Calculation of histone-protein complex structures

the abstract of the PhD Thesis

by

Bayartsetseg Bayarsaikhan

Supervisor
Dr. Habil. Csaba Hetényi
Associate Professor

Doctoral School of Pharmacology and Pharmaceutical Sciences
Head of the Doctoral School
Prof. Dr. Erika Pintér

Department of Pharmacology and Pharmacotherapy
Medical School
University of Pécs
Hungary
2024

Research papers concerning the present thesis

(These papers are referred to as their Roman numerals in the text.)

PAPER I

Zsidó BZ*, **Bayarsaikhan B***, Börzsei R, Hetényi C. Construction of Histone–Protein Complex Structures by Peptide Growing. International Journal of Molecular Sciences. 2023 Jan;24(18):13831. [**IF: 4.9; Q1**]

*Equal contribution.

PAPER II

Bayarsaikhan B, Zsidó BZ, Börzsei R, Hetényi C. Efficient Refinement of Complex Structures of Flexible Histone Peptides Using Post-Docking Molecular Dynamics Protocols. International Journal of Molecular Sciences. 2024 Jan;25(11):5945. [**IF: 4.9; Q1**]

Other research papers

1. Zsidó BZ, **Bayarsaikhan B**, Börzsei R, Szél V, Mohos V, Hetényi C. The Advances and Limitations of the Determination and Applications of Water Structure in Molecular Engineering. International Journal of Molecular Sciences. 2023 Jan;24(14):11784. [**IF: 4.9; Q1**]
2. Börzsei R, **Bayarsaikhan B**, Zsidó BZ, Lontay B, Hetényi C. The Structural Effects of Phosphorylation of Protein Arginine Methyltransferase 5 on Its Binding to Histone H4. International Journal of Molecular Sciences. 2022 Sep 26;23(19):11316. [**IF: 5.6; Q1**]
3. Fliszár-Nyúl E, Faisal Z, Skaper R, Lemli B, **Bayartsetseg B**, Hetényi C, et al. Interaction of the Emerging Mycotoxins Beauvericin, Cyclopiazonic Acid, and Sterigmatocystin with Human Serum Albumin. Biomolecules. 2022 Aug;12(8):1106. [**IF: 5.5; Q1**]

The sum impact factor of the articles as a basis of the thesis: 9.8

Cumulative impact factor: 25.8

Introduction

1. Importance of histones

As epigenetics is fundamental in chromatin-related pathomechanisms of various diseases, understanding the epigenetic regulation is one the key challenges of this century. Among epigenetic mechanisms, post-translational modifications of histone proteins largely determine the epigenetic state of the cell and have crucial therapeutic, as well as diagnostic importance (1,2). Therefore, determining the atomic resolution structure of histone-reader (writer) protein complexes is a key to unraveling epigenetics and designing new drugs. However, determining such nucleosomal complex structures poses a significant challenge even to high throughput crystallographic techniques, due to their large size and the large number of possible complex structures (3). To answer the challenge, fast, complementary theoretical methods can serve as valuable alternatives to experimental techniques. These methods can help bridge the gap, providing insights and predictions that experimental techniques alone may not be able to achieve in a reasonable timeframe.

2. Peptide docking and challenges

Molecular docking tools predict the binding mode (position, orientation, and conformation) of ligands (drug candidates) to their target molecules and score and rank them (4). Despite significant advancements, current molecular docking methods still face several challenges and limitations, (4,5), especially when it comes to large, highly flexible peptide ligands.

The major drawback in docking peptide ligands is their large size and high conformational flexibility, which increase the number of possible conformations and search space, leading to higher computational costs and a greater number of false-positive results (6). In cases of the histone H3 peptides, anchoring residues located at its N-terminal end contribute the most to their often weak interactions with shallow binding pockets on the reader proteins which further complicates the prediction of accurate binding modes. Various strategies have been developed to overcome this limitation.

Fragment-based docking has gained significant attention in drug discovery, leading to clinical trials for several promising candidates (7). Unlike traditional docking methods that involve entire ligand structures, fragment-based docking first divides the ligand into smaller, low-molecular-weight fragments. These fragments are then docked into the binding site and either linked covalently or placed one by one incrementally to grow the complete bound ligand structure. By focusing on smaller fragments, the conformational space of ligand-protein interactions can be more efficiently explored, resulting in more accurate predictions of binding modes and affinities (7,8).

Despite its successes in past studies, current fragment-docking methods exhibit several limitations. Since the success of fragment docking heavily relies on the covalent linking of fragments (7), the primary challenge in fragment docking lies in finding a way to link the fragments with optimal steric alignment, considering the shape-wise match and the gap between the two docked fragments.

3. MD-based refinement methods and their limitations

Fast docking methods can use post-docking refinement steps prior to ranking to address the challenges in this field, specially to introduce structural flexibility and improve the energetics of the interface for accurate scoring (9). Refinement procedures can range from short energy minimizations, which remove steric clashes, to more advanced methods that allow binding site flexibility upon ligand binding, such as molecular dynamics (MD) or Monte Carlo simulations. MD simulations have been used in such refinements, as they can incorporate both ligand and protein flexibility, allowing the binding site to adapt to the ligand, strengthening pre-existing interactions formed during docking and forming new ones (10). Moreover, MD simulations can incorporate the effects of structural water molecules within the binding site using various solvent models. Despite the advantages, MD-based refinement protocols face several challenges, particularly for large peptide ligands, like H3 histone peptides, due to their extensive water-mediated hydrogen-bond networks with their target proteins. The presence of these bridging water molecules is essential for accurate binding mode predictions. For instance, Rastelli et al. reported that incorporating bridging water molecules into MD-based refinement methods significantly improved enrichment factors for adenosine A2A (11). However, only few refinement methods currently incorporate crystal or predicted water molecules to facilitate accurate mediated interactions during simulations.

Objectives of the present thesis

The primary aim of this thesis is to devise an innovative peptide docking protocol, capable of generating atomic resolution structures of histone-reader complexes, without requiring prior knowledge of the binding sites. Additionally, this work seeks to systematically investigate MD-based refinement strategies to enhance the accuracy of predicted ligand binding modes. The specific objectives are:

I. To design PepGrow, incorporating docking and in situ growing of histone H3 peptide fragments within reader protein binding pockets, followed by scoring and ranking the resulting complex models. The performance of PepGrow was evaluated on a set of histone H3 peptide-reader complexes and compared to a benchmark set of ten fast docking engines.

II. To develop MD-based refinement protocols to improve the structural accuracy of initial docking solutions generated by PepGrow, systematically investigating the effects of various MD simulation parameters to identify refinement conditions for enhancing peptide binding mode predictions.

I. Development and Evaluation of PepGrow (Paper I)

Methods

Fragment Selection and Docking: The development of PepGrow starts with the careful selection of an appropriate seed fragment for docking. For the histone H3 peptide and reader protein complexes, the popular fast docking tool AutoDock4 (12) was able to accurately dock short (di-)peptides at high accuracy. Therefore, all possible dipeptides were generated from the H3 peptide of System 1xwh, resulting in nine distinct fragments and docked using AutoDock 4.2.6. The highest-ranked binding modes were carried forward to the fragment growing step.
Fragment Growing: Instead of directly linking docked fragments, PepGrow utilizes Modeller's builder routine for fragment growth (13). We thoroughly evaluate restraints, energy calculation features, and varying seed numbers. Generating 100 models rapidly using default building settings was used for the final PepGrow protocol.
Scoring: To score and rank a large pool of peptide binding modes, the target-ligand intermolecular interaction energy and its separate components (Lennard-Jones and Coulomb terms) were calculated. The top 1% of ranked binding modes was observed to contain the best binding mode, thus, the representative of the top 1% ranked according to different energy values selected as 'Rank 1'. A comparison of the RMSD value of the Rank 1 binding mode obtained using different binding energies showed that E_{inter}-based representative selection method has the best performance.
Final Protocol: Among the nine dipeptides, Fragment 1 (AR) produced the best results and was chosen as the seed fragment for docking. The top-ranked binding mode of AR fragment was used as a seed fragment to produce 100 models rapidly using default building settings. The models were further ranked according to their intermolecular interaction energy and a binding mode with the closest match with the average coordinates of the peptide binding modes ranked in the top 1% was selected as the final solution.
Evaluation: A dataset of ten histone H3 peptide-reader complexes was used to evaluate PepGrow. Its performance was benchmarked against ten fast docking engines using a standardized protocol to avoid bias. Low-energy peptide structure and the unbound (apo) target structures were used in all docking calculations. To assess the sensitivity of the docking methods to target conformational change, we repeated all docking calculations using the bound (holo) forms of the reader proteins as targets.
Evaluation Criteria: The structural accuracy of all docking methods was evaluated using root mean square deviation (RMSD) calculated between the predicted and experimental (reference) binding modes of the peptide. An RMSD below 2 Å is generally considered to represent an accurate binding mode (14). For the histone H3 ligand, this core region corresponds to the first five amino acids. The lowest RMSD of all docked binding modes is referred to as $RMSD_{best}$.

Results

A comparison of the best docking results obtained by all docking methods evaluated in this study showed that PepGrow outperformed other benchmark methods. PepGrow produced a mean $RMSD_{best}$ of 5.36 (±1.47) Å for the full-length histone H3 peptide and achieved an excellent $RMSD_{best}$ of 4.09 (±1.18) Å for the first five amino acids. The acceptable level of

$RMSD_{best}$ is 4.0±3.0 Å based on data collected from publications of benchmark methods where RMSD was calculated only for the peptide backbone. On the other hand, side-chain atoms were also included in the RMSD calculations of this study. Therefore, PepGrow's performance can be considered good or above average compared to RMSD values from benchmark methods. Moreover, PepGrow's performance was consistent across both apo and holo target structures, showcasing the robustness of the method.

Structural accuracy is an important aspect of a docking method, but its ability to accurately rank the docked binding modes is equally important. Although PepGrow achieved good structural accuracy, ranking the best result as a final result is still challenging for all methods. The best binding modes consistently showed better conformational accuracy than the top-ranked modes across all docking methods, highlighting the need for improved ranking schemes. The E_{inter}-based representative selection in PepGrow proved to be a viable alternative but still requires further refinement.

II. Development and evaluation of MD refinement protocols (Paper II)

Methods

Previous studies and the results of PepGrow indicate that for peptide ligands, fast docking methods exhibit moderate structural precision which can be improved by applying a post-docking refinement step. In this study, we developed six MD-based refinement protocols to improve the structural accuracy of docking solutions produced by PepGrow.

Since top-ranked docking solutions are often selected as final solutions in drug development procedures, the top-ranked docking solutions from PepGrow, using apo target structures, were used as starting points for testing these protocols. All refinement protocols consist of two main steps: (i) pre-MD hydration, building the void-free hydration structure at the complex interface followed by five-step refinement of the predicted hydration structure, (ii) consecutive MD simulations. A pre-MD hydration step prepares the system for MD runs and constructs hydration structures for the target-peptide interfaces using the all-inclusive identity-based prediction algorithm of MobyWat (15,16). The hydrated structure underwent a five-step robust equilibration adapted from the HydroDock protocol (17). The main goal of the refinement step was to optimize the orientation of hydrogen atoms in the predicted water molecules, facilitating water network formation. The systems were then subjected to consecutive MD simulations, with parameters systematically varied across protocols, including simulation composition, temperature, length, position restraints, and peptide ligand length.

Results

The performance of the MD-based refinement protocols was evaluated by assessing how close the ligand binding modes come close to the experimental (reference) binding modes upon refinement starting from the initial conformations produced by PepGrow. All protocols improved the initial conformations generated by PepGrow, with Protocol P4 showing the best results. P4 achieved a median improvement of 32% (4.6 Å) over the initial docked structures in terms of the change in RMSD from the experimental references. and consistently delivered improvements greater than 1 Å in nearly all cases, with a maximum improvement of 84%. P4 demonstrated robust performance with apo structures, outperforming its results with holo structures and indicating its ability to handle variable target conformations.

An analysis of the MD parameters showed that extending simulation time or increasing simulated annealing temperatures beyond certain limits did not significantly enhance refinement accuracy. It was also concluded that flexible restraints on the binding site region, as applied in P4 and P6, improved target adaptability to peptide binding modes without destabilizing the complex. Excluding non-interacting C-terminal regions of H3 peptides was shown to facilitate stronger peptide-target interactions and yield more stable complexes.

Additionally, accurate initial positioning of anchoring residues (e.g., R2 in H3 peptides) proved critical, as it maintained strong target-ligand interactions during MD. Furthermore, the pre-MD hydration step played a vital role in constructing accurate hydration structures at the binding interface, enabling optimal target-peptide interactions during refinement.

Summary

This thesis addresses the difficulty of accurately predicting the complex structures of histone H3 peptides with various reader proteins using computational approaches. Histone peptides pose significant challenges to fast docking methods due to their linear N-terminal tail with large conformational flexibility (18) sticking out of the nucleosome structure. Moreover, extensively hydrated peptides like histones are well-known problematic cases for fast docking methods due to their lack of explicit water models (18). To overcome these limitations, we developed PepGrow, a fragment-based docking protocol that combines AutoDock 4.2 and Modeller's fast model-building capabilities. PepGrow was evaluated against ten histone H3 peptide-reader complexes and compared to ten benchmark methods designed for protein-peptide docking. The results demonstrate that PepGrow outperforms these benchmarks, particularly due to two main advantages: (i) the use of a di-peptide seed in the initial docking step, and (ii) growing the remaining peptide fragments using Modeller's robust building routine.

The ranking results showed that ranking the best solutions as a final result is still challenging for all methods tested in this study. The conformational similarity of the best binding mode, measured by RMSD, considerably exceeded RMSD of the top-ranked binding mode in all dockings, indicating poor ranking precisions. The ranking results also showed that E_{inter}-based representative selection implemented in PepGrow is a viable ranking alternative.

The histone complexes tested in this study were shown to be a particularly challenging test set for all docking methods, achieving moderate to poor precision in terms of structural accuracy. The precision drops further when the top-ranked solutions are considered. In such cases, post-docking refinement methods can be used to improve the accuracy of the predicted target-ligand complex structures.

In this study, we constructed six MD refinement protocols in an attempt to improve the precision of the top-ranked docking results produced by PepGrow. Protocol P4 showed the best performance, achieving a median of 32% (4.6 Å) improvement over the docked structures in terms of the change in root mean squared deviations from the experimental references. The results showed that the pre-MD hydration step and the inclusion of simulated annealing within the MD protocol, and the full flexibility of the binding site region made Protocol P4 a robust option for refining initial conformations of a wide range of structural qualities. The pre-MD hydration step was shown to provide an accurately predicted void-free hydration structure at the interface, which is crucial for forming the hydration networks necessary for appropriate target-peptide interactions during MD. Additionally, an accurate positioning of anchoring residues (R2 in the H3 peptides) in the starting peptides conformations were shown to affect the efficiency of the MD refinement significantly.

This study shows that a proper MD-based refinement protocol not only improves the structural accuracy of target–ligand complexes but also enhances the efficiency and reliability of current fast docking methods. Improved precision in predicted binding modes positively influences the estimation of binding affinity energies, therefore potentially enhancing the accuracy of ligand ranking. This advancement holds a potential for accelerating the discovery of new drugs in epigenetics or any design projects working with peptide ligands.

Summary of the New Results

In line with the aims of my research and the results detailed in this dissertation, the most significant findings of my doctoral work are summarized as follows:

1. Histone H3 peptides in complex with their reader proteins were shown to be a particularly challenging test set for existing docking methods. The benchmark methods that were specifically designed for protein-peptide or macromolecular complex docking (except for AutoDock4) showed moderate to poor structural accuracy, with a significant drop in precision when the top-ranked solutions were considered.
2. A novel fragment-based docking protocol, PepGrow, was developed. Instead of linking all peptide fragments, PepGrow docks a histone fragment seed and grows the full peptide tail within the reader protein's binding pocket. This *in situ* growing approach outperformed all benchmark methods in terms of structural accuracy.
3. Accurate ranking of the best binding modes remains a challenge for all docking methods. While the E_{inter}-based representative selection used in PepGrow offered a promising alternative, it still requires further optimization to improve ranking reliability.
4. Post-docking MD refinement protocols were designed to improve the structural quality of PepGrow docking results. Among the tested protocols, Protocol P4 performed best, achieving a median improvement of 32% (the largest improvement of 84%) over the starting docked structures.
5. Systematic exploration of MD simulation parameters revealed that extending simulation time or increasing the maximum simulated annealing temperature beyond certain thresholds did not yield significant improvements in refinement efficiency.
6. Proper initial positioning of anchoring residues, such as R2 in H3 peptides, was shown to significantly enhance the efficiency and accuracy of MD refinements, ensuring stable interactions throughout the simulation.
7. The pre-MD hydration step played a critical role in constructing accurate, void-free hydration structures at the binding interface. This step facilitated optimal target-peptide interactions and was crucial for achieving reliable and accurate MD refinement results.

These findings highlight the potential of PepGrow and MD refinement protocols to address long standing challenges in peptide docking, paving the way for more precise and efficient computational approaches in structural biology and drug discovery.

Acknowledgements

The research was supported by a János Bolyai Research Scholarship of the Hungarian Academy of Sciences. We acknowledge the HPC support of the Governmental Information Technology Development Agency, Hungary.

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Dr. Hetényi Csaba. His tireless support, brilliant and creative ideas have inspired me countless times throughout this journey. I deeply appreciate his thoroughness and patience, and I consider myself fortunate to have had the opportunity to work under his guidance.

I would also like to extend my sincere thanks to my PhD fellows, colleagues, and co-authors, Dr. Balázs Zsidó, Dr. Rita Börzsei, and Viktor Szél. I am particularly grateful for the productive collaboration with Dr. Balázs Zsidó, whose unwavering support has been invaluable during our work together. My gratitude also goes to Prof. Pintér Erika and the entire team at the PTE Department of Pharmacology and Pharmacotherapy.

Last but certainly not least, I extend my deepest thanks to my family and friends for their unwavering belief in me and their continuous emotional support. Without their encouragement and support, this journey would not have been possible.

References

1. Portela A, Esteller M. Epigenetic modifications and human disease. Nat Biotechnol. 2010 Oct;28(10):1057–68.
2. Rodenhiser D, Mann M. Epigenetics and human disease: translating basic biology into clinical applications. CMAJ. 2006 Jan 31;174(3):341–8.
3. Zsidó BZ, Hetényi C. Molecular Structure, Binding Affinity, and Biological Activity in the Epigenome. Int J Mol Sci. 2020 Jun 10;21(11).
4. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. Molecules. 2015 Jul 22;20(7):13384–421.
5. Huang SY. Comprehensive assessment of flexible-ligand docking algorithms: current effectiveness and challenges. Brief Bioinformatics. 2018 28;19(5):982–94.
6. Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O, et al. Protein-peptide docking: opportunities and challenges. Drug Discov Today. 2018;23(8):1530–7.
7. Bian Y, Xie XQS. Computational Fragment-Based Drug Design: Current Trends, Strategies, and Applications. AAPS J. 2018 Apr 9;20(3):59.
8. Lamoree B, Hubbard RE. Current perspectives in fragment-based lead discovery (FBLD). Essays Biochem. 2017 Nov 8;61(5):453–64.
9. Guterres H, Im W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. J Chem Inf Model. 2020 Apr 27;60(4):2189–98.
10. Huo S, Wang J, Cieplak P, Kollman PA, Kuntz ID. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. J Med Chem. 2002 Mar 28;45(7):1412–9.
11. Anighoro A, Rastelli G. Enrichment factor analyses on G-protein coupled receptors with known crystal structure. J Chem Inf Model. 2013 Apr 22;53(4):739–43.
12. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009 Dec;30(16):2785–91.
13. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. Protein Sci. 2000 Sep;9(9):1753–73.
14. Kramer B, Rarey M, Lengauer T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins: Structure, Function, and Bioinformatics. 1999 Nov;37(2):228–41.
15. Jeszenői N, Bálint M, Horváth I, van der Spoel D, Hetényi C. Exploration of Interfacial Hydration Networks of Target-Ligand Complexes. J Chem Inf Model. 2016 Jan 25;56(1):148–58.
16. Jeszenői N, Horváth I, Bálint M, van der Spoel D, Hetényi C. Mobility-based prediction of hydration structures of protein surfaces. Bioinformatics. 2015 Jun 15;31(12):1959–65.
17. Zsidó BZ, Börzsei R, Szél V, Hetényi C. Determination of Ligand Binding Modes in Hydrated Viral Ion Channels to Foster Drug Design and Repositioning. J Chem Inf Model. 2021 Aug 23;61(8):4011–22.
18. Bálint M, Horváth I, Mészáros N, Hetényi C. Towards Unraveling the Histone Code by Fragment Blind Docking. Int J Mol Sci. 2019 Jan 19;20(2):422.