2023 American Control Conference (ACC), San Diego, CA
May 31–June 2 2023

Prediction of Protein Folding Pathways under Entropy-Loss Constraints using Quadratic Programming-Based Nonlinear Control

Alireza Mohammadi† and Mark W. Spong

***Abstract*— This paper investigates the problem of prediction of protein molecule folding pathways under entropy-loss constraints by formulating a control synthesis problem whose solutions are obtained by solving large-scale quadratic programming (QP) optimizations with nonlinear constraints. The utilized *non-iterative* and computationally efficient algorithm, which is based on solving generalized eigenvalue problems, prevents an unpredictable and potentially large number of iterations at each protein conformation for computing the folding control inputs. The synthesized control inputs remain close to the renowned kinetostatic compliance method (KCM) reference vector field while satisfying proper *quadratic inequality constraints* that limit the rate of molecule entropy-loss during folding.**

I. INTRODUCTION

Algorithmic prediction of the *pathways* through which protein folding process takes place hold the key to computer-aided drug discovery [1]. The promising framework of KCM addresses the high computational burden of the physics-based approaches by modeling protein molecules as a large number of rigid nano-linkages folding under the effect of interatomic forces [2]. Despite the significance of entropy-loss constraints during folding, there is a gap in knowledge of encoding these constraints in the KCM framework through a fast and numerically stable algorithm. Based on our earlier results in [3], we have demonstrated that the entropy-loss constraints can be encoded into the KCM-based folding dynamics by using control inputs that guide the folding simulation. These control inputs are generated by solving a large-scale QP that constrains a physically meaningful norm of the control input. To obtain folding control input vectors in our prior work [3], we used interior point algorithms for solving *box-constrained* large-scale QPs. However, encoding such constraints via polytopic inequalities will result in loss of accuracy in predicting the folding pathways since the conventional methods for representing the protein configurational entropy rely on quadratic representations using ellipsoidal norms.

In this paper, we address the aforementioned shortcoming by utilizing a *non-iterative* and computationally efficient algorithm [4], which is based on solving generalized eigenvalue problems. The computed folding control inputs remain close to the renowned KCM reference vector field while satisfying proper *quadratic inequality constraints* that guarantee a constrained rate of protein molecule entropy-loss.

A. Mohammadi is with the University of Michigan, Dearborn. M. W. Spong is with the University of Texas at Dallas. This work is supported by the National Science Foundation (NSF) under Award CMMI-2153744 and Award CMMI-2153901. † Corresponding author: A. Mohammadi, Email: amohmmad@umich.edu

II. TRS-BASED KCM

In contrast to our prior work [3], the proper way of synthesizing folding control input vectors that are close to the KCM reference vector field and satisfy entropy-loss constraints is to solve for the QP

$$\mathbf{u}_c^*(\boldsymbol{\theta}) = \underset{\mathbf{u}}{\operatorname{argmin}} \left\{ \frac{1}{2}\mathbf{u}_c^\top \mathbf{Q}\mathbf{u}_c + \mathbf{G}^\top(\boldsymbol{\theta})\mathbf{u}_c \right\}, \quad \text{subject to} \quad \|\mathbf{u}_c\|_{\mathbf{B}} \leq \Omega. \tag{1}$$

The QP in (1) is a trust-region subproblem (TRS), which is conventionally solved using iterative algorithms requiring an unpredictable and potentially large number of iterations at each conformation. In this paper, we address the aforementioned shortcoming by utilizing a *non-iterative* algorithm [4] based on solving generalized eigenvalue problems.

Figure 1 demonstrates the numerical simulation results associated with folding of the backbone chain of a molecule with 10 peptide planes (i.e., $N = 22$). While the conventional KCM folding scheme cannot guarantee a constrained control input, the proposed TRS-based KCM folding input, which is synthesized by solving (1) with $\mathbf{Q} = \sqrt{2N}\mathbf{I}_{22}$, $\mathbf{B} = \mathbf{I}_{22}$, $\Omega = \frac{\zeta c_0}{\sqrt{2N}}$, $\zeta = 0.45$, and c_0 is a constant chemical conversion factor, satisfies the imposed quadratic constraints.

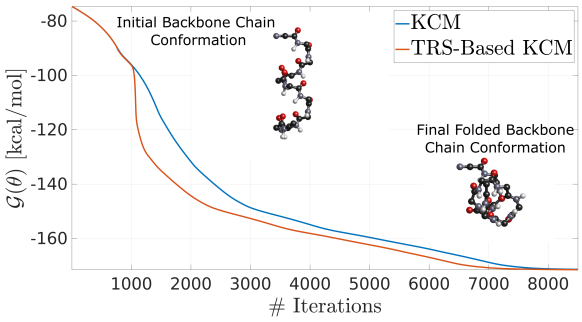

Fig. 1: The protein backbone chain free energy during folding.

REFERENCES

[1] F. Bergasa-Caceres and H. A. Rabitz, "Interdiction of protein folding for therapeutic drug development in SARS CoV-2," *J. Phys. Chem. B*, vol. 124, no. 38, pp. 8201–8208, 2020.

[2] P. Tavousi, M. Behandish, H. T. Ilieş, and K. Kazerounian, "PROTOFOLD II: Enhanced model and implementation for kinetostatic protein folding," *J. Nanotechnol. Eng. Med.*, vol. 6, no. 3, 2015.

[3] A. Mohammadi and M. W. Spong, "Quadratic optimization-based nonlinear control for protein conformation prediction," *IEEE Control Syst. Lett.*, vol. 6, pp. 373–378, 2022.

[4] S. Adachi, S. Iwata, Y. Nakatsukasa, and A. Takeda, "Solving the trust-region subproblem by a generalized eigenvalue problem," *SIAM J. Optim.*, vol. 27, no. 1, pp. 269–291, 2017.