to extract a new basis whose axes are ordered by the amount of variance they capture in the structure data. These axes are also referred to as principal components (PC). A subset of them that captures more than 90% of the variance but represent over a hundred-fold reduction over the number of Cartesian coordinates of the protein under investigation are selected as variables, and each new computed structure is effectively represented as a point in this variable space. The reader is directed to Ref. [5] for further details.

Unlike prior work, in which an EA evolves individuals in this variable space, starting from a collection of individuals (points) corresponding to known structures, the new EA evolves paths utilizing only two given (experimentally-known) structures. These given structures initialize the start and end points for all sought paths in the variable space. Specifically, a path individual is represented as a (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. In addition to cost, a second quantity is associated with each path, its resolution. The resolution of a path is the maximum Euclidean distance between two consecutive points in the path.

Initially, $n_$points are obtained by linear interpolation between the given start and end points. Each obtained point undergoes a transformation, which effectively converts it to an all-atom protein structure corresponding 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 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 prior work, and we direct the reader to Ref. [4] for details.

Once the initial population of paths is defined, successive generations evolve 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.

All existing (of paths in the current population) and resulting points (new points yielded by the variation operator) are then inserted into a nearest-neighbor graph (nngraph) which connects a point to others within a ball of radius ϵ; the radius ϵ is measured via the Euclidean distance in the variable space. After initialization, this value undergoes increase or decrease as follows. Dijkstra's algorithm is invoked on the nngraph to obtain the lowest-cost path. All internal points of the path are removed, and Dijkstra's algorithm is invoked again in the induced subgraph; prior to any application of Dijkstra's algorithm, ϵ is increased from its current value until the start and end points are in a connected component in the induced subgraph. This process of applying Dijkstra's algorithm is repeated until N lowest-cost paths are obtained to initialize the next generation. In addition, ϵ is subjected to a gradual decrease over the generations so that the algorithm evolves both low-cost and high-resolution paths that faithfully follow the actual energy landscape.

The described process for selecting the paths for the next generation is a novel mechanism in several ways. First, the paths in a population are effectively rewired, and a graph structure is used to centralize the view. Second, novel points are added onto the graph to increase diversity. This selection mechanism circumvents the issue of comparing two paths to determine which one is better. Comparing two paths is not trivial, as (lower) path cost is not the only consideration. The resolution of a path is also very important.

Cost and resolution are competing optimization objectives, but the adversarial relationship is non-trivial. A path with low resolution may have low cost because it may "dig a tunnel" or "draw a bridge" to connect two structures rather than follow the curvature of the landscape. Such considerations are ignored in existing robotics-inspired methods, particularly in tree-based ones, but the proposed EA addresses both cost and resolution in its selection operator.

2.1 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 $n\_points = 10$ points obtained by linear interpolation between the given start and end points, and the population size is $N = 15$ paths. The initial value for the ϵ radius is set to d/n_points, with d measuring the Euclidean distance between the start and goal points (corresponding to projections of the known start and goal structures in the variable space). In the results related in Section 3, the resolution of paths, however, is related in terms of root-mean-square-deviation (rmsd) [16] and reported in Å, as there is more domain-specific insight on what rmsd values correspond to low or high resolutions. 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 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, with a significant portion of this time devoted to conduct N lowest-cost path searches on the induced nngraph.

3 RESULTS

The performance of the path-evolving EA is showcased here on several variant sequences of the H-Ras enzyme, which is a protein central to cell growth and various human cancers. Specifically, we consider here the healthy variant, also referred to as the wildtype (WT) from now on, and two oncogenic variants, G12C and Q61L. The naming convention indicates in which position the variant differs from the WT sequence; for instance, G12C indicates that in this variant, the W amino acid at position 12 in the WT sequence is mutated to the C amino acid.

The path-evolving EA is run to obtain paths connecting two known structures of H-Ras that correspond to two different functional states. Specifically, the two structures are those deposited in the Protein Data Bank (PDB) [2] under PDB ids 1qra and 4q21, corresponding to the on/active and the off/inactive state of H-Ras, respectively. For each of the test cases, the computational budget is fixed to 100,000 fitness evaluations, which is 10 times less than that used in prior work that first reconstructs energy landscapes with an EA and then uses graph-based representations to answer specific path queries [18].

The results are organized as follows. We first demonstrate that even with a much more modest budget, the path-evolving EA is able to find among its top solutions paths that have similar energetic costs and similar resolutions as the more computationally-demanding landscape-reconstructing EA from previous published work. We additionally demonstrate that with further computational resources, the path-evolving EA is able to further improve its solutions. This detailed analysis focuses on the WT H-Ras sequence. We then relate the performance of the path-evolving EA on the other two variants of H-Ras.

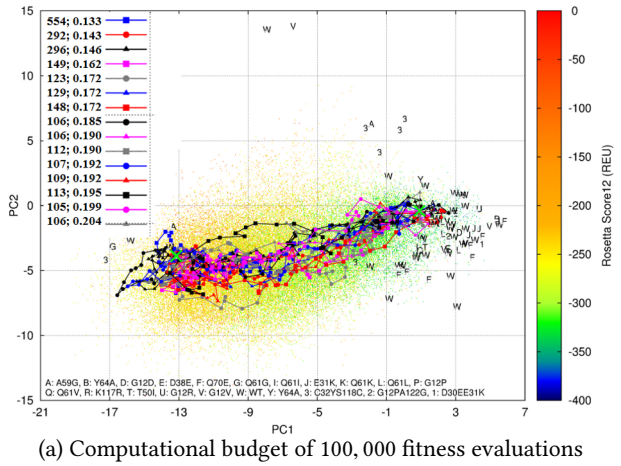
3.1 Analysis of the Performance of Path-evolving EA on H-Ras WT

Figure 1 draws the top 15 paths found by the path-evolving EA on the WT H-Ras. Panel (a) shows the results obtained when the algorithm is limited to a computational budget of 100,000 fitness evaluations. The paths are drawn by connecting consecutive structures with edges. Structures are drawn as dots, projecting them onto the top two variables in the variable space. As described briefly in Section 2, the variables are principal component (PC) axes extracted from analysis of experimentally-known structures of a protein's healthy and pathogenic variants. The color-coding in Figure 1 additionally conveys the energy of each structure, with a red-to-blue color scheme tracking high-to-low energy values measured with the *score12* energy function in Rosetta [12, 13].

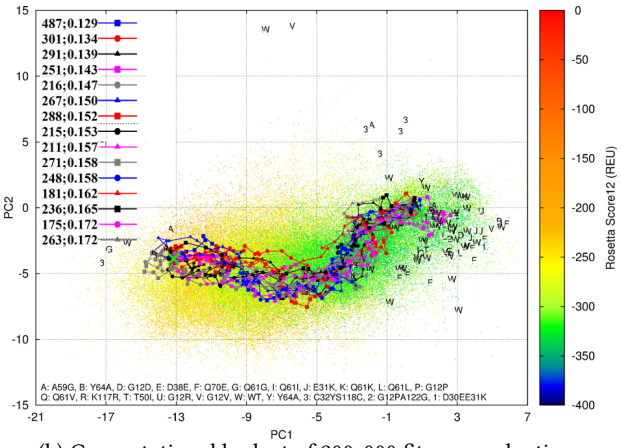
Figure 1(a) shows that even a modest budget of 100,000 fitness evaluations reaches low costs and high resolutions. Figure 1(b) shows that the quality of the paths when the path-evolving EA is provided with a longer computational budget of 200,000 fitness evaluations improves even further (higher resolutions). Visual comparison shows that the longer budget allows the paths to phenotypically converge to a narrower region of the fitness landscape, as well, but their average cost increases from 170.33REUs to 260.07REUs when the budget is doubled. We describe this phenomenon as *tunneling*; paths with higher resolution do not follow the curvature of the landscape but instead connect by an edge two structures that may be far apart to one another. Doing so underestimates the true energetic cost; rather than climbing a mountain, such paths draw tunnels through an energetic mountain in the landscape.

Table 1 enhances the analysis by comparing the path-evolving EA to the landscape-reconstructing EA [18] of prior work by juxtaposing the costs and resolutions of the ten lowest-cost paths produced by each algorithm; the performance of the path-evolving EA is shown under both computational budgets of 100,000 and 200,000 fitness evaluations (respectively 1/10 and 1/5 of the budget used by the landscape-reconstructing EA [18], respectively). We restrict the comparison to the 10 lowest-cost paths, as the resolution of the lowest-cost paths obtained by the path-evolving EA restricted to a computational budget of 100,000 fitness evaluations deteriorates afterwards. While all paths obtained by post-analysis of the reconstructed map in prior work have the same resolution, the ones obtained by the path-evolving EA have varying resolution.

The juxtaposition in Table 1, where paths are ordered from high to low energy, shows that the path-evolving EA is able to obtain high-resolution (0.133Å and 0.129Å under the two budgets) paths with much less computational budget (and consequently fewer actual structures). These resolutions are better than the ones obtained with the landscape-reconstruting EA. As also related earlier, path costs at high resolution typically increase due to the high ruggedness of protein energy landscapes, but the lowest costs at similar resolutions are similar; for comparison, the lowest-cost path reported by the landscape-reconstructing EA has a cost of 266REUs



(a) Computational budget of 100, 000 fitness evaluations



(b) Computational budget of 200, 000 fitness evaluations

Figure 1: The path-evolving is allowed a budget of (a) 100, 000 **or (b)** 200, 000 **fitness evaluations; the** 15 **lowest-cost paths are shown by drawing an edge between two consecutive structures. Dots show 2D projections of structures over the top two variables (described in Section 2). Dots outside the drawn paths are color-coded and note structures generated during the execution of the algorithm. 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, with WT referring to the structures detected in the wet laboratory for the healthy form/variant of the protein, and others to pathogenic forms (variants found in cancer and other disorders). The legend in each plot lists the path costs in Rosetta energy units (REUs) and their resolutions in angstroms (Å).**

(all paths have a resolution of 0.145Å). The costs of paths found at around that resolution (0.143-0.146Å) from the path-evolving EA with a budget of 100, 000 fitness evaluations also have similar costs (292 and 296REUs). At a similar resolution (0.143), the path-evolving EA with twice the budget finds a lower-cost path of 251REUs. As also related earlier, the path-evolving EA with a higher computational budget but still about 1/5 of the budget afforded to

Table 1: Top ten paths obtained by each algorithm.

Path-evolving EA 200K fitness evals		Path-evolving EA 100K fitness evals		EA [18] 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

the landscape-reconstructing EA reports much better resolutions (as low as 0.129Å) that invariably increase the costs of paths due to the ruggedness of the landscape. Moreover, the average value over the resolution of of the paths improves from 0.168Å (when the budget is 100, 000 fitness evaluations) to 0.151Å when the budget is doubled.

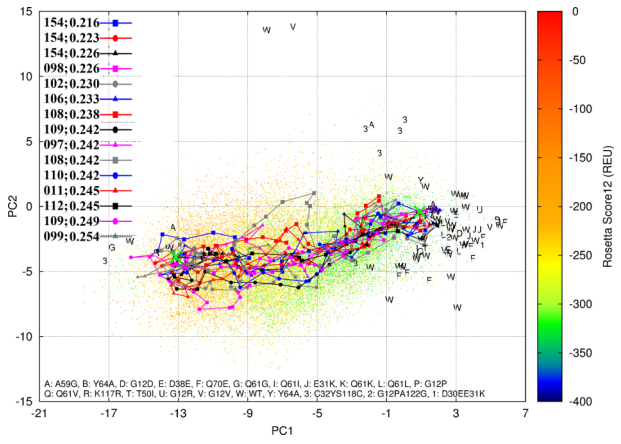
3.2 Comparative Analysis on More Variants

Figure 2 shows the top 15 paths obtained by the path-evolving EA on two more test cases that are oncogenic variants of H-Ras. The resolutions are generally worse in comparison with what the path-evolving EA can find for the H-Ras WT, which helps reduce the total cost of a path, as described earlier. The resolutions obtained for the Q61L variant are better than those obtained for the G12C, pointing to possibly more complex landscapes (and even more so for G12C) where more fitness evaluations may be needed to uncover local minima in higher-energy regions. 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 in the G12C variant than the Q61L variant.

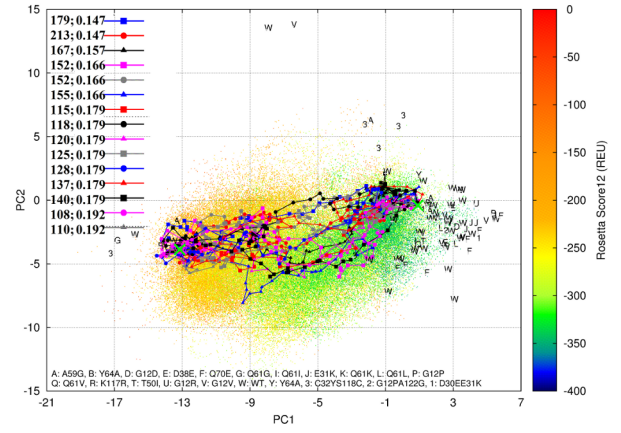
4 CONCLUSION

The evaluation of this path-evolving EA suggests that it represents a complementary approach to uncovering structural excursions in proteins of interest without relying on an *a priori* reconstruction of energy landscapes. The results provided on the H-Ras WT demonstrate the capability of the proposed EA to improve over such landscape-reconstructing EAs that are now considered state of the art [18], as well as the ability of the proposed EA to further improve the quality of its solutions when allowed larger computational budgets. The results on the two other pathogenic variants of H-Ras suggest that there may be complex landscapes where larger computational budgets may needed to uncover the true energetic costs of specific structural excursions.

The results related in this paper warrant further research on frameworks for modeling protein structural dynamics with reasonable computational budgets. The emphasis on lower computational budgets in this paper 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



(a) G12C



(b) Q61L

Figure 2: The top 15 paths obtained by the path-evolving EA with a budget of 100,000 fitness evaluations is shown here for two more test cases, the G12C and the Q61L variants of H-Ras. The same plotting style is followed, as in Fig. 1.

allow understanding the impact of mutation-altered dynamics on protein function.

It is worth noting that the techniques presented here are more general than the specific domain of protein modeling and thus potentially useful for a broad range of problems on fitness landscapes. In addition, 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 operators and selection mechanisms in such settings.

5 ACKNOWLEDGMENTS

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

REFERENCES

[1] Oliver Beckstein, Elizabeth J. Denning, Juan R. Perilla, and Thomas B. Woolf. 2009. Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open-closed transitions. *J Mol Biol* 394, 1 (2009), 160–176.

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

[3] 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.

[4] 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.

[5] 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.

[6] J. Evenäs, S. Forsén, A. Malmendal, and M. Akke. 1999. Backbone dynamics and energetics of a calmodulin domain mutant exchanging between closed and open conformations. *J Mol Biol* 289, 3 (1999), 603–617.

[7] F. Feixas, S. Lindert, W. Sinko, and J. A. McCammon. 2014. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys Chem* 186 (2014), 31–45.

[8] H. Frauenfelder, P. W. Fenimore, and R. D. Young. 2007. Protein Dynamics and Function: Insights from the Energy Landscape and Solvent Slaving. *IUBMB Life* 59, 8-9 (2007), 506–512.

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

[10] 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.

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

[12] Kristian W. Kaufmann, Gordon H. Lemmon, Samuel L. DeLuca, Jonathan H. Sheehan, and Jens Meiler. 2010. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. *Biochemistry* 49, 14 (2010), 2987–2998. DOI : https://doi.org/10.1021/bi902153g arXiv:http://pubs.acs.org/doi/pdf/10.1021/bi902153g PMID: 20235548.

[13] 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.

[14] D. Manocha and Y. Zhu. 1994. 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. AAAI, Stanford, CA, 285–293.

[15] 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.

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

[17] Y. Okamoto. 2004. Generalized-ensemble algorithms: Enhanced sampling techniques for Monte Carlo and molecular dynamics simulations. *J Mol Graph Model* 22, 5 (2004), 59–64.

[18] 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.

[19] 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.

[20] 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.

[21] A. Shehu. 2013. 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.

[22] 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.