Kino-Geometric Modeling: Insights into Protein Molecular Mechanisms

Dominik Budday1,*, Sigrid Leyendecker1,**, and Henry van den Bedem2,3,***

1 Chair of Applied Dynamics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany
2 Biosciences Division, SLAC Nat. Acc. Lab., Stanford University, Menlo Park, CA 94025, USA
3 Department of Bioengineering and Therapeutic Sciences, UCSF, San Francisco, CA 94158, USA

Proteins are dynamic macromolecules that perform an immense variety of biological functions on a broad range of spatiotemporal scales. Their conformational ensemble is a fundamental determinant of functionality in health and disease. While computational advances have increasingly enabled the computation of atomically detailed trajectories from Molecular Dynamics (MD) simulations, there remain considerable drawbacks when aiming for fast, yet elaborate insights into the molecular mechanisms of function. Here, we explore the potential of kinematics and geometry based methods, inspired from traditional robotics, to study protein conformational dynamics. Using geometric tools, we demonstrate insights into molecular mobility from instantaneous rigidity and flexibility analysis on selected example systems. Resulting motions from kinematically sampling along collective degrees of freedom show qualitative and quantitative agreement with motions from MD simulations. Coupled to sophisticated motion planning strategies, our approach is capable of providing structural ensemble representations from sparse experimental data such as double electron-electron resonance (DEER) that remain difficult to interpret otherwise. Overall, we establish our Kino-Geometric Sampling tool KGS as an efficient alternative to obtain high-level insights into molecular mechanisms across scales, with ample applications in protein design and human health.

1 Introduction

Proteins are fundamental to life on our planet. They are dynamic biomolecules that fold into a complex three-dimensional structure, usually termed their native conformation. This conformation, together with dynamic, structural rearrangements governed by the underlying free-energy surface, dictates their biological functions and interactions with different binding partners. The efficient and reliable characterization of the spatial and temporal characteristics of this conformational ensemble is critical to decipher the molecular mechanisms of health and disease, which can inform on pathways for therapeutic intervention. Molecular Dynamics (MD) based methods can provide atomically detailed trajectories, but often face limitations regarding spatial and temporal scales of protein motion due to the enormous computational resources necessary. Our method takes an alternative, kinematics based route towards high-level insights into the molecular mechanisms of function.

2 The Kino-Geometric Sampling framework

Our Kino-Geometric Sampling (KGS) and modeling approach originates in traditional robotics and efficiently captures small- and large-scale collective motions in the molecule. It models dihedral (torsional) angles of rotatable, covalent single bonds as degrees of freedom and non-covalent interactions such as hydrogen bonds or hydrophobics as constraints [1–4]. Thereby, our model takes advantage of the collective, coordinated motions of dihedral angles that entirely preserve the global constraint network [1], or perturb it only slightly in a hierarchical manner [2]. Our approach is inherently time independent and can provide insights on various levels into molecular stability and flexibility important for protein function. To graphically display the capabilities of KGS, we use the well-studied example of Adenylate Kinase (ADK), an important enzyme which undergoes large opening and closing motions during its catalytic cycle, associated with maintaining energy levels in our cell (see e.g. [5]).

Starting from a protein or RNA structure as deposited in the protein data bank (PDB) (Fig. 1, top center), KGS automatically identifies significantly strong hydrogen bonds and hydrophobic interactions, based on geometric and energetic criteria typically used in the protein rigidity community. On an instantaneous level, we can then identify larger rigid substructures that emerge from this covalent and non-covalent interaction network when only motions that respect constraints are accessible (Fig. 1, top left [1]). Individual colors denote collectively moving substructures in ADK that stabilize the enzyme, but still allow for its characteristic domain opening and closing motions. When extending the analysis towards motions that increasingly perturb the constraint network (Fig. 1, top right), we observe more global motions spanning the entire protein [2]. The overall motion pattern, colored by atomic root-mean-squared fluctuations (RMSF) increasing from blue to red, matches the underlying enzyme domains: the stable CORE domain (mostly blue), and the two flexible LID (left) and NMP (right) binding domains (green/red) [5]. Both analyses are instantaneous in nature and report these global stability and flexibility measures in seconds.

To analyze entire molecular conformation spaces and potential activation pathways in more detail, we have integrated various sophisticated motion planning strategies in KGS. Our algorithms are based on a rapidly exploring random tree (RRT [6]) that is guided by the principal of minimal frustration to automatically explore energetically accessible regions of conformation

* Corresponding author: e-mail dominik.budday@fau.de
** e-mail sigrid.leyendecker@fau.de
*** e-mail vdbedem@stanford.edu

© 2019 The Authors Proceedings in Applied Mathematics & Mechanics published by Wiley-VCH Verlag GmbH & Co. KGaA Weinheim
space. It does so by forming dynamic Clash-avoiding Constraints (dCC) whenever two atoms are in close contact, biasing exploration towards more favorable directions. For example, providing the closed form of ADK as a second structural input (Fig. 1, bottom center), KGS is capable of identifying a connective transition pathway between the open and closed state (Fig. 1, bottom left) in a matter of hours [3]. When comparing RMSF from this dCC-RRT pathway with an MD trajectory (Fig. 1, center left), we find striking qualitative and quantitative agreement [4]. In comparison with unbiased, unguided sampling around the open state alone (blue line), we observe nearly identical trends, but much less amplitude in motion. Thus, clever motion planning is crucial to overcome obstacles and broadly cover conformation space, while our underlying kino-geometric model captures stability and flexibility characteristics similar to MD, at much lower cost and much higher speed.

Finally, we are extending our current framework towards integration and structural interpretation of sparse experimental data, as obtained from double electron-electron resonance (DEER) or Förster resonance energy transfer (FRET) experiments. These data provide distance distribution between selected probe pairs in solution, thereby sparsely characterizing the dynamics between different stabilized states (see e.g. [7]). However, their sparsity demands for complementary analysis to defuse the lack of a structural basis. Integrating FRET/DEER probe distances as target distance distributions of selected atom pairs in KGS will allow us to structurally characterize associated excited states in detail, potentially delivering new insights into important drug targets. We can further illuminate the limitations of DEER/FRET and help select informative probe pairs to generate better data. As shown on the example of ADK (Fig. 1, bottom right), at least three well-selected probe pairs are necessary to obtain a close representation of its open state. Otherwise, conformational flexibility of the protein, as well as flexibility of the attached probe itself, prevent reliable identification of the correct protein fold.

Overall, our results establish kino-geometric protein modeling and sampling via KGS as an efficient alternative to obtain high-level insights into molecular mechanisms across scales. Setting up a computational pipeline for protein and RNA dynamics, these results may serve as proficient input to fine-tune analysis using more detailed, costly methods such as MD, with broad applications in protein engineering, drug design, and human health. Source code, scripts and examples are freely available for download at https://github.com/ExcitedStates/KGS.

References

[1] D. Budday, S. Leyendecker, and H. van den Bedem, J. Mech. Phys. Solids 83, 36–47 (2015).
[2] D. Budday, S. Leyendecker, and H. van den Bedem, J. Chem. Inf. Model. 58(10), 2108–2122 (2018).
[3] D. Budday, R. Fonseca, S. Leyendecker, and H. van den Bedem, Proteins: Struct. Funct. Bioinf. 85(10), 1795–1807 (2017).
[4] D. Budday, High-Dimensional Robotics at the Nanoscale - Kino-Geometric Modeling of Proteins and Molecular Mechanisms, PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2019.
[5] O. Beckstein, E. J. Denning, J. R. Perilla, and T. B. Woolf, J. Mol. Biol. 394(1), 160–176 (2009).
[6] S. M. LaValle, Rapidly-exploring random trees: A new tool for path planning, Tech. Rep. 98-11, Iowa State Univ., 1998.
[7] B. Carrington, W. K. Myers, P. Horanyi, M. Calmiano, and A. D. Lawson, Biophys. J. 113(2), 371–380 (2017).