Systematic classification of kinase conformations and binding sites to rationalize virtual screening

Eric Chen, Yingkai Zhang
Department of Chemistry – New York University, New York, NY

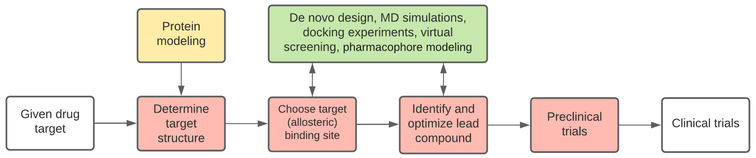
NYU SCCPC

Objective

- Develop rational computer-aided drug design (CADD) platform
- Analyze collections of PDB structures give intuition into drug design decisions

Background

- Protein structural data is readily available for many proteins.
- A general and portable method that analyzes and classifies structures of protein families without prior knowledge
- Use within CADD to guide target structure determination



- Design (allosteric) inhibitors to target unique binding sites on specific conformations of proteins

Case study: Cyclin Dependent Kinase (CDK) 2

- CDK2 activity is involved in cell cycle regulation and is dysregulated in many human cancers
- Target for development of non-hormonal contraceptives[1] and anti-cancer drugs[2]
- Most kinase inhibitors target the ATP binding site, but this is challenging due to limited on-target potency and selectivity

Methods

Classification

- Represent protein structures as pairwise residue–residue shortest distance matrices
  - Shortest distance is a proxy for interactions
- Use unsupervised machine learning methods (hierarchical clustering and principal component analysis (PCA)) to classify the structures

Analysis

- Standardized mean difference to determine cluster defining residue pairs

$$\mathrm{SMD}_{ij}^{x|y} = \frac{\overline{d_{ij}^{x}} - \overline{d_{ij}^{y}}}{\sqrt{\sigma_{ij}^{x} * \sigma_{ij}^{y}}}$$

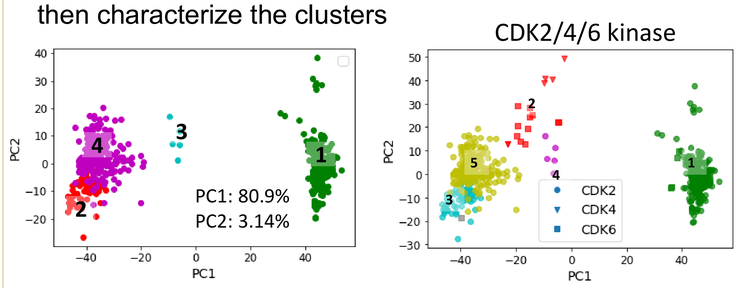
- Compare pockets across structures by local environment

Docking and Scoring

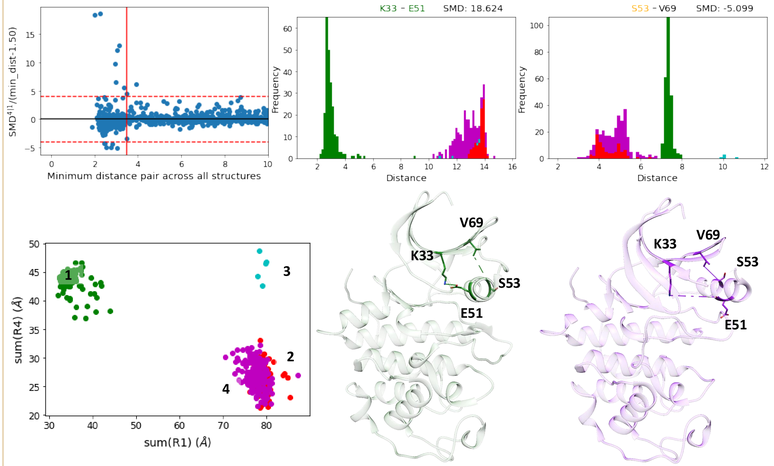
- Use deep learning (DiffDock)[3]/traditional (Lin_F9)[4] docking models, and machine learning scoring functions ($\Delta_{\text{Lin_F9}}$XGB)[5]

Package Highlights

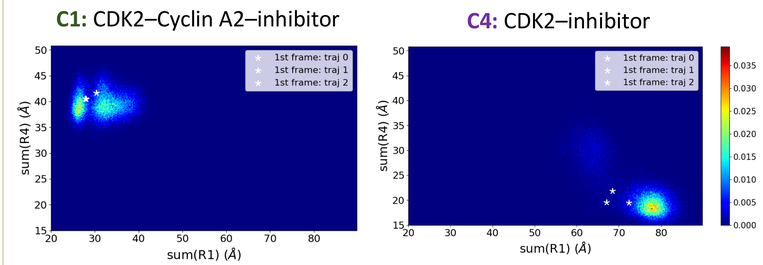
- Analyze CDK2 X-ray crystal structures by using PCA and then characterize the clusters



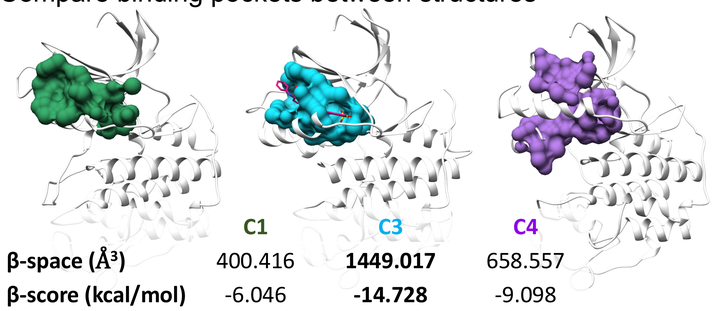
- Highlight the residue pairs that distinguish each cluster conformation



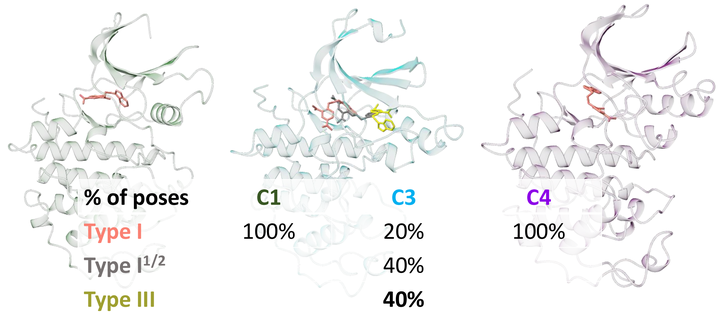
- Intuitively classify molecular dynamics simulations



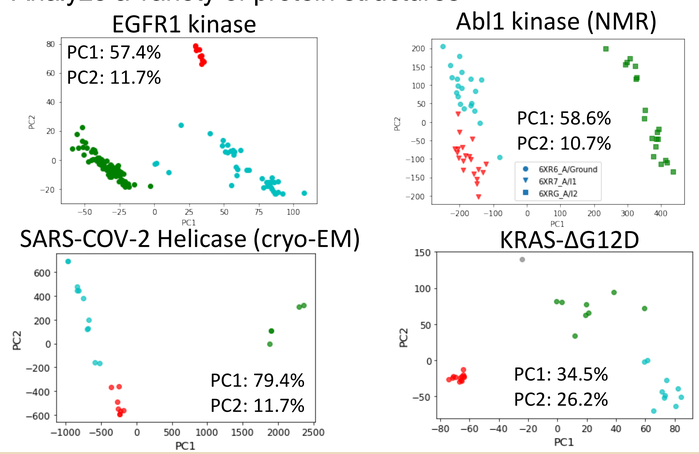
- Compare binding pockets between structures



- Dock allosteric compounds[6] to unique conformations to reveal binding preferences



- Analyze a variety of protein structures



Future steps

- Elucidate how experimental conditions bias structures
- Orthosteric and allosteric structural analysis and docking
- Enhance conformational diversity of protein structure predictions
- Characterize binding site location with ligand–receptor vector

Acknowledgements

Yingkai Zhang group, NYU HPC team and the Simons Center for Computational Physical Chemistry
PI: Yingkai Zhang
Group members: Alida Besch, Song Xia, Zixuan Feng, Charlotte Infante, Thomas Kelly, Xuhang Dai, Chao Han, Shih-Chiang Lo, Yuting Wu

[1]Faber, E. B., Wang, N. & Georg, G. I. *Biology of Reproduction* **103**, 357–367 (2020).
[2]Roskoski, R. *Pharmacol. Res.* **139**, 471–488 (2019).
[3]Corso, G., Stärk, H., Jing, B., Barzilay, R. & Jaakkola, T. Preprint at http://arxiv.org/abs/2210.01776 (2022).
[4]Yang, C. & Zhang, Y. *J. Chem. Inf. Model.* **61**, 4630–4644 (2021).
[5]Yang, C. & Zhang, Y. *J. Chem. Inf. Model.* **62**, 2696–2712 (2022).
[6]Faber, E. B., Georg, G. I., *et al. Nat Commun* **14**, 3213 (2023).