Semiempirical quantum mechanical evaluation of target – ligand interactions using an end-point formalism

PhD thesis

Viktor Szél

Doctoral School of Pharmacology and Pharmaceutical Sciences
Neuropharmacology Program

Leader of doctoral school: **Prof. Dr. Erika Pintér**
Leader of doctoral program: **Prof. Dr. Erika Pintér**
Supervisor: **Dr. Csaba Hetényi**

University of Pécs
Faculty of Medicine
Department of Pharmacology and Pharmacotherapy
Pharmacoinformatics Unit

Pécs, 2025.

The author's scientific publications

Articles forming the basis of the thesis:

Article I

Szél, V.; Zsidó, B.Z.; Jeszenői, N.; Hetényi, C. Target – Ligand Binding Affinity from Single Point Enthalpy Calculation and Elemental Composition. *Phys. Chem. Chem. Phys.* **2023**, *25* (46), 31714–31725. **[IF: 2.9; Q2]**

Article II

Szél, V.; Zsidó, B.Z.; Hetényi, C. Enthalpic Classification of Water Molecules in Target - Ligand Binding. *J. Chem. Inf. Model.* **2024**, *64* (16), 6583-6595. **[IF: 5.7; Q1]**

Sum of impact factors of the articles forming the basis of the thesis: 8.6

Cumulative impact factor (all articles): 22.4

Independent citations (mtmt.hu, all articles): 14

Articles related to the topic of the thesis

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**, *61* (8), 4011–4022. **[IF: 6.2 ; Q1]**

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. *Int. J. Mol. Sci.* **2023**, *24* (14), 11784. **[IF: 4.9; Q1]**

Presentations and conference papers related to the topic of the thesis

Szél, V.; Hetényi, Cs.
A quantum mechanical scoring scheme for drug design (2020) – oral presentration, Online Medical Conference for PhD Students and Experts of Clinical Sciences (MedPECS 2020)

Hetényi, Cs.; Horváth, I.; **Szél, V.**; Zsidó, B. Z.; Börzsei R.

Szükségesek-e az explicit vízmolekulák a célmolekula-ligandum kötési affinitások számításánál? (2022) – oral presentration, Az MKE Szerves- és Gyógyszerkémiai Szakosztályának QSAR és Modellezési Szakcsoportja és az MTA Szegedi Akadémiai Bizottságának Kemometria és Molekulamodellezés Munkabizottságának szimpóziuma (KeMoMo-QSAR szimpózium), Szeged, Magyarország

Zsidó, B. Z.; Börzsei R.; **Szél, V.**; Hetényi, Cs.
HydroDock protokoll célpont-ligandum dokkoláshoz explicit vízmolekulákkal (2022) – oral presentration, Az MKE Szerves- és Gyógyszerkémiai Szakosztályának QSAR és Modellezési Szakcsoportja és az MTA Szegedi Akadémiai Bizottságának Kemometria és Molekulamodellezés Munkabizottságának szimpóziuma (KeMoMo-QSAR szimpózium), Szeged, Magyarország

Hetényi, Cs.; Zsidó, B. Z.; **Szél, V.**; Börzsei R.; Bayarsaikhan, B.
Exploration of water structure of protein-ligand complexes (2022) – oral presentration, Annual Meeting of the Hungarian Biochemical Society, Pécs, Magyarország

Hetényi, Cs.; **Szél, V.**; Zsidó, B. Z.; Börzsei R.
Optimization of Drug-Target Interactions Using Fast Pharmacoinformatics Protocols (2023) – oral presentration, 18th Annual Congress of International Drug Discovery Science & Technology (IDDST-2023), Amszterdam, Hollandia

Hetényi, Cs.; **Szél, V.**; Bayarsaikhan, B.; Zsidó, B. Z.; Börzsei R.
A hidratáció szerepe a gyógyszertervezésben (2023) – oral presentration, Bioinformatika 2023: Bioinformatika a Mesterséges Intelligencia Korában, Budapest, Magyarország

Zsidó, B. Z.; Börzsei R.; **Szél, V.**; Mohos, V.; Hetényi, Cs.
Az adamantán típusú vegyületek kötődési mechanizmusa a SARS-COV-2 ioncsatornájához (2023) – oral presentration, 52. Membrán-Transzport Konferencia, Sümeg, Magyarország

Other articles

Szél, V.; Csekő, G.; Horváth, A. K. Kinetics and Mechanism of the Oxidation of Bromide by Periodate in Aqueous Acidic Solution. *J. Phys. Chem. A* **2014**, *118* (45), 10713–10719.

1. Introduction

One of the key challenges in target-based computational drug design is selecting ligands that bind to the target molecule with the appropriate strength, forming a starting point for further development. The interaction strength between the ligand and the target is typically assessed using scoring functions, which have been continuously developed over nearly four decades. These scoring functions are mathematical models that sum the various components of the interaction to generate a numerical score. Ligands can then be ranked based on these scores, allowing for the selection of the best matches. However, a well-known limitation of these models is their limited accuracy, particularly due to an inadequate description of the electronic nature of non-covalent interactions and the water structure at the target-ligand interface. In this work, our goal was to create a quantum mechanical scoring function that overcomes these limitations by incorporating the effects of the surrounding water medium through a combined implicit-explicit solvation model. Speed was also an essential factor, as scoring functions generally need to enable virtual screening of large compound libraries within reasonable time frames. In Article I, we presented the methodological development of a scoring function that combines end-point semi-empirical PM7/1SCF calculations, the implicit COSMO solvation model, predicted explicit interfacial water structures, and ligand-based descriptors. This approach was applied to a diverse set of target-ligand systems that span a wide range of ligand sizes. During complex formation, explicit water molecules in the target binding pocket rearrange and contribute uniquely to the resulting binding affinity. Water molecules can either be retained, helping to mediate the target-ligand interaction, or be displaced from the binding pocket. These roles are associated with characteristic thermodynamic profiles and their spatial positions are valuable for designing optimized ligands. However, measuring the individual thermodynamic contributions of these water positions is not feasible experimentally, making theoretical methods essential. Thus, our goal was to develop a method to individually score water molecules located in the binding pocket, which could identify potentially conserved or displaced waters without prior knowledge of the ligand. In Article II, we explored the issue of water conservation by using the semi-empirical PM7/1SCF method to estimate the binding enthalpy of water positions. We defined the concept of conservation, validated our method through statistical analysis, and provided drug design examples to demonstrate its applicability. In our studies all targets were protein molecules, but the applied procedures can easily be extended to other macromolecules.

2. Methods

During this work, we applied computational chemistry methods, including molecular dynamics simulations, structural optimization, and end-point calculations at the molecular mechanical and semi-empirical quantum mechanical levels.

In Article I, we investigated the free energy of binding by compiling a dataset of target-ligand systems. This dataset included experimental binding enthalpy (ΔH_b) and binding free energy (ΔG_b) pairs obtained from isothermal titration calorimetry measurements, as well as atomic-resolution structures. After completing the missing parts of the structures, we generated the water structures of the target-ligand interfaces using a two-step, molecular dynamics-based protocol. The resulting structures were minimized at the molecular mechanical level, and the binding regions were extracted through fragmentation. To estimate the binding enthalpy, we performed end-point calculations using the PM7/1SCF semi-empirical quantum mechanical method, and fitted linear regression models between the computed reaction enthalpy ($\Delta_r H$) and the experimental data. To assess the impact of water structure on binding enthalpy, we compared implicit, explicit, and implicit-explicit (hybrid) water models. The binding enthalpy was calculated for structures minimized at both molecular and quantum mechanical levels, allowing us to compare the time requirements and achievable accuracy of the methods. The binding free energy was estimated by combining the predicted binding enthalpy ($\Delta_r H$) with an additional variable, systematically selected from a descriptor set based on ligand the structures, using multilinear regression.

In Article II, we explored the role of water positions in target-ligand binding by generating water structures for the ligand-free target (apo) and the target-ligand complex (holo) for the selected systems, using the same protocol as before. By aligning and structurally comparing the apo-holo pairs, we classified the apo water positions into conserved and displaced groups. After molecular mechanical minimization and structural fragmentation, we calculated the end-point binding enthalpy (ΔH_b) for each apo water position using the PM7/1SCF semi-empirical method and then organized the values into a list in ascending order. We further investigated the possibility of classifying the conserved and displaced groups based on the calculated ΔH_b, which we evaluated using ROC analysis and the enrichment factor (EF) for both the entire dataset and individual systems. We also calculated the fractional polar molecular surface area of the binding pockets for each system to characterize the polarity of the pockets. As an application of our method, we determined and classified the apo water positions for the HIV-1 protease and Influenza A ion channel targets, and tested its effectiveness by comparing water positions with the structures of known ligands.

3. Results

3.1. Target- ligand binding affinity from single point enthalpy calculation and elemental composition

We established a linear relationship between the experimental values of the target-ligand binding enthalpy and the endpoint reaction enthalpy ($\Delta_r H$) calculated using the PM7/1SCF method. Among the approaches used to model the water structure of the complex interface, the combined implicit-explicit (hybrid) model proved to be the most accurate. From these, we selected the Hybrid 1 version with no axis intercept ($\beta=0$) as the representative model. After removing outliers, we observed a good agreement with the experimental values ($R^2 = 0.93$, RMSE = 2.7 kcal/mol). We further demonstrated that the binding enthalpy values calculated on QM minimized structures are comparable to those obtained via the 1SCF method. However, due to its significantly higher computational cost, using QM minimization is not practical. For estimating the binding free energy (affinity), quantities NHA/NA and NHB/NB were selected from the ligand-based descriptor set. These descriptors have the advantage of being predictive for ΔG_b, while not correlating with $\Delta_r H$ and are primarily entropic in nature. We showed that a multilinear model, combining the calculated 1SCF binding enthalpy and one descriptor (NHA/NA) as independent variables, gives a good fit to the experimental ΔG_b values ($R^2 = 0.60$, RMSE = 0.94 kcal/mol), which was named QMH-L.

3.2. Enthalpic classification of water molecules in target-ligand binding

In case of the classification performed on the entire dataset, we demonstrated that based on the calculated binding enthalpy values of the individual apo water positions, conserved and displaced groups can be distinguished (ROC AUC = 0.74). For our method, we established a general classification threshold based on the maximum F1 metric at $\Delta H_b = -1.9$ kcal/mol. According to this threshold, positions with a more negative binding enthalpy are identified as conserved, while those with higher binding enthalpy are categorized as displaced. When examining individual systems, we showed that the classification efficiency increases with the increasing polarity of the target binding pocket, which aligns with the well-defined hydration sites of polar pockets and their role in stabilizing target-ligand interactions. In terms of classification efficiency based on ΔH_b, the two endpoints among the examined systems were represented by 1lcj (ROC AUC = 0.85) and 2jm6 (ROC AUC = 0.67). We validated our method on two pharmacologically relevant targets to identify water positions that are crucial in ligand binding. For the HIV-1 protease target, we identified the W1 and W2 water positions at the top of the ΔH_b list, which are well-known for their role in drug binding. W1 is positioned between two aspartate units (D25) and mostly integrated into the ligand, while W2 is positioned between two isoleucine units (I50) and utilized forming a bridging contact between target and ligand. We also identified futher water positions that could be matched with the structure of other known inhibitors. For the Influenza A virus ion channel, we interpreted the optimized hydrogen bonding network of the binding pocket in the case of the bound spiro-adamantyl-amine using the W14 and W19 water positions identified as displaced.

4. Summary

In this work, we focused on the computational investigation of target-ligand binding affinity (ΔG_b) and the role of water structure in binding interactions. To address the limitations of classical MM-based methods, particularly the inadequate treatment of electronic effects and solvent interactions, we turned to quantum mechanical approaches, which allow for accurate calculation of molecular charge distributions. While the more precise description comes with significant computational demands, we bridged this gap by combining the PM7 semi-empirical method with a single-point (1SCF) approach, calculating the enthalpic component of target-ligand interactions (ΔH_b) for challenging systems containing large peptide ligands. To account for solvent effects, we used the COSMO implicit model and predicted explicit water molecules with the MobyWat program at the complex interface. By supplementing the binding enthalpy term with a descriptor describing the density of heavy atoms of the ligand (NHA/NA), we developed the QMH-L scoring function for estimating binding affinity. Our models were validated against experimental ΔH_b and ΔG_b data from ITC measurements, showing a good fit. The advantage of the QMH-L model is that the semi-empirical method used does not require additional parameterization and provides a good balance between accuracy and computational cost, making it a reasonable compromise within the range of scoring methods from molecular mechanics to ab initio quantum mechanics. Semi-empirical approaches, given current computational capabilities, can be routinely applied with high throughput opening up a wide range of applications. The broad applicability of quantum mechanical methods in drug design is emphasized in current review articles citing our work supporting the importance of this area. We also applied semi-empirical quantum mechanical calculations to predict hydration, which often plays a crucial role in target-ligand interactions. We determined and classified over 1,000 water positions in the binding pockets of various targets, using the PM7/1SCF method to estimate the binding enthalpy of individual water molecules. By examining systems with varying polarities, we showed that classification efficiency increases with the increasing polarity of the target binding pocket. The knowledge of conserved and displaced water positions can be useful, for example, in the design of optimized ligands and also in virtual screening tasks. Conserved water positions optimize hydrogen bonding networks between the target and ligand, either by mediating interactions or by being incorporated isosterically into the ligand structure, while displaced waters contribute a favorable entropic component to the overall binding affinity. We validated our method on the two pharmacologically relevant drug targets of HIV-1 protease and Influenza A ion channel. For the HIV-1 protease, we identified key conserved water positions the utilization of which was structurally confirmed in known inhibitors. On the other hand, for the Influenza A ion channel, we demonstrated that the water molecules identified as displaced contribute to the optimization of the hydrogen bond network in the binding pocket, which was interpreted through the binding modes of ligands.

5. Theses points

According to the objectives outlined in the Introduction, I summarize the results of my thesis in the following points:

1. Binding enthalpy (ΔH_b) values for target-ligand complexes were estimated using the calculated end-point reaction enthalpy ($\Delta_r H$) of binding at the PM7/1SCF semi-empirical quantum mechanical level. A good agreement with experimental values was achieved (Article I).

2. Fast 1SCF calculations were shown to be a suitable alternative to the computationally more demanding QM minimization approach for calculating target-ligand binding enthalpy (ΔH_b) (Article I).

3. A systematic comparison of approaches for handling the water structure of the target-ligand interface was performed, which resulted in the hybrid (implicit-explicit) model providing the best description. This was achieved by combining the COSMO implicit solvent model with explicit water molecules predicted by the MobyWat program (Article I).

4. The QMH-L calculator was developed to estimate target-ligand binding affinity (ΔG_b) using the calculated binding enthalpy (Point 1) along with a ligand-based descriptor to provide accurate results. The NHA/NA descriptor was selected for the final model which represents the density of the ligand's heavy atoms and is potentially generalizable across the chemical space (Article I).

5. Target (apo) and target-ligand (holo) water structures were generated using MobyWat protocols. Based on structural apo-holo comparisons apo water positions were classified into conserved and displaced groups according to their roles during ligand binding (Article II).

6. A method was developed to distinguish between conserved and displaced water groups based on the binding enthalpy of individual apo water positions, calculated using the PM7/1SCF semiempirical quantum mechanical method ($\Delta_r H$). Additionally, we showed that the classification efficiency based on binding enthalpy increases with the polarity of the target's binding pocket (Article II).

7. Our classification method was applied to identify key conserved water positions that have been utilized in the design of known inhibitors for the HIV-1 protease drug target. We also interpreted the role of displaced water molecules in ligand optimization for the Influenza A ion channel target (Article II).

Acknowledgement

I would like to thank my supervisor, Dr. Csaba Hetényi , who was always available to answer my professional questions and was very flexible in dealing with everyday issues. I would also like to thank the colleagues of the Unit of Pharmacoinformatics, Dr. Balázs Zsidó, Dr. Rita Börzsei and Bayartsetseg Bayarsaikhan, for always being available to help with practical matters. I would like to thank my mother for the support I have received over the years. I would like to thank my father for the sports afternoons we had together, and my friends for the get-togethers, which all contributed to maintaining balance.