Zoltán Simon

Drug Discovery by Polypharmacology-based Interaction Profiling

PhD theses

Supervisor: András Málnási-Csizmadia DSc.

Structural Biochemistry Doctoral Program
Doctoral School in Biology

Program Leader: Prof. László Gráf DSc.
Head of the School: Prof. Anna Erdei DSc.

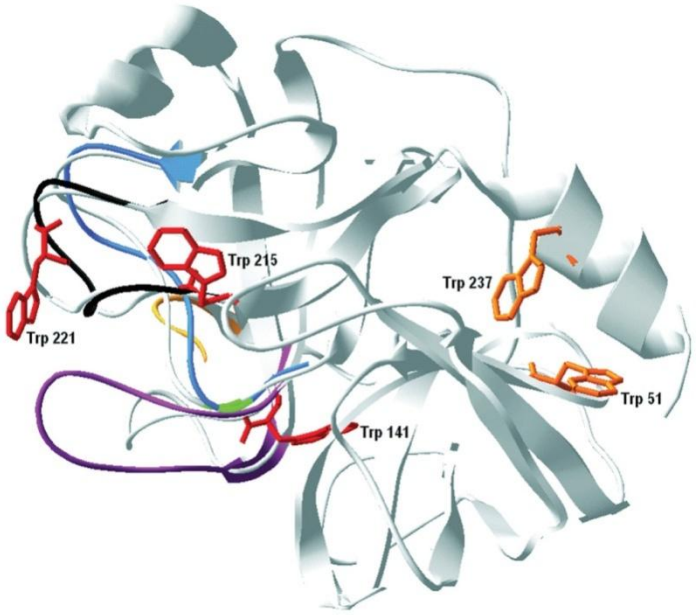
Eötvös Loránd University
Budapest, Hungary
2010

Introduction

The early pharmacologic theories explained drug-effect associations in a mechanistic manner, implying that a drug selectively acts on a specific biological target and affects its function like a magic bullet (Ehrlich). Now the tide seems to turn: drugs are recognized as affecters of complex biological networks. Systems-based approaches are gathering larger and larger ground, bringing a holistic view into drug research. Polypharmacology is a newly emerging approach which reflects the high complexity of the mechanism of actions of drugs, i.e., the fact that many drugs affect multiple proteins. This aspect of pharmacology has not been fully exploited in drug development. Consequently, the entire effect profiles of drugs and drug candidates have remained unrevealed. We hypothesized that complex molecular feature sets of drugs, e.g., *in silico* generated interaction patterns against a set of protein binding sites, correlate with the known part of effect profiles and may therefore provide predictive power to reveal the entire effect profiles of drugs. We also assumed that no target proteins are needed to obtain a high level of correlation between the interaction profiles (IPs) and the effect profiles (EPs) since the interactions of a drug with a diverse protein set mimics its possible interaction pattern with the human proteome.

A shift of scientific viewpoint, similar to the one in pharmacology, occurred in a much smaller but not simpler system: a protein. In Emil Fischer's time, proteins were considered as static objects whose interactions with other molecules could be described as the interaction of two complementary shapes, a key and a lock. In contrast with the "molecular machinery" approach, it is now revealed that proteins show more flexibility than anything engineered by a human being. In the second part of my thesis, the complex problem of protein flexibility was studied in the model system of the activation of human trypsin 4. During activation, four distinct regions of the protein undergo conformational change. The protein parts involved in the conformational change, i.e., the activation domain is bordered by hinge glycines that secure the large conformational freedom needed for the rearrangement. In our study, the kinetic and thermodynamic parameters of the activation were modified by introducing side chains with different sizes at a hinge position. We examined the temperature and external viscosity dependence of the rate of the conformational change and evaluated the results applying Kramer's theory. Based on this, a measure reflecting protein flexibility, i.e., the internal viscosity parameter of the activation of the wild type trypsin and two hinge mutants were determined. Our results suggest that the flexibility of the studied protein can be modulated by introducing point mutations in the hinge region.

The two studied topics are related in the level of complexity; consequently, a similar reduction in the system is needed to study them. In pharmacological effect prediction, complex drug-protein interaction patterns and bioactivity profiles, representing an enormous amount of information, must be handled. Dimension reduction and capturing of the important factors are the keys to solve the main problems of pharmacology. On the other hand, the practically infinite conformational space of a protein makes it hard to understand protein flexibility; dimension reduction is needed to find a smaller system in which flexibility can be examined. The activation of a human trypsin isoform, fine-tuned by a single residue at a specific position in the protein, offers us an ideal model system.



Superimposed structure of bovine trypsinogen (PDB ID: 1tgn) **and human trypsin 4** (PDB ID: 1h4w)**.** The peptide backbone segment of human trypsin 4 whose conformation differs from the conformation of bovine trypsinogen (gray) corresponding to the activation domain is shown as a colored ribbon. The 16-19 peptide segment is shown in yellow, the 142-152 segment is colored purple and the 184-194 and 216-223 segments are represented by a blue and a black ribbon, respectively. The backbone of residue 193, which goes through large dihedral angle transition in the course of activation, is shown in green. Tryptophan residues are also highlighted, Trp141, 215 and 221 that might account for the fluorescence intensity change during the conformational change are colored red, while Trp51 and Trp237 is shown in orange (chymotrypsin numbering system is used to identify the residues).

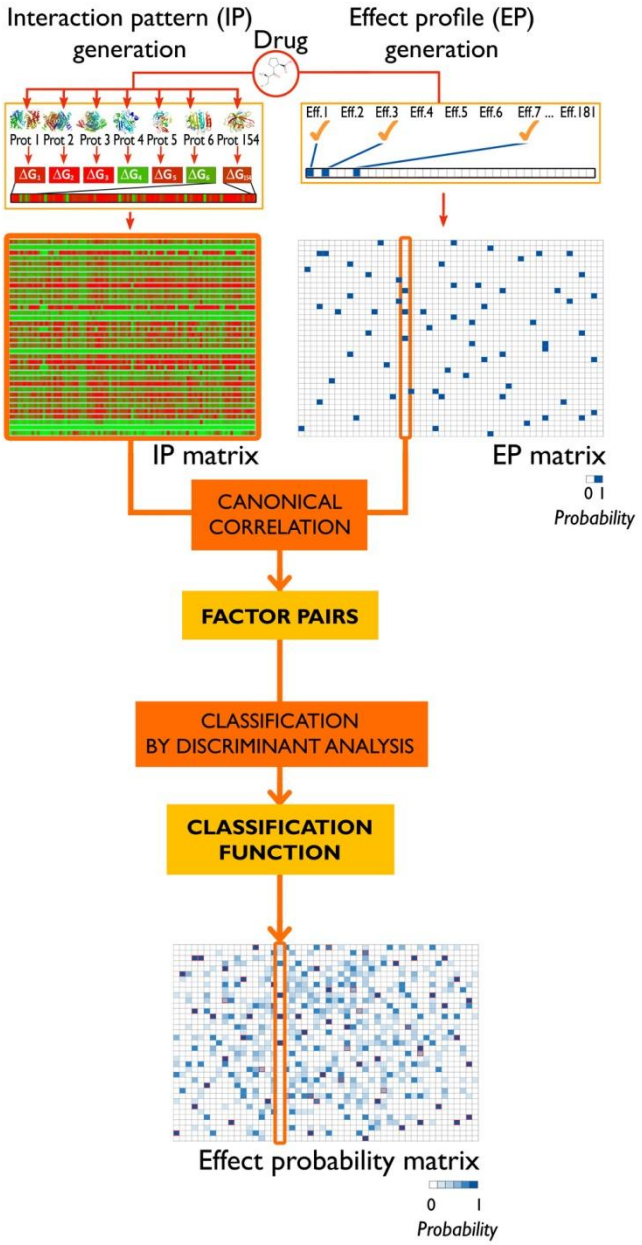
Aims

1. The primary aim of my PhD work was to develop an *in silico* system for the prediction of bioactivity properties of small-molecule compounds, applying the paradigm of polypharmacology.
2. We aimed to study the diversity of the protein set used to generate the interaction profiles. We also aimed to examine the importance of binding site shape in the determination of the affinity profiles of proteins.
3. Another aim of my work was to characterize the thermodynamics of a special monomolecular structural rearrangement and to unravel the role of the internal viscosity by introducing residues with different sizes at a specific hinge point.
4. In order to increase the accuracy of the thermodynamic parameters calculated from Arrhenius plots, we aimed to develop a combined heat-jump/stopped-flow equipment which allows the extension of the Arrhenius plots to higher temperatures without the denaturation of the studied enzyme.

Methods

Examination of the associations between Interaction Profiles, Effect Profiles and binding site descriptors

A drug molecule is *docked* to a set of 154 proteins (using different scoring functions) and the calculated lowest binding free energies are entered into a row vector, i.e. the Interaction Pattern (IP) (see figure). IPs of the 1,226 studied drugs form the IP matrix. The Effect Profile (EP) matrix contains the therapeutic effects of the drugs in a binary coded form (blue and white cells represent the presence and the absence of a given effect from the 181 categories, respectively). Then, *canonical correlation analysis* (CCA) is performed in order to generate highly correlating factor pairs that serve as the input for *linear discriminant analysis* (LDA). This way, classification functions are produced that yield the probability for each drug-effect pair, resulting in the effect probability matrix. (Note that the values in this matrix are continuous.) Classification accuracy was assessed by *Receiver Operating Characteristic* (ROC) analysis. In order to evaluate robustness of our results, a *cross-validation* with the leave-one-out procedure was performed. In order to assess the importance of binding site shape in binding affinity determination, a geometric descriptor matrix was produced for the



154 proteins. The "overlap" between the geometric descriptor set and the IP matrix was determined by *canonical redundancy analysis* (CRA). This method was applied to sensitivity tests, i.e., the examination of the role of the scoring functions. *Principal component analysis* (PCA) was used to study the complexity of the resulting data matrices.

Internal viscosity: the role of hinge residues in trypsin activation

We expressed wild type human trypsinogen 4 and its R193G/A/Y/F mutants and monitored their conformational change upon activation in *pH-jump stopped-flow* experiments by detecting the intrinsic tryptophan fluorescence change. (The pH jump triggers a conformational change identical to the activation of the zymogen form.) *Temperature dependence* of the rate of the conformational change was measured using conventional stopped flow between 5-38 °C and our newly developed heat-jump/stopped-flow between 34-60 °C. *Dependence* of the rate constants *on the relative external viscosity* was measured at 20 °C, using buffers with different amount of maltose as a viscogen, yielding relative external viscosity values ranging from 1 to 8.18. Data were evaluated using Kramers' rate theory modified by Ansari *et al* in which the preexponential term of the Arrhenius equation contains a friction parameter that can be separated to external (solvent) viscosity and an internal, viscosity-like parameter. (However, in contrast with solvent viscosity, this internal viscosity always belongs to a specific conformational change.) The model proposes that the rate of a conformational change is inversely proportional to the external viscosity.

Principle of the heat-jump/stopped-flow equipment

A heating loop was inserted between the substrate syringe and the mixing chamber of a conventional stopped-flow apparatus. The cuvette house was heated by another heating element. Reactant syringes are kept at room temperature. After pushing the syringes, the non-heat-sensitive reactant (generally, the substrate) passes the heating loop and reaches a temperature higher than the measuring temperature. The heat-jump occurs when this high-temperature solution is mixed with the other solution (generally, the enzyme) that was kept at room temperature. After thermal equilibration in the mixing chamber, the reactants reach the cuvette which is set to the resulting (i.e., measuring) temperature. (Otherwise, temperature equilibration would occur in the cuvette that would interfere with the optical signal used to follow the process of the reaction.)

Results

1. We constructed an IP matrix containing interaction profiles of 1,226 drugs on 154 proteins, applying AutoDock4 and X-SCORE scoring functions. We also constructed an EP matrix, containing the 181-bit-long bioactivity pattern of the same set of drugs. We built a third data matrix for the description of the binding site geometry of the applied proteins, using PocketPicker. This matrix contains 405 descriptors for 154 proteins each.
2. In order to determine the structural diversity of the applied protein set and to examine the importance of binding site shape in binding affinity determination, the MAF matrix (the transposed IP matrix, considering proteins as observations instead of variables) and the geometric descriptor matrix was subjected to PCA, CCA and CRA. Our analysis revealed that the protein set possesses a fair diversity in terms of binding site shape. PCA of the MAF and geometric descriptor variables revealed that the MAF set possesses a higher level of complexity than the geometric descriptor set. CCA resulted in 3 statistically significant canonical factor pairs with correlation values of 0.87, 0.84 and 0.77, respectively. CRA indicated that geometric descriptor factors explain 6.9 % of the variance of the MAF factor set while MAF factors explain 15.9 % of the total variance of geometric descriptor factors. Based on the salient structures of the factor pairs, a clear-cut association was identified between the shape and bulkiness of the drug molecules and the protein binding site descriptors.
3. Different scoring functions provided similar binding affinity patterns in terms of the level of canonical correlation between the IPs based on the different scoring functions and the

binding site geometry descriptor data. A high level of correlation was obtained between the AutoDock4 and the X-SCORE-based IP set.

4. Using a one-dimensional analysis procedure, i.e., an IP similarity measurement, a clear association was revealed between the IPs and EPs of drug molecules, i.e., molecules possessing similar IPs have similar bioactivity properties.

5. The resulting association was evaluated using a truncated dataset which did not contain any known target proteins of the drugs and practically the same results were obtained.

6. The method allows a highly efficient identification of similarities in effects and mechanisms of action of drugs despite structural diversity as it was revealed in many cases.

7. The association between IPs and EPs was quantified using CCA and LDA between the IP matrix and each effect fingerprints in the EP dataset (Drug Profile Matching). The accuracy of the resulting classification function for each effect was evaluated using ROC analysis. 82% of the studied effects resulted in an AUC value larger than 0.95, indicating that an excellent classification was obtained. Leave-one-out cross-validation was also performed in order to determine the robustness of our findings. In average, a 25-fold increase was yielded compared to a random classification.

8. Examination of the canonical loading structure of the target proteins in the classification functions revealed that their importance in the respective effect predictions is similar to other, non-target proteins. This finding is in concert with Result No. 5.

9. Several case studies, *in vitro* and cell culture tests were carried out to evaluate our predictions. Based on their accuracy and robustness results, "ACE inhibitor" and "COX inhibitor" predictions were selected for *in vitro* testing. More than 50% of predictions were confirmed. The predicted adrenergic and dopaminergic profile of two compounds were also confirmed in cell culture tests.

10. Thermodynamic analysis assessing the role of internal friction in a specific conformational rearrangement, i.e., trypsin 4 activation, revealed that the trypsin 4 mutants R193G/A/Y/F differ only in the preexponential term of the Arrhenius equation, while the activation energy is unaffected by the mutations.

11. Solvent viscosity dependence of the rate of the conformational transition of the wild-type trypsin and its R193G and R193A mutants was also determined. It was revealed that the rate is inversely proportional to the solvent viscosity.

12. The aforementioned phenomenon was interpreted in terms of the Kramers' theory and the relative internal viscosity of these mutants (regarding the conformational change upon activation) was determined ($\sigma_{R193G} = 0.27$, $\sigma_{R193A} = 0.81$, $\sigma_{wild\text{-}type} = 1.67$). It was revealed

that a bulkier side chain in the hinge region allows less conformational freedom for the peptide backbone in the conformational transition as compared to glycine which can be detected as an increase in internal viscosity.

13. A new, combined heat-jump/stopped-flow equipment was constructed in order to improve the accuracy of the thermodynamic parameters calculated from Arrhenius plots. The system was calibrated utilizing the temperature sensitivity of fluorescence, i.e., mixing the same fluorescent reactant from both syringes will result in a fluorescence change if the temperature of the heating loop and the cuvette house is miscalibrated. Based on these results, temperature pairs were determined for a number of measuring temperatures.

14. Dead time of the newly developed equipment was also determined and it was found to be similar to that of the unmodified equipment (0.9 ms).

Conclusions

Two levels of complexity were addressed in my thesis. In the first part, I presented an approach to predict the effect profiles of drug molecules. Then, I assessed protein flexibility by determining the internal viscosity of an interdomain conformational rearrangement. The methodology applied to study the two problems was different; nevertheless, we were facing the same theoretical problem in both cases, i.e., the handling of complexity and the challenge of the extraction of the relevant features of a system.

1. One-dimensional and multidimensional analyses unveiled a strong correlation between the EPs and IPs. We proved that target proteins are not necessarily for effect prediction as it was proved using one-dimensional and multidimensional analyses. To our knowledge this is the first method which directly relates distant levels of information, i.e., the information from the atomic interactions with the information from physiological effects. Based on these quantitative correlations, hidden effects can be revealed for the existing drugs and their entire effect profiles can be predicted. The accuracy and the robustness of our effect prediction method were evaluated by successful *in vitro* analyses. Unlike other similarity-based approaches, no direct topological similarity information on drug molecules is involved in the method; therefore, our approach is able to detect effect profile similarities even in the case of limited structural similarity between compounds. The good predictive power of our approach gives an opportunity to its use with marketed drugs or as a preclinical screen, increasing the efficacy of drug development.

2. Our analysis of the importance of binding site geometry revealed that, except for few specific cases, the shapes of the binding pockets have relatively low weights in the determination of the affinity profiles of proteins. Since the MAF profile is closely related to the target specificity of ligand binding sites we can conclude that the shape of the binding site is not a pivotal factor in selecting drug targets. Nonetheless, based on strong specific associations between certain MAF profiles and specific geometric descriptors we identified, the shapes of the binding sites do have a crucial role in virtual drug design for certain drug categories, including morphine derivatives, benzodiazepines, barbiturates and antihistamines. Therefore we conclude that the application of shape-based drug design methodologies might prove better performance on this drug set than that for others.

3. The role of internal friction was also studied in the thesis, applying human trypsin 4 activation as a model system. We found that an increase in the size of the side chain is associated with the decrease of the reaction rate constant. Our data show that the mutations do not affect the activation energy (the exponential term) of the reaction, but they significantly alter the preexponential term of the Arrhenius expression. The effect of solvent viscosity on the rate constants of the conformational change during activation of the 193G and 193A mutants were determined and evaluated by Kramers' theory. Based on this, we determined the internal viscosity parameter of the activation of the wild type trypsin and its R193A and R193G mutants experimentally. We conclude that the reaction rate of the studied conformational transition is regulated by the internal molecular friction which can be specifically modulated by mutagenesis in the hinge region. Extended by the investigation of the temperature dependence of internal viscosity in the future, our results can be the basis of a new scientific model explaining the physical background of internal friction of enzyme reactions.

4. In order to expand the available temperature range for the thermodynamic characterization of the activation of the wild-type trypsin and its mutants, a novel heat-jump/stopped-flow apparatus was constructed and calibrated. The main advantage of this setup is that the fast heat-jump occurs simultaneously with the rapid mixing of the reactants so dead time of mixing and heat-jump are identical if the equipment is calibrated correctly. Consequently, all enzyme reactions can be followed that are faster than the generally slow temperature denaturation.

Papers related to the work presented in the thesis:

1. Zoltán Simon, Margit Vigh-Smeller Ágnes Peragovics, Gábor Csukly, Gergely Zahoránszky-Kőhalmi, Anna Á Rauscher, Balázs Jelinek, Péter Hári, István Bitter, András Málnási-Csizmadia, Pál Czobor: **Relating the shape of protein binding sites to binding affinity profiles: is there an association?**
BMC Struct Biol 2010, **10**:32.

2. Júlia Tóth*, Zoltán Simon*, Péter Medveczky, Linda Gombos, Balázs Jelinek, László Szilágyi, László Gráf, András Málnási-Csizmadia: **Site directed mutagenesis at position 193 of human trypsin 4 alters the rate of conformational change during activation: role of local internal viscosity in protein dynamics**.
Proteins 2007, **67**(4):1119-1127.
*equal contribution

3. Bálint Kintses, Zoltán Simon, Máté Gyimesi, Júlia Tóth, Balázs Jelinek, Csaba Niedetzky, Mihály Kovács, András Málnási-Csizmadia: **Enzyme kinetics above denaturation temperature: a temperature-jump/stopped-flow apparatus.**
Biophys J 2006, **91**(12):4605-4610.

Other publications related to the work presented in the thesis:

1. Zahoránszky Gergely, Simon Zoltán, Zhenhui Yang, Hári Péter, Málnási-Csizmadia András: **„MIF: a gyógyszerkutatás új korszakának hírnöke"** *Biokémia*, XXXII. évf., 4. szám (2008), Budapest, Hungary
2. Simon Zoltán, Zahoránszky Gergely, Zhenhui Yang, Jelinek Balázs, Hetényi Csaba, Hári Péter, Bitter István, Málnási-Csizmadia András: **„Gyógyszerhatások előrejelzése molekuláris interakciós ujjlenyomat segítségével"** Proceedings of the Hungarian Biochemistry Association (2008), Szeged, Hungary
3. Rauscher Anna, Simon Zoltán, Gráf László, Málnási-Csizmadia András: **„A konformációváltozás során fellépő belső súrlódás hőmérsékletfüggése a humán tripszin 4 modellrendszerben"** Proceedings of the Hungarian Biochemistry Association (2008), Szeged, Hungary
4. Zoltán Simon, Júlia Tóth, Péter Medveczky, Linda Gombos, Balázs Jelinek, László Szilágyi, László Gráf, András Málnási-Csizmadia: **"Site Directed Mutagenesis at Position 193 of Human Trypsin 4 Alters the Rate of Conformational Change during Activation: Role of Local Internal Viscosity in Protein Dynamics"** IV. International Conference on Molecular Recognition (2007), Pécs, Hungary
5. Simon Zoltán: **„Molekuláris Interakciós Ujjlenyomat: Új megközelítés a gyógyszertervezésben"** Proceedings of the Hungarian Biochemistry Association (2006), Pécs, Hungary