

Diss. ETHNr. 14975

Current Problems in Molecular Docking

A dissertation submitted to the
Swiss Federal Institute of Technology Zurich

for the degree of
Doctor of Natural Sciences

presented by

Pavel Pospíšil

Chemical Engineer
Institute of Chemical Technology, Prague, Czech Republic
&
License de Biochimie
Université Joseph Fourier, Grenoble, France

born September 23rd, 1971
citizen of Prague, Czech Republic

accepted on the recommendation of

Prof. Dr. G. Folkers, examiner
Prof. Dr. L. Scapozza, co-examiner



2002

Summary

The present dissertation is concerned with the methodology and open issues of molecular docking and virtual screening. Virtual screening is a modern, computer-based application, which consists in the screening of chemical libraries on therapeutically interesting macromolecular targets. Due to its speed and cost-effectiveness, the pharmaceutical industry has shown recently a considerable interest in this technique. One of the critical requirements to ensure accurate performance in virtual screening is a deep elucidation of the active site of the targeted molecule.

Several contemporary achievements in molecular docking and virtual screening are reviewed in the introduction. Virtual screening, standing for a fast serial docking of small molecules to a target site can lead to very different results. These depend on the kind of algorithm implemented in the docking program, on the simulation of the flexibility of both the ligand and the receptor molecules and on the type of chemical library, which is to be used for screening. In particular, a correct preparation of the active site of the target molecule has a crucial impact on the accuracy of the docking and the results of the scoring function.

The second part of this work shows the methodology of docking in a case study of three different therapeutically interesting enzymes: Human dihydroorotate dehydrogenase (DHODH), human phosphodiesterase 4 (PDE4) and Herpes Simplex virus type 1 thymidine kinase (HSV1 TK). Whereas the binding modes of substrate and inhibitors within the active site of DHODH and HSV1 TK are known, for PDE4, the binding orientation of its natural substrate cAMP has only been subject of speculations. Docking into DHODH, whose active site is mainly hydrophobic, led relatively easy to accurate docking results, even though the presence of cofactors in the active site turned out to be crucial for the prediction of the binding mode. The crystal structure of PDE4 was resolved without a bound ligand. It showed that this enzyme

contains in its active site a magnesium and a zinc ion as well as a water molecule that plays a crucial role in the catalytic mechanism. Several docking experiments using cAMP and the PDE4 inhibitor rolipram were performed and the results showed a clear-cut dependency between the ligand orientation and the presence of metals in the active site.

Herpes Simplex virus type 1 thymidine kinase is the main protein studied in this work. HSV1 TK has a broad substrate diversity as compared to its cellular isoenzyme being able to phosphorylate diverse pyrimidine and purine analogs. HSV1 TK in combination with ligands acting as prodrugs is used as suicide enzyme in virus-directed enzyme prodrug gene therapy of cancer and stem cell transplantation. The active site of HSV1 TK is small and polar, and contains discrete crystal water molecules, which mediate the binding mainly between pyrimidines and the protein. Because of the crystallographic information on the presence or absence of water within the binding sites of thymidine and the clinically used purine derivatives aciclovir and ganciclovir, docking of known binders to HSV1 TK was performed with or without waters present in the active site. The docking into the water-containing site showed a higher accuracy and more exact prediction of free energy of binding for pyrimidine analogs. Similar results were obtained for purine analogs docked into the water-empty site. These structural insights were used in subsequent virtual screening studies. Because the water molecules in the docking site were left in place during the docking procedure, the compound CFBT was found to be a HSV1 TK ligand. Subsequent experimental testing confirmed CFBT as a new lead compound of HSV1 TK, which can be further optimized as a substrate and inhibitor.

In the context of broadening the knowledge on substrate specificity of HSV1 TK and another essential kinase, the Varicella Zoster virus thymidine kinase (VZV TK), new 9-(2-hydroxypropyl)substituted purine derivatives were docked into the crystal structure of HSV1 TK. The docking orientations and calculated free energy of binding were compared to their experimentally determined binding constants and *in vitro* phosphorylation rates. A set of criteria for the evaluation of molecular docking results was developed and allowed the qualitative prediction of a given ligand to be a good or poor substrate or a potential inhibitor. In particular, docking results showed multiple

binding modes, which is in line with the observations that these purine derivatives bind to thymidine kinases with mM or sub-mM affinities.

The third part of this work is dealing with the open question of what may be the impact of tautomeric states in virtual screening. Because of the fact that pyrimidine and purine analogs are able to acquire several tautomeric states, HSV1 TK is a suitable target for addressing this question. A program named AGENT to generate tautomers of nucleoside derivatives was developed. AGENT created rapidly the correspondent number of tautomers to every nucleoside entry in the database. The two databases, one containing the original compounds and the other their tautomers were screened together on the active sites of HSV1 TK in three different states. The comparative evaluation of scoring of the compounds and tautomers revealed that tautomers and compounds achieved different scores as compared to earlier screening experiments, in which tautomerism was disregarded. One of the active sites used in the screening contained a 180° flipped rotamer of the glutamine 125 side-chain, which is a key residue in nucleobase recognition. It builds a hydrogen bond network with a donor-donor-acceptor pattern, which is favored by the majority of the purine and cytidine derivatives. Especially in this case, subsequent comparative ranking of compounds versus tautomers revealed potential for new hit detection when including tautomeric forms in virtual screening procedures.

The last section of the present work is dedicated to a case study showing the docking to a model of Epstein-Barr virus thymidine kinase (EBV TK). The three dimensional structure of EBV TK is not known. The enzyme is implicated in several EBV-associated diseases. To gain further structural insights and to speed up drug development, a model was built using homology modeling. The model was subsequently refined by the means of a 1 ns molecular dynamic simulation and prepared ready to be used for molecular docking studies of nucleoside derivatives. Superimposition of EBV TK and its template HSV1 TK showed why thymidine and other nucleobase derivatives bind more weakly to EBV TK than to HSV1 TK. Molecular docking of 29 essential and known nucleobase derivatives used in clinical trials uncovered the binding patterns of the compounds. The docking of purine analogs demonstrated that the enzyme does not phosphorylate ganciclovir; a prodrug largely discussed in association with EBV TK. Molecular docking revealed the structure-

activity relationship of investigated compounds towards EBV and gives new insights for the design of new EBV TK substrates and inhibitors.

Zusammenfassung

Die vorliegende Dissertation behandelt die Methodik und offene Fragestellungen des molekularen Dockens und des *Virtual Screenings*. *Virtual Screening* ist eine moderne, computergestützte Methode zur Auffindung neuer Leitstrukturen. Dabei werden Bibliotheken von potentiell aktiven Wirkstoffen mit therapeutisch interessanten makromolekularen Zielstrukturen gesiebt. Da sich mit Hilfe dieser Technik schnell und kostengünstig aus einer grossen Anzahl von Molekülen Leitstrukturen für die Wirkstoffentwicklung herausfiltern lassen, ist das *Virtual Screening* auch in der pharmazeutischen Industrie auf grosses Interesse gestossen. Unabdingbar ist allerdings die genaue Kenntnis der strukturellen Eigenschaften des aktiven Zentrums des Zielproteins.

In der Einführung werden die Grundlagen und die wichtigsten aktuellen Trends des molekularen Dockens und des *Virtual Screenings* gestreift. Die Zuverlässigkeit des Dockens hängt davon ab, was für ein Dockingalgorithmus verwendet wurde, wie die Flexibilität des aktiven Zentrums und der Ligandmoleküle simuliert wurden und wie gross die strukturelle Vielfalt der in der Datenbank enthaltenen Moleküle ist. Ein weiteres, nicht zu vernachlässigendes Detail ist die korrekte Vorbereitung des aktiven Zentrums des Zielmoleküls vor der Dockingprozedur.

Im zweiten Teil dieser Arbeit wird die Methodologie des Dockings an drei im Mittelpunkt therapeutischen Interesses stehenden Zielenzymen erläutert: Humane Dihydroorotatdehydrogenase (DHODH), humane Phosphodiesterase 4 (PDE4) und Herpes Simplex Virus Typ 1 Thymidinkinase (HSV1 TK). Während die Bindungsmodi von Substraten und Inhibitoren in den aktiven Zentren von DHODH und HSV1 TK bekannt sind, kann über die genaue Orientierung der Bindungen des natürlichen Substrates cAMP an PDE4 nur spekuliert werden. Docking von Molekülen in das aktive Zentrum von DHODH, welches hauptsächlich von hydrophoben Gruppen

gebildet wird, führte ohne grossen Aufwand zu strukturell vernünftigen Resultaten. Es konnte jedoch gezeigt werden, dass die Anwesenheit des Cofaktors Voraussetzung für die korrekte Bindungsvorhersage war. Die Kristallstruktur von PDE4 wurde ohne gebundenen Liganden gelöst. Das aktive Zentrum von PDE4 enthält neben einem Magnesium- und einem Zinkion ein Wassermolekül. Mehrere Dockingstudien wurden mit dem natürlichen Liganden cAMP und dem PDE4-Inhibitor Rolipram durchgeführt. Dabei konnte ein klarer Unterschied zwischen den von den in das aktive Zentrum ohne Metallionen gedockten Molekülen im Vergleich zu den in das aktive Zentrum mit Metallionen gedockten Verbindungen gewählt Bindungsmodus festgestellt werden.

Herpes Simplex Virus Typ 1 Thymidinkinase ist das im Rahmen dieser Arbeit am intensivsten studierte Enzym. HSV1 TK zeichnet sich im Unterschied zu der humanen zellulären Thymidinkinase dadurch aus, dass es eine ausserordentlich breite Palette von Pyrimidin- und Purinderivaten zu phosphorylieren in der Lage ist. HSV1 TK wird in Kombination mit als *Prodrugs* agierenden Substraten als Suizidenzym in der Suizidenzym/Prodrug Gentherapie gegen Krebs und andere, vom immunkompromitierten Zustand von Transplantationspatienten profitierenden Krankheiten eingesetzt. Das aktive Zentrum der HSV1 TK ist klein und polar und enthält Wassermoleküle, welche hauptsächlich die Bindung von Pyrimidinen unterstützen. Aufgrund der wichtigen Rolle, welche vorhandene oder nicht vorhandene Wassermoleküle beispielsweise bei der Bindung von Thymidin oder den antiviral wirkenden Purinderivaten Aciclovir und Ganciclovir spielen, wurden die bekannten HSV1 TK Substrate in ein aktives Zentrum der TK mit und in eines ohne die erwähnten Wassermoleküle gedockt. Während die Resultate der Dockingstudie mit dem wasserenthaltenden aktiven Zentrum eine höhere Genauigkeit und eine bessere Vorhersage der freien Bindungsenergie für Pyrimidinanaloga erlaubte, konnte für Purinanaloga das gleiche beobachtet werden, wenn das wasserfreie Zentrum verwendet wurde. Die so gewonnenen strukturellen Erkenntnisse wurden in weiteren Dockingstudien verwendet. Weil die Wassermoleküle bei einem Teil der Studien im aktiven Zentrum belassen wurden, wurden wir auf die Verbindung CFBT aufmerksam. Experimentelle Analysen zeigten, dass es sich bei CFBT um einen neuen HSV1 TK Liganden handelte, welcher nun weiter optimiert werden kann.

Um das bestehende Wissen über die Substratspezifität der HSV1 TK sowie der Varicella-Zoster Virus Thymidinkinase (VZV TK) zu erweitern, wurden neue 9-(2-hydroxypropyl)-substituierte Purinderivate in das aktive Zentrum der HSV1 TK gedockt. Die Orientierung der gedockten Moleküle innerhalb des aktiven Zentrums und die berechnete freie Bindungsenergie wurden mit experimentell bestimmten Bindungskonstanten und *in vitro* Phosphorylierungsraten verglichen. Ein Kriterienset zur Evaluation der Resultate wurde entwickelt. Mit Hilfe dieses Kriteriensets war es möglich, qualitativ vorherzusagen, ob es sich bei einem gegebenen Liganden um ein gutes oder ein schlechtes Substrat oder um einen potentiellen Inhibitor handelt. Die Resultate der Dockingstudie zeigten, dass für einige der Verbindungen mehrere energetisch ähnliche Bindungsmodi (Multipler Bindungsmodus) existieren, was die experimentell beobachteten relativ schwachen Bindungskonstanten (mM oder sub-mM) dieser Purinderivate erklärt.

Der dritte Teil dieser Arbeit behandelt das noch weitgehend unerforschte Gebiet des Tautomerismus im *Virtual Screening*. HSV1 TK eignet sich vorzüglich für solche Studien, da sowohl Purin- wie auch Pyrimidinanaloga eine Reihe von tautomeren Zuständen annehmen können. Ein Computerprogramm namens AGENT zur Generation von Tautomeren von Nukleosidderivaten wurde entwickelt. AGENT kreierte eine neue Datenbank mit – grob im Durchschnitt gerechnet – ca. vier Tautomeren pro Eintrag in der Ausgangsdatenbank. Die beiden Datenbanken wurden für Screeningstudien mit drei verschiedenen aktiven Zentren der HSV1 TK verwendet. Eine komparative Auswertung der im *Scoring* erreichten Werte der Verbindungen und ihrer Tautomeren ergab, dass mehrere Tautomere besser abschnitten als ihre Ursprungsverbindungen. Eine der erwähnten in der Dockingstudie verwendeten drei HSV1 TK Zentren enthielt ein 180° Rotamer der Glutamin 125 Seitenkette, einer Schlüsselstelle bei der Erkennung von Nukleobasen. Zusammen mit einem benachbarten Wasserstoffmolekül bildet Glutamin 125 ein Wasserstoffbrückennetzwerk mit einem Donor-Donor-Akzeptor Muster, welches von der Mehrheit der Purin- und Pyrimidinderivate bevorzugt wird. Speziell in diesem Fall schnitten besonders viele Tautomere im *Scoring* besser ab als ihre Ursprungsverbindungen.

Der letzte Teil der vorliegenden Arbeit ist einer Fallstudie, welche die beim Docking ins aktive Zentrum eines Modells der Epstein-Barr Virus Thymidin Kinase (EBV TK) auftretenden Problem beschreibt, gewidmet. Die dreidimensionale Struktur der EBV TK ist unbekannt. Um die strukturellen Eigenschaften des aktiven Zentrums der EBV TK abschätzen zu können und so die Wirkstoffentwicklung voran zu bringen, wurde mit Hilfe von *Homology Modelling* ein Modell der EBV TK konstruiert. Das Modell wurde anschliessend mit Hilfe einer 1 ns dauernden Molekulardynamiksimulation verfeinert und für Dockingstudien mit Nucleosidderivaten vorbereitet. Durch Überlagerung der EBV TK mit der beim *Homology Modelling* als Vorlage verwendeten HSV1 TK konnte gezeigt werden, warum Thymidin und andere Nucleobasen schwächer an die EBV TK als an die HSV1 TK binden. Molekulare Dockingstudien mit 29 bekannten Nucleobasederivaten, welche in klinischen Versuchen verwendet worden waren, zeigten die Bindungsmodi der erwähnten Substanzen. Durch das Docken von Purinderivaten konnte gezeigt werden, warum die EBV TK Ganciclovir nicht phosphoryliert. Molekulare Dockingstudien am EBV TK Modell führten zu neuen strukturellen Erkenntnissen und beschleunigten die Entwicklung neuer EBV TK Substrate und Inhibitoren.