

## Welcome

It is the sixth anniversary of our workshop, far from yet being routine (which it will never turn into!) but close of turning into a tradition as a well-established scientific event challenged by the kind enough claim of being the “European Gordon Conference on CAMD”. My organization of the “US CAMD Gordon Conference” in 2009 forced us to depart from our biannual frequency and so we took a gap of three years. In the years before we passed through “Virtual Screening”, “ADMET”, “Omics”, “Fragments”, and after “Merging Chemical and Biological Spaces” we want to focus this time on the “Interactome”. The interactome starts at its lowest level with atoms, proceeds via molecules to larger and larger assemblies to end up with cells and entire organisms. What controls the interplay of all these components and what is their role in a disease situation? Can we learn from these considerations for our own discipline, the design and the development of novel drugs? Hopefully, with this meeting we can initiate a discussion on these aspects.

As for the last events, the charming castle of Rauischholzhausen is the venue for our meeting. Once again the applications for the workshop exceeded the number of available places, and we tried to accept a maximum of colleagues to generate this well-received dense and creative atmosphere that favours long discussions and fruitful exchanges of ideas from early breakfast until late glasses of beer in the Schlosskeller. We hope that new contacts established during the meeting may constitute the starting point for novel joint-ventures and long-lasting friendships. Since some faces turn up again every second year, we are confident that some success is achieved along these lines.

Also this year we are extremely grateful for the generous support of various companies even though the number of potential donors is decreasing due to ongoing mergers and the economic situation is tough throughout. As the budget of public funding for such events hardly exists, such meetings would be impossible without donations from industry and kind sponsoring of travelling costs.

We are also extremely grateful for the valuable assistance of Ms. Lydia Hartleben and all members of the Marburg drug design group: without their help, this event could not have been brought to life!

Gerhard Klebe



## Acknowledgement

The workshop has been generously sponsored by the companies listed below:

Actelion	Basel, Switzerland
Astra Zeneca	London, United Kingdom
Bayer-Schering AG	Leverkusen, Germany
Boehringer Ingelheim	Biberach, Germany
Hoffmann - La Roche AG	Basel, Switzerland
Intervet Innovation GmbH	Schwabenheim, Germany
Merck KGaA	Darmstadt, Germany
Sanofi Aventis	Frankfurt, Germany
Solvay Pharmaceuticals GmbH	Hannover, Germany

# Contents

<b>1 Program</b>	<b>8</b>
<b>2 Talks</b>	<b>14</b>
1. Martin Stahl . . . . .	15
2. Johan Åquist . . . . .	17
3. Bernd Engels . . . . .	19
4. Rebecca C. Wade . . . . .	21
5. Shoshana J. Wodak . . . . .	23
6. Roberto Mosca . . . . .	25
7. Volkhard Helms . . . . .	27
8. Michael Kuhn . . . . .	29
9. Jordi Mestres . . . . .	31
10. Johanna Fehr . . . . .	33
11. Jürgen Bajorath . . . . .	35
12. Xavier Barril . . . . .	37
13. Helena Danielson . . . . .	39
14. Steve Homans . . . . .	41
15. Marc-Emmanuel Dumas . . . . .	43
16. Günter Mayer . . . . .	45
17. Carsten Schultz . . . . .	47
18. Daniel Bur . . . . .	49
19. Christoph W. Müller . . . . .	51
20. Jane A. Endicott . . . . .	53
21. Stefan Knapp . . . . .	55
22. Eric Meggers . . . . .	57
23. Hans-Werner Mewes . . . . .	59
<b>3 Posters</b>	<b>62</b>
1. P. Czodrowski . . . . .	63
2. P. Daldrop, R. Brenk . . . . .	64
3. H.W.H. Göhlmann, B. Wroblowski, W. Talloen, D. Wuyts, S. Dickens, K. Smans, P. Peeters, S. Hochreiter, G. Klambauer, Z. Shkedy . . . . .	65
4. E. Gregori-Puigjané, J. Hert, M. J. Keiser, C. Laggner, J.J. Irwin, B.K. Shoichet . . . . .	66
5. S. Forte, G. Rappazzo, G.L. Volti, S. Guccione . . . . .	67
6. B. Grüning, C. Senger, S. Günther . . . . .	69
7. A. Jahn, G. Hinselmann, N. Fechner, C. Henneges, A. Zell . . . . .	70
8. G.E. Kellogg, V.N. Koparde, P.D. Mosier, J.N. Scarsdale . . . . .	71
9. O. Koch, L. Flohé, T. Jaeger, P.M. Selzer . . . . .	72
10. O. Korb, S. Bowden, T. Olsson, D. Frenkel, J. Liebeschuetz, J. Cole . . . . .	73

11. C. Luca, A. Hauser, M. Jung, W. Sippl . . . . .	74
12. D. Muthas, A. Krasowski, R. Brenk . . . . .	75
13. A. Perdih, M. Brvar, M. Oblak, L.P. Mašič, T. Solmajer . . . . .	77
14. M. Pippel, R. Meier, W. Sippl . . . . .	78
15. L. Rosenbaum, A. Jahn, G. Hinselmann, N. Fechner and A. Zell	79
16. L. Ruddigkeit, R.v. Deursen, L. Blum and J.-L. Reymond . . . . .	80
17. P. Schmidtke, X. Barril . . . . .	81
18. J. Seco . . . . .	82
19. S. Shekfeh, S.D. Furdas, M. Jung, W. Sippl . . . . .	83
20. H. Rode, M. Sos, R. Thomas and D. Rauh . . . . .	84
21. J. Simard, C. Grütter, M. Getlik and D. Rauh . . . . .	85
22. E. Specker, M. Neuenschwander, M. Lisurek, B. Rupp, R. Kühne, J.P.v. Kries and R. Frank . . . . .	86
23. A. Tomić, J. Špoljarić, M. Abramić, S. Tomić . . . . .	87
24. S. Vogel, M. Bauer, F. Boeckler . . . . .	88

**4 List of Participants 90**



# 1 Program

**Monday, March 22, 2010**

- 15:00      Opening and Introduction  
15:15      **Martin Stahl**, Roche Basel, Switzerland  
*A Medicinal Chemist's Guide to Molecular Interactions*
- 16:00      Coffee
- 16:30      **Johan Åquist**, University of Uppsala, Sweden  
*Force Field Based Scoring of Protein-Ligand Binding Affinities*
- 17:15      **Bernd Engels**, University of Würzburg, Germany  
*Are Molecular Crystals indeed good Approximation for Protein Surroundings?*
- 19:00      Dinner
- 20:30      Poster session

**Tuesday, March 23, 2010**

- 9:00 **Rebecca Wade**, Heidelberg Institute for Theoretical Studies (HITS), Germany  
*Molecular Interaction Fields: From Drug Design to Systems Biology and back*
- 9:45 **Shoshana Wodak**, University of Toronto, Dept. of Biochemistry, Canada  
*Interaction Proteomics, the Bounty and the Challenges*
- 10:30 Coffee
- 11:00 **Roberto Mosca**, Institute for Research in Biomedicine, Barcelona, Spain  
*Structural Systems Biology: Modelling Protein Interactions and Complexes*
- 11:45 **Volkhard Helms**, Univ. of Saarbrücken, Germany  
*Designing Transient Binding Pockets on Protein Surfaces*
- 13:00 Lunch
- 15:00 **Michael Kuhn**, TU Dresden, Germany  
*Predicting Drug Targets and Finding the Molecular Basis of Side Effects*
- 15:45 **Jordi Mestres**, IMIM Institut Municipal d'Investigació Mèdica, Barcelona, Spain  
*Chemical Probes for Systems Biology*
- 16:30 Coffee
- 17:00 **Johanna Fehr**, Inst. of Pharmacology, Philipps-University Marburg, Germany  
*Common Pathways Database – Development of a Tool to Systematically Support the Discovery of Multiple-Indication Drugs*
- 17:45 **Jürgen Bajorath**, B-IT, LIMES, University of Bonn, Germany  
*Molecular Networks and SAR Analysis*
- 19:00 Dinner
- 20:00 Musical Intermezzo: UFA meets Jazz

**Wednesday, March 24, 2010**

- 9:00 **Xavier Barril**, Dept. Fisicoquímica, Facultat de Farmàcia, Barcelona (Spain)  
*Binding Site Detection and Druggability Predictions: the Solvent's Perspective*
- 9:45 **Helena Danielson**, Uppsala University, Sweden  
*SPR Biosensor Generated Interaction Kinetic Data for Identification and Characterization of Leads*
- 10:30 Coffee
- 11:00 **Steve Homans**, Univ. of Leeds, United Kingdom  
*Prediction of Affinity from Structure, Entropic Considerations of Protein-Ligand Binding*
- 11:45 **Marc E. Dumas**, Imperial College London, United Kingdom  
*Genetic Mapping of Metabolic Profiles: Genomic and Metagenomic Influences and Consequences for Identification of Drug Targets*
- 13:00 Lunch
- 15:00 **Günter Mayer**, University of Strathclyde, Glasgow, Scotland  
*The Chemical Biology of Aptamers*
- 15:45 **Carsten Schultz**, EMBL-Heidelberg, Germany  
*New Tools for Imaging and Modulation of Intracellular Signaling Networks*
- 16:30 Coffee
- 17:00 **Daniel Bur**, Actelion, Basel, Switzerland  
*Search for Multifunctional Plasmeprin Inhibitors as Antimalarial Drugs*
- 17:45 **Christoph Müller**, EMBL-Heidelberg, Germany  
*Molecular Mechanisms of Chromatin Targeting*
- 19:00 Dinner
- 20:30 Round-table discussion

**Thursday, March 25, 2010**

- 9:00      **Jane A. Endicott**, Univ. of Oxford, United Kingdom  
*Targeting the CDK Family for Anti-Cancer Drug Design*
- 9:45      **Stefan Knapp**, Structural Genomics Center, Oxford,  
United Kingdom  
*Targeting Signalling Networks in Cancer*
- 10:30     Coffee
- 11:00     **Eric Meggers**, Univ. of Marburg, Germany  
*Organometallic Complexes as Emerging Scaffolds for the De-  
sign of Enzyme Inhibitors*
- 11:45     **Hans-Werner Mewes**, Helmholtz Center Munich, Ger-  
many  
*Drug Development: Puzzle or Mystery?*
- 13:00     Lunch
- 14:00     End of Workshop



## 2 Talks

Monday, 15:15

## **A Medicinal Chemist's Guide to Molecular Interactions**

**Martin Stahl**

Roche Basel, Switzerland

Detailed knowledge about non-covalent intermolecular interactions is an essential precondition for structure-based design. At the same time, solvent, conformational, entropy and cooperativity contributions all contribute to molecular recognition events, making it hard to find a clear relationship between structural and thermodynamic information. So what guidance can be given to medicinal chemists without the risk of simplification and over-interpretation?

We show that a combination of multiple approaches - careful analysis of structure-activity relationships, of crystal structures, and experimental and computational work on model systems - can effectively be utilized to derive knowledge useful for prospective applications. Search results from Cambridge Structural Database and the Protein Data Bank are presented and discussed with a focus on unbiased searches and intuitive visualization of geometric relationships. Multiple examples from medicinal chemistry programs will be shown.

Notes

Monday, 16:30

## **Force Field Based Scoring of Protein-Ligand Binding Affinities**

**Johan Åquist**

Univ. of Uppsala, Sweden

We will discuss recent advances in applying molecular mechanics based scoring methods to protein-ligand complexes. Some key issues that will be addressed are sensitivity to the 3D receptor model, treatment of solvation, effects of conformational sampling, discrimination between binding modes and the prospects for high-throughput applications.

Notes

Monday, 17:15

## **Are Molecular Crystals indeed good Approximations for Protein Surroundings?**

**Bernd Engels**

Institut für Physikalische und Theoretische Chemie, Universität Würzburg,  
Germany

Beside nuclei positions high-resolution X-ray experiments can also provide information about the electron density of the studied molecule. According to the Hohenberg-Kohn theorem this density determines the bonding situation within the molecule but also entails its interactions with the surrounding including properties as the chemical reactivity and possible intermolecular bondings. The former is important to predict the chemistry of a given molecule. The latter could be a useful tool for rational drug design since the molecular recognition process is determined by the intermolecular interactions between target and active compounds. However, such investigations depend on the supposition that crystal and enzyme environments influence the electron density of a given compound in a similar manner. In the present talk we investigate this assumption. As an example for reversible ligands the trans-4-(aminomethyl)cyclohexane-1-carboxylic acid (AMCHA) is used while E64c is studied to provide information for irreversible ligands. Their densities are computed in different environments including crystal, enzyme, polar solvent and gas phase. The analysis of the variations shows indeed small differences for reversible inhibitors but considerably larger one for irreversible inhibitors.

Notes

Tuesday, 9:00

## Molecular Interaction Fields: From Drug Design to Systems Biology and back

Rebecca C. Wade

Molecular and Cellular Modeling Group, Heidelberg Institute for  
Theoretical Studies (HITS),  
Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany  
E-mail: [rebecca.wade@h-its.org](mailto:rebecca.wade@h-its.org)

Molecular interaction Fields (MIFs) describe the interactions possible between a chemical probe and a target molecule. The probe may be a small molecule, e.g. water, a fragment of a molecule, e.g. an aromatic ring, or have specific physicochemical properties, e.g. a hydrophobic probe. The MIF for a proton corresponds to the molecular electrostatic potential. MIFs are usually computed on a three-dimensional grid of points surrounding the target molecule. The analysis of MIFs has a wide range of applications<sup>1</sup> and some of our recent developments based on PIPSA will be discussed in this talk. PIPSA (Protein Interaction Property Similarity Analysis) is a procedure to compare the interaction properties of a set of proteins of similar three-dimensional structure (see [pipsa.eml.org](http://pipsa.eml.org), including the webPIPSA server). PIPSA can aid the identification of selective sites to target in drug design<sup>2</sup>, the functional annotation of protein families<sup>3</sup>, and the analysis of proteins along a biochemical pathway<sup>4</sup>. qPIPSA (quantitative PIPSA) is a development to aid the estimation of kinetic parameters for mathematical modeling of biochemical pathways by using protein structural information<sup>5,6</sup>. qPIPSA has been incorporated into the SYCAMORE platform (Systems Biology Computational and Modelling Research Environment; [sycamore.eml.org](http://sycamore.eml.org)) which guides the scientist through the process of setting up and carrying out biochemical network simulations<sup>7</sup>.

### References:

- <sup>1</sup> Wade, R.C. In 'Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction', Ed. Cruciani, G., *Wiley-VCH*, Weinheim. (2005) Ch. 2, pp27-42.
- <sup>2</sup> Henrich, S., Richter, S., Wade, R.C. *ChemMedChem* (2008) **13**, 413-417.
- <sup>3</sup> Winn, P.J., Religa, T.L., Battey, J.D., Banerjee, A., Wade, R.C. *Structure* (2004) **12**, 1563-74.
- <sup>4</sup> Stein, M., Gabdoulline, R.R., Wade, R.C. *Mol. BioSyst.* (2010) **6**, 162-174.
- <sup>5</sup> Gabdoulline, R. R., Stein, M., Wade, R.C. *BMC Bioinformatics* (2007) **8**, 373.
- <sup>6</sup> Stein, M., Gabdoulline, R.R., Wade, R.C. *Curr. Opin. Struct. Biol.* (2007) **17**, 166-172.
- <sup>7</sup> Weidemann, A., Richter, S., Stein, M., Sahle, S., Gauges, R., Gabdoulline, R., Surovtsova, I., Semmelrock, N., Besson, B., Rojas, I., Wade, R., Kummer, U. *Bioinformatics*, (2008) **24**, 1463-4.

Notes

Tuesday, 9:45

## Interaction Proteomics, the Bounty and the Challenges

**Shoshana J. Wodak**<sup>1,2,3</sup>, Jim Vlasblom<sup>1,2</sup>, Shuye Pu<sup>1</sup>, Andrei Turinsky<sup>1</sup>  
and Brian Turner<sup>1</sup>

<sup>1</sup> Molecular Structure and Function Program, Hospital for Sick Children,  
555, University Avenue, Toronto, ON M5G 1X8, Canada,

<sup>2</sup> Department of Biochemistry University of Toronto, Toronto, ON, Canada

<sup>3</sup> Department of Molecular Genetics, University of Toronto, Toronto, ON,  
Canada

Virtually all cellular processes are carried out by groups of proteins that interact with one another and with other cellular components. Large-scale characterization of these interactions is key for gaining system-level understanding of the functional context of individual proteins. This understanding can then be exploiting for more accurate target identification and ultimately for designing effective drugs to cure diseases. We will review the major high throughput techniques for probing protein association. Their complementary nature, the data they produce, and some of the biases from which they suffer will be discussed. The important role of computational methods in processing, analyzing, and validating the large body of noisy data produced by the experimental procedures will be highlighted, and the status of literature curated data protein interactions stored in specialized databases will be evaluated.

Notes

Tuesday, 11:00

## **Structural Systems Biology: Modelling Protein Interactions and Complexes**

**Roberto Mosca**, Patrick Aloy

Institute for Research in Biomedicine, Barcelona, Spain

Much of systems biology aims to predict the behavior of biological systems on the basis of the set of molecules involved. Understanding the interactions between these molecules is therefore crucial to such efforts. Although many thousands of interactions are known, precise molecular details are available for only a tiny fraction of them. The difficulties that are involved in experimentally determining atomic structures for interacting proteins make predictive methods essential for progress. Structural details can ultimately turn abstract system representations into models that more accurately reflect biological reality.

Notes

Tuesday, 11:45

## Designing Transient Binding Pockets on Protein Surfaces

Volkhard Helms

Univ. of Saarbrücken, Germany

Protein surfaces involved in the formation of protein-protein contacts are typically quite flat and unstructured both in the unbound and in the complex forms. Still, small molecules are sometimes able to induce the opening of suitable pockets at this interface and bind. We showed that in molecular dynamics simulations of solvated unbound proteins transient pockets open frequently nearly everywhere on the protein surface<sup>1</sup>. When testing alternative methods to generate conformational ensembles, the tCONCOORD method turned out as a computationally efficient alternative to the costly molecular dynamics simulations<sup>2</sup>. In order to induce pockets with defined properties at given locations, we have developed a pocket design algorithm based on the A\* algorithm<sup>3</sup>. As an application, we will present results from a combined experimental and computational study to characterize how polyamine molecules affect the electron transfer chain between adrenodoxin reductase, adrenodoxin and a member of the cytochrome P450 family<sup>4</sup>.

### References:

<sup>1</sup> Eyrisch, S., Helms, V. (2007) *J. Med. Chem.* **50**, 3457-3464

<sup>2</sup> Eyrisch, S., Helms, V. (2009) *J. Comp. Aid. Mol. Des.* **23**, 73-86

<sup>3</sup> Eyrisch, S. and Helms, V. (2008) *GCB 2008 Proceedings*.

<sup>4</sup> Berwanger, A., Eyrisch, S., Scchuster, I., Helms, V., Bernhardt, R. (2009, in press) *J. Inorg. Biochem.*

Notes

Tuesday, 15:00

## **Predicting Drug Targets and Finding the Molecular Basis of Side Effects**

**Michael Kuhn**

TU Dresden (Biotec, Andreas Beyer group) and  
MPI-CBG (Tony Hyman group)

Side effects of medicines seem to be an entirely unwelcome feature of drug treatment. However, the information stored in drug labels actually provides insights into the phenotypes upon chemical perturbation in humans. In a first approach, we developed a measure of side-effect similarity for drugs to find drugs that are likely to share targets. Applied to 746 marketed drugs, a network of 1018 side effect-driven drug-drug relations became apparent, 261 of which are formed by chemically dissimilar drugs from different therapeutic indications. We experimentally verified 13 out of 20 predictions for shared drug targets. Of these, 11 revealed inhibition constants equal to less than 10 micromolar. Nine of these were tested and confirmed in cell assays. Secondly, we performed a large-scale analysis of the connections between proteins and side effects. We thus show that the phenotypic data generated during clinical trials makes it possible to uncover causal associations between targeting proteins and side effects. The causality of the associations is confirmed in the literature and by occurrence of the predicted phenotypes in knockout mice. Furthermore, we were able to predict new drug targets by building side effect profiles for proteins. Underlying tools for this work are STITCH (<http://stitch.embl.de>), a database of protein-chemical interactions, and SIDER (<http://sideeffects.embl.de>), a collection of drugs and their side effects. Both databases can be browsed online, or downloaded for computational analyses.

Notes

Tuesday, 15:45

## Chemical Probes for Systems Biology

**Jordi Mestres**

IMIM Institut Municipal d'Investigació Mèdica, Barcelona, Spain

The identification of a small-molecule modulator for each individual function of all human proteins has been proposed as one of the grand challenges for chemical biology in the years to come. In an attempt to coordinate efforts towards this ambitious goal, the US National Institutes of Health (NIH) launched the Molecular Libraries and Imaging (MLI) initiative that, in a pilot phase (2004-2008), led to the development of ten high-throughput screening (HTS) centers that screened 691 assays, covering 171 targets and resulting in the identification of 64 chemical probes.

However, taking a decision on the exact range of values for the physico-chemical and pharmacological properties that a small molecule should have to be considered a useful research tool to probe biology is difficult. The current potency and selectivity criteria for qualifying a small molecule as chemical probe are focussed on studying the role of biological targets but the question is whether these same criteria are the most adequate to probe biological systems with small molecules.

Notes

Tuesday, 17:00

## **Common Pathways Database – Development of a Tool to Systematically Support the Discovery of Multiple-Indication Drugs**

**Johanna Fehr**

Philipps-University Marburg, Institut für Pharmakologie und Toxikologie,  
Philipps-Universität Marburg

Drugs that can be used to treat more than one disease not only benefit patients, but they also have proven economic advantages for the pharmaceutical companies that develop them. Although a great deal of information is available about signaling pathways and functional relationships between drug targets, so far the development of these multiple-indication drugs has been a result of chance. The goal of this work was to establish a systematic methodology for identifying drug targets and signaling pathways that are common in more than one disease (common targets/common pathways).

A Microsoft® Office Access-based database was built that automatically and systematically analyzes connections of therapeutic areas and indication groups via common drug targets and signaling pathways.

This common pathway database allows for the identification of hot-spots of connectivity between therapeutic areas and indication groups. These hot-spots encompass common physiological functions and therefore have the highest chance to be targetable by multiple-indication R&D projects. Furthermore the tool highlights those pathways that are most important for the connection of selected indication groups. For a selected combination of indication groups and pathways, the tool automatically creates a list of both known common targets as well as potential new common target candidates. Due to their role in the common pathway, the common target candidates have an increased likelihood to be of use for both indication groups in the future. An adjacent prioritization methodology evaluates common target candidates and allows separating more promising candidates from less optimal ones.

Pharmaceutical companies and academic research groups could use this novel methodology to systematically support research and development of common targets and pathways, which may ultimately lead to the development of multiple-indication drugs.

Notes

Tuesday, 17:45

## Molecular Networks and SAR Analysis

**Jürgen Bajorath**

Department of Life Science Informatics, B-IT, LIMES Program Chemical  
Biology and Medicinal Chemistry, University of Bonn, Dahlmannstrasse 2,  
D-53113 Bonn

E-mail: [bajorath@bit.uni-bonn.de](mailto:bajorath@bit.uni-bonn.de)

The exploration of structure-activity relationships (SARs) plays a fundamental role in medicinal chemistry. Traditionally, SARs have been studied on a case-by-case basis, both in computation and experiment. Computational methods and graphical analysis tools are introduced for systematic mining of SAR information in sets of active compounds including HTS data. Similarity-based molecular network representations play a central role in graphical exploration. Using numerical SAR analysis functions and molecular networks, SAR information can be extracted from compound data sets of any source and SAR rules can be derived. Results of systematic SAR analysis show that SAR information contained in large data sets is usually heterogeneous and multi-layered. Many activity landscapes contain SAR microenvironments representing different local SAR components.

Notes

Wednesday, 9:00

## **Binding Site Detection and Druggability Predictions: the Solvent's Perspective**

**Xavier Barril**

ICREA Research Professor at Barcelona University  
Dept. Físicoquímica, Facultat de Farmàcia  
Av. Joan XXIII, s/n. 08028 - Barcelona (Spain)

As shown by X-ray crystallography and NMR experiments, binding sites have a natural tendency to recognize small organic molecules. This property has been used to characterize binding sites (solvent mapping) and to estimate their druggability (fragment screening by NMR). We have shown that molecular simulations with explicit solvent reproduce this behavior and can be used to detect binding sites and predict the maximal affinity that a drug-like molecule could attain for them (*J. Med. Chem.* 2009, 2363). More recently, we have evaluated how different approximations (e.g. implicit solvent, rigid solute) impact on the quality of the results, providing valuable clues to understand the contributions of various terms to the binding events. This will be presented along with other research carried out in our group, highlighting the leading role played by the solvent in molecular recognition processes.

Notes

Wednesday, 9:45

## **SPR Biosensor Generated Interaction Kinetic Data for Identification and Characterization of Leads**

**Helena Danielson**

Uppsala University, Sweden

The use of SPR-biosensor technology for lead identification and characterization will be presented. The focus will be on screening and the subsequent steps in the hit to lead discovery process. An overview of experimental design and data analysis strategy will be illustrated together with examples of different approaches that can be used to validate hits. This includes distinguishing between specific/distinct interactions to a defined binding site from weaker non-specific interactions to other sites or the general protein surface, the identification of the binding site of interest, the stoichiometry and the basic characteristics of the interaction. Methods for more detailed characteristics of hits/leads and the use of analogues series, protein variants and modifications of experimental conditions will also be illustrated. A variety of examples will be used to demonstrate how SPR biosensor-based strategies are implemented for lead discovery.

Notes

Wednesday, 11:00

## **Prediction of Affinity from Structure: Entropic Considerations of Protein-Ligand Binding**

**Steve Homans**

Astbury Centre for Structural Molecular Biology, Faculty of Biological Sciences, University of Leeds LS2 9JT UK

All biological processes depend critically on highly specific recognition between molecules with carefully tuned affinities. Despite the universal nature of these interactions, our understanding of their molecular basis is severely limited. For example, despite a number of successes, it is still extraordinarily difficult to exploit routinely high-resolution structural data for a given complex in order to design molecules that inhibit binding. In other words, it is not trivial to predict binding affinity from structure. In the face of the emergence or re-emergence of diseases such as age-related neurodegenerative disorders or the relentless progress of antibiotic resistant bacterial strains that may soon reach epidemic proportions, the ability to design novel ligands at will that inhibit biomolecular interactions remains one of the major challenges in contemporary science.

While the crystal or NMR structure of a protein is unquestionably thought provoking in the process of ligand design, the key to understanding the affinity of a ligand for its receptor lies in the dynamics and thermodynamics of the association rather than a simple static picture. With the advent of technologies such as isothermal titration calorimetry (ITC), it is possible under ideal circumstances to obtain reliable experimental data on the global thermodynamic parameters governing a biomolecular association. However, from the point of view of ligand optimisation, it would be of immeasurable benefit to obtain these thermodynamic parameters experimentally on a per-residue, rather than global basis. This presentation will describe some of our recent results using NMR in combination with ITC, protein crystallography, computational chemistry and site-directed mutagenesis in order to delineate the thermodynamics of ligand-protein associations in model systems.

Notes

Wednesday, 11:45

## **Genetic Mapping of Metabolic Profiles: Genomic and Metagenomic Influences and Consequences for Identification of Drug Targets**

**Marc-Emmanuel Dumas**

Imperial College London, Biomolecular Medicine, Dept of Surgery and Cancer, Sir Alexander Fleming Building, Exhibition Road, South Kensington, London SW7 2AZ, UK.

The study of human multifactorial diseases like insulin resistance, or complex biological processes such as ageing, represents a real healthcare challenge for the western and developing world. In this regard, high-throughput “-omics” biotechnologies like genomics, transcriptomics and metabolomics are invaluable tools for these pathologies. Integration of metabolic profiles with genome-wide genotyping and expression profiling data provides a platform to identify biomarkers and susceptibility genes for pathological components of the cardio-metabolic syndrome (glucose intolerance, insulin resistance, dyslipidemia, hypertension, obesity). Metabolomic Quantitative Trait Locus (mQTL) mapping consists of the robust and accurate statistical integration of genome-wide genotyping (single nucleotide polymorphisms, microsatellites) and metabolome-wide profiling by NMR spectroscopy and Mass spectrometry, to identify candidate biomarkers and susceptibility genes in rodent models of human disease. Several signal processing and statistical developments were performed in order to enhance signal recovery, locus detection and biomarker identification. Validation of candidate mQTL studies is now underway with eQTL studies. From a network biology angle, candidate genes and metabolites are then mapped onto biological networks, allowing an efficient visualisation of complex data, allowing the identification of mechanistic arguments, explaining the influence of gene variants on metabolic profiles and eventually disease phenotypes.

Notes

Wednesday, 15:00

## The Chemical Biology of Aptamers

**Günter Mayer**

University of Strathclyde, Glasgow, UK

Our research focuses on the application of combinatorial chemistry methods for the functional analysis of biomolecules. In combination with chemical biology approaches we aim at the development of biomolecular tools allowing both the precise investigation of signal transduction cascades and the development of novel therapeutic strategies.

We apply in vitro selection techniques for obtaining aptamers that bind to a defined target molecule. Aptamers are short single stranded nucleic acids that fold into a well-defined 3D-structure thereby binding with high affinity and specificity to diverse targets, ranging from small molecules to proteins and even living cells. We identified aptamers that bind to cell-surfaces, proteins, and non-coding RNA molecules. Employing organic synthesis and based on these aptamers we develop novel molecular tools for the investigation of the function and the identity of the cognate target molecules. The aptamers are further applied to validate the targets as pharmaceutically relevant and for the development of novel therapeutic approaches.

Notes

Wednesday, 15:45

## **New Tools for Imaging and Modulation of Intracellular Signaling Networks**

**Carsten Schultz**

EMBL-Heidelberg, Germany

With the development of new methods and models describing the complex molecular networks in cells, new possibilities arise for a more efficient drug development and improved predictions of drug side effects. A concise modelling requires experimental evaluation of model structures and ideally of predictions provided by the model. Therefore our lab develops new methods to image and modulate intracellular events. In particular, we are interested in new methods to dissect signaling branches by specifically activating enzymes downstream of the membrane receptor level.

In this lecture, a variety of methods to detect signaling events in living cells as well as new methods to activate enzymes by small molecules will be presented. This includes synthetic membrane-permeant second messenger molecules such as phosphoinositides and protein constructs that are rapidly activated by chemical dimerizers such as rapamycin and its analogues. While these new tools help to reveal surprising new findings in receptor endocytosis, endosomal fusion, and calcium signal generation, the methods are also employed to validate a new model, called DynaCellNet, for G-protein-coupled receptor signaling featuring 45 elements and 290 molecular interactions. The model has predictive power and recently led to the discovery of a novel feedback loop in the regulation of intracellular calcium transients in our lab.

Notes

Wednesday, 17:00

## Search for Multifunctional Plasmepsin Inhibitors as Antimalarial Drugs

**Daniel Bur**, Christoph Binkert, Christoph Boss, Olivier Corminboeuf, Walter Fischli, Corinna Grisostomi, Andrew F. Jones, Solange Meyer, Lars Prade, Sylvia Richard-Bildstein and Thomas Weller.

Actelion Pharmaceuticals Ltd, Hegenhaimermattweg 91, CH-4123 Allschwil, Switzerland

The malaria parasite *Plasmodium falciparum* degrades host cell hemoglobin inside an acidic food vacuole during blood stage of the infectious cycle. A number of aspartic proteinases called plasmepsins play important roles in this degradation process and have therefore generated significant interest as new anti-malarial targets. Several X-ray structures of plasmepsin II (PMII) are available in the public domain but for a long time structure guided drug design was hampered by the fact that only inhibitors comprising a statine moiety or derivatives thereof were published. Our drug discovery efforts to find innovative, cheap and easily synthesized inhibitors against aspartic proteinases yielded some highly potent non-peptidic achiral inhibitors. Multiple well resolved X-ray structures of PMII will be presented featuring potent achiral inhibitors in an unprecedented orientation, contacting the catalytic aspartates in a new way. Major side chain rearrangements in the active site can influence the binding modes of the inhibitors. Moreover, in one case a second inhibitor molecule could be located unambiguously in the active site of PMII. Synthesis strategy as well as structural details will be discussed and comparisons with previously solved PMII-inhibitor complexes will be presented.

Notes

Wednesday, 17:45

## Molecular Mechanisms of Chromatin Targeting

**Christoph W. Müller**

European Molecular Biology Laboratory, Structural and Computational  
Biology Unit Meyerhofstrasse 1, D-69117 Heidelberg, Germany  
E-mail: *cmueller@embl.de*

The structural organization and accessibility of chromatin is in many cases controlled by histone-modifying enzymes and ATP-dependent chromatin remodeling factors. A key step in the recruitment of these enzymes to chromatin is the recognition of histone post-translational modifications (PTMs) by effector modules such as bromodomains, chromodomains, PHD fingers, MBT and SANT domains.

We will present recent work of our laboratory on the TAF1 homolog Brdt, a member of the BET protein family<sup>1</sup>. Brdt reorganizes acetylated chromatin by associating with hyperacetylated histone H4. Remarkably, a single bromodomain (BD1) of Brdt is responsible for selectively recognizing histone H4 tails bearing two or more acetylation marks. The crystal structure of BD1 bound to a diacetylated H4 tail shows how two acetyllysine residues cooperate to interact with one binding pocket and provides insight into the combinatorial readout of histones bearing multiple PTMs

Further examples concern the Polycomb group of proteins involved in the stable repression of different control genes during development. The Polycomb proteins *Drosophila* Sex comb on midleg (Scm)<sup>2</sup> and Sfm<sup>3</sup> contain two and four MBT repeats, respectively. Both proteins preferably bind mono- or dimethyl-lysine containing peptides, while unmethylated or trimethyl-lysine containing peptides are bound with significantly lower affinity. Crystallographic analysis of the MBT repeat domains of Scm and Sfm combined with isothermal calorimetry measurements provides insight into their ability to recognize methyl-lysine containing peptides. In the Pho repressing complex (PhoRC) the methyllysine-binding Sfm protein is associated with the DNA-binding factor Pho, thereby combining recognition of histone PTMs with DNA-sequence specific recognition to recruit Polycomb group proteins to their correct target sites.

### References:

<sup>1</sup> J. Morinière et al., Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* **461**, 664-668 (2009).

<sup>2</sup> C. Grimm et al., Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg. *EMBO Reports*. **8**, 1031-1037 (2007).

<sup>3</sup> C. Grimm et al., Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J*. **28**, 1965-1977 (2009).

Thursday, 9:00

## Targeting the CDK Family for Anti-Cancer Drug Design

**Jane A. Endicott**

Professor of Structural Biology, Department of Biochemistry,  
South Parks Road, Oxford OX1 3QU, U.K.  
E-mail: [jane.endicott@bioch.ox.ac.uk](mailto:jane.endicott@bioch.ox.ac.uk)

The cyclin-dependent kinases (CDKs) have been intensely studied because of their involvement in regulating essential cellular activities that include proliferation and transcription. The model of sequential CDK activation controlling cell cycle progression has been elaborated by studies of CDK and cyclin knock-out mice that have shown that although the CDK and cyclin families exhibit considerable functional redundancy, certain cell cycle CDKs have essential cell-type specific roles through their ability to phosphorylate specific substrates or to bind to specific cyclins. Recent structural studies of CDK1, 2, 4 and 9 complexes have also provided greater insights into substrate and inhibitor binding modes and uncovered a diversity of mechanisms of CDK activation and regulation that challenge the paradigm of CDK activation derived from structural studies of CDK2. Taken together, these recent developments may offer novel potential opportunities for the development of more effective CDK inhibitors.

Notes

Thursday, 9:45

## Targeting Signalling Networks in Cancer

**Stefan Knapp**

Phosphorylation Dependent Signalling Group, Structural Genomics  
Consortium, Oxford University, UK

In recent years our understanding of cellular signalling has changed from linear and directional pathways models to complex dynamic systems of protein interaction networks. This new view of cellular signalling has profound consequences on the selection of target molecules for drug discovery. The variable and diverse mutational landscape of each cancer subtype and considerable cellular heterogeneity in tumours makes rational target selection even more complex for this therapeutic area.

To address these issues we used a combination of unbiased kinome wide RNAi screens together with parallel inhibitor screening to interrogate the role of protein kinases in a specific cancer subtype. Surprisingly many so far unstudied kinases targets emerged as novel promising entry points of pharmacological intervention. Large scale structural comparison of generated co-crystal structures was subsequently used to obtain more specific inhibitors and new strategies for the structure based design of ATP competitive kinase inhibitors.

In addition, I will also briefly discuss our effort designing open access chemical probes for epigenetic targets.

Notes

Thursday, 11:00

## Organometallic Complexes as Emerging Scaffolds for the Design of Enzyme Inhibitors

Eric Meggers

Fachbereich Chemie, Philipps-Universität Marburg,  
Hans-Meerwein-Strasse, 35032 Marburg, Germany

Medicinal chemistry and chemical biology are dominated by organic chemistry, while metal-containing compounds play only a minor role. This is well illustrated by a review of drugs approved by the US Food and Drug Administration during 2007 in which not a single compound contains a metal atom, with most compounds being reversible enzyme inhibitors. However, our laboratory recently demonstrated that chemically inert metal complexes can serve as promising scaffolds for the design of enzyme inhibitors and we reported several compounds with high affinities and promising selectivity profiles for protein kinases and lipid kinases.<sup>1,2,3</sup> These organometallic compounds are formally derived from the class of ATP-competitive indolocarbazole alkaloids (e.g., staurosporine) in which the substitutionally inert metal center allows to access unconventional structures with defined and rigid shapes in an economical fashion. This presentation will illustrate how the structural features of metal complexes can be applied to the design of molecules with astonishing potencies and selectivities.

### References:

- <sup>1</sup> E. Meggers, G. E. Atilla-Gokcumen, H. Bregman, J. Maksimoska, S. P. Mulcahy, N. Pagano, D. S. Williams, *Synlett* 2007, **8**, 1177 (Account article).
- <sup>2</sup> E. Meggers, *Curr. Opin. Chem. Biol.* 2007, **11**, 287.
- <sup>3</sup> J. Maksimoska, L. Feng, K. Harms, C. Yi, J. Kissil, R. Marmorstein, E. Meggers, *J. Am. Chem. Soc.* 2008, **130**, 15764-15765.

Notes

Thursday, 11:45

## Drug Development: Puzzle or Mystery?

**Hans-Werner Mewes**

Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich  
National Research Center for Environment and Health, GmbH  
Chair of Genome Oriented Bioinformatics, Life Science Centre  
Weihenstephan  
Technische Universität München, Am Forum 1, 85354  
Freising-Weihenstephan, Germany

The process of drug development involves the entire spectrum of chemistry and biology. Success stories and disasters are part of its history. An indefinite chemical space is confronted to a widely unknown universe of cellular processes. The rationale selection of drug targets and candidate compounds is a complex process dominated by trial and error. Modern technologies generate massive amounts of data that need to be interpreted with respect to the effect of the drug investigated. Since a few years, a basic shift towards the analysis of biological systems using mathematical models, was coined as Systems Biology. Systems Biology uses models in a way that give the impression to be deterministic, in other words to predict “what if” relations between the action of the drug and its impact. The questions to be discussed in my presentation are:

- (i) are quantitative models suitable for the investigation of drug/target interactions?
- (ii) is the problem solvable in a rational way, as promised by systems biology?
- (iii) how can the available information from millions of publications be used and build into the drug development process?

Notes



### 3 Posters

## pKa Calculations on Staphylococcal Nuclease

**Paul Czodrowski**

Merck KGaA, Department of Bio- and Chemoinformatics  
Frankfurter Straße 250, D-64293 Darmstadt, Germany  
E-mail: [paul.czodrowski@merck.de](mailto:paul.czodrowski@merck.de)

The enzyme staphylococcal nuclease is an enzyme responsible for the cleavage of a phosphodiester bond within a polynucleotide chain. The titratable residues of the active site show highly shifted pKa values, e.g. Asp21 with a pKa value of 6.1<sup>1</sup>. Results of pKa calculations<sup>2</sup> with a structure-based interpretation will be presented.

### References:

- <sup>1</sup> C.A. Castaneda et al., *Proteins.*, **77**, 570 (2009)
- <sup>2</sup> P. Czodrowski et al., *Proteins.*, **65**, 424 (2006)

## Virtual Ligand Screening for the Purine Riboswitch

**Peter Daldrop and Ruth Brenk**

University of Dundee, Division of Biochemistry and Drug Discovery, College of Life Sciences, Dow Street, Dundee DD1 5EH, Scotland, UK  
E-mail: p.daldrop@dundee.ac.uk

Structure-based ligand screening has emerged as an important tool in rational drug design<sup>1</sup>, and in silico virtual ligand screening (VLS) is becoming increasingly popular owing to the enhanced speed and reduced cost. While VLS has been predominantly applied to protein targets, its application on RNA species remains relatively unexplored. With many RNA species validated as drug targets<sup>2</sup>, such as the ribosomal RNA, we are interested in developing a reliable method for RNA-ligand screening. To achieve this, we focus on a well-studied RNA species to validate and experiment our ligand screening protocol, both retro- and prospectively.

Riboswitches are cis-acting gene regulatory elements which constitute part of the 5' untranslated region (5'UTR) of many, predominantly bacterial, mRNAs<sup>3</sup>. Ligand-induced structural changes lead to the formation of two mutually exclusive three-dimensional structures, of which one permits gene expression while the other does not. The *Bacillus subtilis* xpt-pbuX G-riboswitch binds guanine as its natural ligand, whereas the C74U mutant has the ligand specificity changed to adenine. The availability of high-resolution crystal structures (1.7 Å) of this riboswitch<sup>4</sup>, together with binding data of a variety of small-molecule ligands in the literature, render it a promising template for the feasibility study of RNA-ligand screening.

Initially, published binding data was used to compile a test set of molecules, which were docked into the binding pocket of the riboswitch C74U mutant structure (PDB ID 2g9c) using DOCK 3.5.54. The algorithm was generally successful in distinguishing ligands from non-binders, and correctly predicted the co-crystallised binding conformations of published riboswitch-ligand complexes in the PDB. To further verify the method, a set of 2579 molecules was selected from an in-house database of commercially available compounds. This data set was then docked into the corresponding binding pocket using the same algorithm. The top-scoring 100 compounds were visually inspected, and four compounds selected for riboswitch binding assessment. These compounds were subjected to X-ray crystallisation trials with the riboswitch, and the binding constants determined. Binding was detected for all four compounds (90-690  $\mu\text{M}$ ), and three ligands of them successfully crystallised in complex with the riboswitch.

### References:

- <sup>1</sup> P. Kolb, R. S. Ferreira, J. J. Irwin, B. K. Shoichet, *Curr Opin Biotechnology*, **20**(4):429-36, 2009
- <sup>2</sup> J. Gallego, G. Varani, *Acc Chem Res*, **34**(10): 836-43, 2001
- <sup>3</sup> J.E. Barrick and R.R. Breaker, *Genome Biology*, **8**(11): R239, 2007
- <sup>4</sup> S.D. Gilbert, S.J. Mediatore and R.T. Batey, *J Am Chem Soc*, **128**(44): 14214-5, 2006

## Cell Line Based Compound Profiling and Response Prediction

**Hinrich W.H. Göhlmann<sup>2</sup>, Berthold Wroblowski<sup>2</sup>, Willem Talloen<sup>2</sup>, Dirk Wuyts<sup>2</sup>, Stijn Dickens<sup>2</sup>, Karine Smans<sup>1</sup>, Pieter Peeters<sup>2</sup>, Sepp Hochreiter<sup>3</sup>, Günter Klambauer<sup>3</sup>, Ziv Shkedy<sup>4</sup>**

1 Ortho Biotech Oncology Research and Development and

2 Johnson & Johnson Pharmaceutical Research and Development, divisions of Janssen Pharmaceutica, Turnhoutseweg 30, 2340 Beerse, Belgium,

3 Institute of Bioinformatics, Johannes Kepler University Linz, Altenberger Str. 69, A-4040 Linz, Austria,

4 Center for Statistics, Hasselt University, 3590 Diepenbeek, Belgium

E-mail: [hgoehlma@its.jnj.com](mailto:hgoehlma@its.jnj.com)

In the past decade drug discovery has been largely oriented towards target-directed drug discovery, but has not been as successful as expected. A target-based approach limits discovery to known molecular targets, and front-loads the time and expense of target-driven biological validation prior to the identification of a drugable compound. As an alternative, compounds screened in cellular assays, seeking for defined phenotypic changes, may have inherently better drug-like properties, such as solubility and cell permeability and should have the desired initial biological effect. After repetition of the screen, IC<sub>50</sub> determination and first-line toxicity testing the scientist is faced with the problem to select active compounds for further development.

We have developed an experimental approach to aid the selection with biological data. Using whole genome gene expression profiling of compounds from the phenotypic screen, we attempt to contribute information that will allow us to identify and distinguish different or common mechanisms leading to the phenotypic readout. First insights into potential side effects as well as data that can be utilized for deciding on a backup strategy are intended to be generated as well.

In a first pilot study a kinase directed 21K small molecule library has been screened using ATP-GLO to determine cell survival on colon-205 cells. The assay identified several hundred hits which have no or only limited biological target information. The hits can be clustered into distinct chemical classes and selections from each cluster were put into high throughput gene expression profiling.

## Towards Completing the Pharmacological Profile of Drugs

Elisabet Gregori-Puigjané\*, Jérôme Hert\*, Michael J. Keiser, Christian Laggner, John J. Irwin and Brian K. Shoichet

Department of Pharmaceutical Chemistry, University of California San Francisco  
1700 4th street, San Francisco, CA 94143-2550  
E-mail: [elisabet.gregori@gmail.com](mailto:elisabet.gregori@gmail.com)

Recent studies<sup>1,2</sup> have pointed out that a significant number of approved drugs have no reported primary target. While Drews et al. report a 7% of drugs with no reported primary target, Overington et al., through an exhaustive analysis of 1,357 unique drugs, concluded that 292 (21.5%) of them had no known primary target. In the MDL Drug Data Report (MDDR) database alone there are 75,623 compounds annotated with general pharmacological effects such as antineoplastic or antiarrhythmic, i.e., with no assigned protein target. In this work, we will examine both the approved and the experimental drugs sets in DrugBank<sup>3</sup> to prove that, by mining all the information contained in annotated chemical libraries, literature and patents, the actual percentage of drugs for which the primary target is not known can be reduced to 1%. We will also look into the compounds in MDDR with no target to try to predict the protein through which they act.

In the first part of this study, we took the approved set of the DrugBank database<sup>3</sup>, consisting of 1,382 drugs, and analyzed those compounds with no reported primary target. This led to a set of 107 compounds (8%), a figure in agreement with the existing literature<sup>1,2</sup>. From this set, 36 were discarded from the study for not being absorbed or not having a protein target (i.e. contrast media, sun blockers or agents that work by osmotic force in the digestive tract). For the remaining set of 71 compounds we could identify known targets for 57 of them in literature or one of the annotated chemical libraries (WORLD of Molecular BioAcTivity (WOMBAT)<sup>4</sup>, the StARliteTM or the MDDR database<sup>5</sup>). This left a set of only 14 drugs (1%) for which the primary target could not be found.

Finally, working with the much bigger set of MDDR containing 172,731 compounds, we tried to deorphanize those with annotations to indications but not to specific protein targets (75,623 compounds) using the Similarity Ensemble Approach (SEA)<sup>6</sup>, a method that scores single molecules against a panel of ligand-sets for with a specific macromolecular target based on chemical similarity. A few of these predictions were tested experimentally leading to the discovery that, for instance, Naftopidil is a Dopamine D2 antagonist (185 nM) and PRE-084 is a Muscarinic M1 and M2 antagonist (543 and 2681 nM, respectively).

### References:

- <sup>1</sup> J. P. Overington, B. Al-Lazikani, A. L. Hopkins, *Nat Rev Drug Discov* **5**, 993 (2006).
- <sup>2</sup> J. Drews, *Science* **287**, 1960 (2000).
- <sup>3</sup> D. S. Wishart et al., *Nucl Acids Res* **36**, D901 (2008).
- <sup>4</sup> M. Olah et al., in *Chemoinformatics in Drug Discovery*, Wiley-VCH, Ed. (New York, 2004), pp. 223-239.
- <sup>5</sup> I. S. L. MDL Information Systems, CA., (2006). <sup>6</sup> M. J. Keiser et al., *Nat Biotech* **25**, 197 (2007/02//print, 2007).

## Heme Oxygenase-1 and -2 Interactions and their Possible Implications in Drug Design

Stefano Forte<sup>1,5</sup>, Giancarlo Rappazzo<sup>2</sup>, Giovanni Li Volti<sup>3\*</sup> and Salvatore Guccione<sup>4,5</sup>

<sup>1</sup> IOM Ricerca, via Penninazzo 7, I-95029 Viagrande (Italy);

<sup>2</sup> Dipartimento di Biologia Animale, University of Catania, via Androne 81, I-95124 Catania (Italy);

<sup>3</sup> Dipartimento di Chimica Biologica, Chimica Medica e Biologia Molecolare and  
<sup>4</sup> Dipartimento di Scienze Farmaceutiche, University of Catania, v.le A Doria 6 Ed. 2  
Città Universitaria, I-95125 Catania (Italy);

<sup>5</sup> ETNALEAD s.r.l. via S. Nullo 5/i c/o Scuola Superiore, I-95123 Catania (Italy).

\*E-mail: livolti@unict.it

Protein–Protein interactions (PPIs) play a key role in most biological processes—from intracellular communication to programmed cell death—and therefore represent a large and important therapeutic targets. The wealth of information obtained from genome and proteome programs has increased the number of known protein–protein interactions involved in the pathogenesis of various human diseases. Nevertheless, these interactions were left aside for a long time, since development of drugs against the classical targets like enzymes and receptors has so far been more cost effective. Since the expenses for developing these classical drugs is on the increase, new target proteins and/or protein–protein interactions have come within reach. The research on protein–protein interactions might boost the development of technologies to identify and validate interactions relevant for human disease and the development of assays to screen for inhibitors of protein–protein interactions.

Taking into account the financial risk associated with this basic research, to date initial steps of target validation and hit identification might predominantly be taken by academia. However, further lead optimization and clinical development will have to be undertaken in collaboration with industry.

**Adiponectin** is a relatively abundant protein secreted from the circulating adipocytes in the micrograms per milliliter range. It acutely lowers blood glucose levels through suppression of hepatic glucose levels and chronically has potent insulin–sensitizing effects through reduction of hepatic lipids. Various diseases are associated with lower plasma adiponectin levels. Specifically, type 2 diabetic subjects and patients with early-stage cardiovascular disease have lower levels of adiponectin. Adiponectin circulates in three different forms: high molecular weight (HMW) (18–36 mer), low molecular weight (LMW) (hexamer), and a trimeric form. Levels of the HMW form are tightly connected to insulin sensitivity. The relevance of the hexameric and trimeric forms has not yet been systematically addressed. Stable higher-order complex formation requires the presence of disulfide bond formation at position 39 by a process which is highly regulated and critically dependent on a number of oxidoreductases that catalyze the disulfide bond exchange.

Recently, we showed that heme oxygenase-2 (HO-2)  $-/-$  (knockout) mice exhibited low levels of adiponectin thus resulting in insulin resistance. Interestingly, pharmacological induction of heme oxygenase-1 (HO-1) restored adiponectin levels in heme oxygenase-2  $-/-$  animals and rescued them from the diabetic condition. Therefore, we hypothesized that the heme oxygenase system can be involved in the secretory mechanisms of adiponectin. Based on our previously published results computational methods for protein interactions prediction were used to identify the possibility for both HO-1/adiponectin and HO-2/adiponectin complexes to form at the level of the monomeric adiponectin. These

oxygenases show, in fact, extended regions that are complementary in shape and functions to corresponding areas in adiponectin, which reflects in the strongly favourable free binding energy. The modelled complexes were used to select the interacting portions of these proteins. An additional filtering procedure was applied to select a shortlist of fragments which are likely to keep the original structural integrity. The corresponding coding regions might be cloned and assayed in a Two Hybrid Yeast System for a direct evaluation of their binding potential and the design of drugs.

## Target Fishing in the Sea of Literature

**Björn Grüning, Christian Senger, Stefan Günther**

Inst. Pharm. Science, Pharm. Bioinformatics, Albert-Ludwigs-University  
Hermann-Herder-Str. 9, D-79104 Freiburg, Germany  
E-mail: [stefan.guenther@pharmazie.uni-freiburg.de](mailto:stefan.guenther@pharmazie.uni-freiburg.de)

To understand the cellular effect of natural or new synthesized compounds it is important to identify potential target proteins, a problem that is also termed as Target Fishing. Typically, gathering information about the compound from literature, databases or pharmaceutical experts precedes experimental tests. Suggestions for molecular interactions are usually derived from compounds with similar structure. However, a vast amount of current knowledge about compound interacting partners is only described in literature and just available in non-standardized text format. Extracting relevant molecular relations based on compound similarities is difficult to perform on text information and the probability to miss important references that describe potential target proteins is enormous. To face this problem we have build up a web service that provides a very efficient method to screen literature for compounds and interacting proteins. Starting with the name or the structure of a molecule, about 16 million medical research articles (PubMed-entries) are screened and proteins that are mentioned in the same context are identified. Furthermore, structurally similar compounds stored in the PubChem database (26 million compounds) are subjected to the same screening procedure. A detailed output table displays the compound of interest and their relatives as well as the frequency of co-occurrence of certain proteins. As a result, it is possible to derive suggestions for new targets based on the included compound similarity screening. Links to the PubMed abstracts enable in-depth analyses of the underlying literature.

## Probabilistic Modeling of Conformational Space for 3D Kernels

**Andreas Jahn, Georg Hinselmann, Nikolas Fechner, Carsten Henneges and  
Andreas Zell**

Center for Bioinformatics (ZBIT), University of Tübingen  
Sand 1, D-72076 Tübingen, Germany  
E-mail: [andreas.jahn@uni-tuebingen.de](mailto:andreas.jahn@uni-tuebingen.de)

A fundamental idea of QSAR methodologies is that reasonable models can only be derived from descriptors containing useful information for a specific problem.<sup>1</sup> Transferred to 3D-QSAR approaches, it is essential for the quality of the model to have active conformations of the molecules. This information can be obtained from X-ray diffraction or NMR experiments, but it is rarely possible to compile a complete data set that contains the necessary information. Therefore, a common approach is to minimize the molecules and trying to compute a useful 3D-QSAR model based on the minimized structures. To reduce the dependency on active conformations, Hopfinger et al.<sup>2</sup> developed a 4D-QSAR formalism that incorporates information of a conformational sampling into a CoMFA like approach. In this work we introduce a 4D formalism that is based on a probabilistic encoding of distance profiles of intramolecular atom-pairs. The use of probability kernels allows the integration of this formalism into 3D kernels for machine learning approaches, like the Support Vector Machine, without violating the kernel properties.

Our algorithm operates on intramolecular atom-pairs and creates distance profiles using the information of the conformational sampling. To reduce the computation time, atom-pairs with a constant distance profile (e.g. neighbored atoms or atoms of the same ring system) were neglected. The next step approximates the distance profiles of the atom-pairs by Gaussian mixture models. For this purpose, an expectation-maximization algorithm determines the parameters of the Gaussian mixture models. We developed a heuristic method that determines the number of components of the Gaussian mixture model based on a modified all pairs shortest path algorithm. The Expected Likelihood Kernel calculates similarity values based on the Gaussian mixture models of the distance profiles of the atom-pairs. These similarity values describe the behavior of the corresponding atom-pair distances in the conformational space of the molecules. Finally, the results of the Expected Likelihood Kernel can be integrated into 3D kernels.

We integrated this formalism into a 3D kernel that is based on local atom pair environments.<sup>3</sup> Results on several data sets show that the formalism is capable to improve the model quality in comparison to the solely 3D based kernel on minimized structures. Additional experiments demonstrate that the probability of choosing useful structures of the conformation ensemble by chance, which yield a better 3D based model, is below 0.01. Another advantage of the formalism is that only the variables of the Gaussian mixture models of the atom-pair distance profiles have to be stored in contrast to the complete conformational ensemble of the molecules.

### References:

- <sup>1</sup> T. A. Martinek, F. Ötvös, M. Dervarics, G. Tóth, F. Fülöp, *J. Med. Chem.*, **48**, 3239-3250 (2005).
- <sup>2</sup> A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.*, **119**, 10509-10524 (1997).
- <sup>3</sup> A. Jahn, G. Hinselmann, N. Fechner, A. Zell, *J. Cheminf.*, **1**, (1):14 (2009).

## Hydropathic Interaction Networks

Glen E. Kellogg, Vishal N. Koparde, Philip D. Mosier, J. Neel Scarsdale

Department of Medicinal Chemistry, Center for the Study of Biological Complexity, and  
Institute for Structural Biology & Drug Discovery, Virginia Commonwealth University  
800 E. Leigh Street, Suite 212, Richmond, Virginia, 23219 USA.

E-mail: [glen.kellogg@vcu.edu](mailto:glen.kellogg@vcu.edu)

We have long been interested in the hydropathic interactions of proteins with ligands, DNA and other protein subunits. Our hydropathic interaction model (HINT)<sup>1,2</sup> has enabled us to develop a quantitative scoring function that has correlated well with free energy of binding in numerous studies. The HINT scoring function is empirically derived and based on measured small molecule LogP for 1-octanol and water, and thus quantitates hydrophobic interactions and desolvation energy. Of parallel interest to us has been quantifying the roles, conservation and energy contributions of water molecules at these interfaces,<sup>3,4</sup> with particular interest in protein active sites. Similarly, we have investigated in detail methods to optimize the ionization states of acidic and basic residues and ligand functional groups at these active sites, in order to accurately model and score the protein-ligand interactions. We are terming this latter analysis “computational titration”,<sup>5,6</sup> and have recently made available a public web server at <http://hinttools.isbdd.vcu.edu/CT>.<sup>7</sup>

Another key application of these technologies is in building molecular models for the target protein itself. Even with atomic resolution crystallographic structures, the placement of protons within their models is often ambiguous. We coined the term *isocrystallographic* to describe the often rather extensive set of structural models that can be fit within the experimental electron density envelopes – mainly differing in proton positions. When these protons are properly placed, through protonation/deprotonation of residue side chains and optimized rotation of water molecules, the intricate nature of the structure is revealed in terms of extensive polar hydrogen bonding networks. Also, hydrophobic interaction networks which stabilize proteins in a complementary and orthogonal way are revealed. We have programmed a novel genetic algorithm-driven extension to computational titration that optimizes ionization states, polar hydrogen (-XH<sub>n</sub>) rotations and water networking for entire proteins– yielding the isocrystallographic model with optimal hydropathic interaction networks for that protein. Most importantly, molecular models for target proteins that are properly configured with optimum or near optimum hydropathic interaction networks will be preferable for structure-based drug discovery applications. This also implies that ligands should bind *holistically* and add to, rather than disrupt, the networks. This poster will illustrate a few applications of this tool in optimizing protein structures and their active site environments.

### References:

- <sup>1</sup> G.E. Kellogg, D.J. Abraham, *Eur. J. Med. Chem.* 2000, **35**, 651 (2000)
- <sup>2</sup> P. Cozzini, M. Fornabaio, A. Marabotti, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, *J. Med. Chem.*, **45**, 2469 (2002).
- <sup>3</sup> A. Amadasi, J.A. Surface, F. Spyraakis, P. Cozzini, A. Mozzarelli, G.E. Kellogg, *J. Med. Chem.*, **51**, 1063 (2008).
- <sup>4</sup> A. Amadasi, F. Spyraakis, P. Cozzini, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, *J. Mol. Biol.*, **358**, 289 (2006).
- <sup>5</sup> M. Fornabaio, P. Cozzini, A. Mozzarelli, D.J. Abraham, G.E. Kellogg, *J. Med. Chem.*, **46**, 4487 (2003).
- <sup>6</sup> F. Spyraakis, M. Fornabaio, P. Cozzini, A. Mozzarelli, D.J. Abraham, G.E. Kellogg, *J. Am. Chem. Soc.*, **126**, 11764 (2004).
- <sup>7</sup> A.S. Bayden, M. Fornabaio, J.N. Scarsdale, G.E. Kellogg, *J. Comput.-Aided Mol. Design*, **23**, 621 (2009).

## Thioredoxin Reductase – A Promising New Drug Target in *M. Tuberculosis*

**Oliver Koch<sup>1,2</sup>, Leopold Flohé<sup>2</sup>, Timo Jaeger<sup>2</sup>, Paul M. Selzer<sup>1</sup>**

<sup>1</sup> Intervet Innovation GmbH, Zur Propstei, 55270 Schwabenheim

<sup>2</sup> MOLISA GmbH, Brenneckestraße 20, 39118 Magdeburg, Germany

E-mail: [Oliver.Koch@sp.intervet.com](mailto:Oliver.Koch@sp.intervet.com)

The resurgence of Tuberculosis, caused primarily by *M. tuberculosis*, and the appearance of multi-drug resistant (MDR) and extensively drug resistant (XDR) *M. tuberculosis* strains strengthen the need for new drugs with alternative modes of action.<sup>1</sup> The thioredoxin reductase (TrxR) / Thioredoxin (Trx) System is part of the *M. tuberculosis* antioxidant system that inactivates peroxides, contributes to ribonucleotide reduction, and thus guarantees the survival within macrophages.<sup>2</sup> *M. t.* lacks the common glutathione system and the *M.t.* TrxR shows a substantial difference in sequence, mechanism and structure to the human TrxRs. This makes the TrxR a promising new target of drugs for the treatment of tuberculosis.

Two important hydrogen bonding interactions were identified at the protein–protein interface of the TrxR-Trx complex. A high-throughput docking using constraints was applied to filter the Intervet in-house compound library (~ 6.5 million compounds) for possible hits that interact with these two hydrogen bonding acceptors. This reduced the number of interesting compounds to ~ 150000 that were redocked without any constraints and most accurate docking settings. For the final compound selection an automated ranking was applied based on a normalisation and consensus scoring strategy.

So far, 17 out of the first 170 tested compounds showed an activity with an IC<sub>50</sub> value upto  $\mu\text{M}$  range. Especially four different scaffolds with a low molecular weight are promising candidates for further developments. Slight modification of the best hit lead already to an improved activity with an IC<sub>50</sub> value of 1.6  $\mu\text{M}$ .

### References:

<sup>1</sup> World Health Organization Tuberculosis Programme; <http://who.int/tb/>

<sup>2</sup> T. Jaeger, L. Flohé, The thiol-based redox networks of pathogens: unexploited targets in the search for new drugs. *Biofactors*, **27**, 109-120 (2006)

## Ensemble Docking Revisited

Oliver Korb\*, Simon Bowden, Tjelvar Olsson, David Frenkel,  
John Liebeschuetz, Jason Cole

Cambridge Crystallographic Data Centre,  
12 Union Road, Cambridge CB2 1EZ, UK \*E-mail: [Oliver.Koch@sp.intervet.com](mailto:Oliver.Koch@sp.intervet.com)

In recent years, the importance of considering induced fit effects in molecular docking calculations has been widely recognised in the molecular modelling community. While small-scale protein side-chain movements are now accounted for in many state-of-the-art docking strategies, the explicit modelling of large-scale protein motions such as loop movements in kinase domains is still a challenging task. For this reason ensemble-based methods have been introduced taking into account several discrete protein conformations in the conformational sampling step. Our protein-ligand docking program GOLD<sup>1,2</sup> has been extended to search such conformational ensembles time-efficiently. The performance of the approach has been assessed on several protein targets using different scoring functions. A detailed analysis of pose prediction and virtual screening results and the dependence of these on the number of protein structures considered in the conformational ensemble; will be presented and limitations of the approach will be highlighted.

### References:

- <sup>1</sup> G. Jones, P. Willett and R. C. Glen, *J. Mol. Biol.*, **245**, 43 (1995).
- <sup>2</sup> G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *J. Mol. Biol.*, **267**, 727 (1997).

## Structure-Based Design of Histone Demethylase Inhibitors

**Carlino Luca<sup>1</sup>, Alexander Hauser<sup>2</sup>, Manfred Jung<sup>2</sup>, Wolfgang Sippl<sup>2</sup>**

<sup>1</sup> Institut für Pharmazeutische Chemie, Martin-Luther-Universität  
Halle-Wittenberg Wolfgang-Langenbeck-Str.4, 06120 Halle (Saale)

<sup>2</sup> Institut für Pharmazeutische Wissenschaften, Albert-Ludwigs-Universität  
Freiburg, Albertstr. 25, 79104 Freiburg.  
E-mail: carlino.luca@gmail.com

The genetic information in eukaryotic cells is organised in a specific structure called chromatin. The basic unit of chromatin is the nucleosome, which consists of four histone proteins and ~ 147 bp of DNA.<sup>1</sup> There are different types of histones known, and the N-terminal tails of these proteins contain important sites for post-translational modifications directly linked to gene expression. The modifications include phosphorylation, acetylation, ubiquitinylation, methylation sumoylation and ribosylation. Each modification is mediated by specific enzymes. LSD1 (Lysine Specific Demethylase 1) is one of the histone demethylases which removes one methyl group from methylated lysine residues. Lysine demethylation may have a dual role in genetic expression: gene repression (H3K4) and gene activation (H3K9).<sup>2</sup>

It has recently been demonstrated that the androgen receptor (AR)-LSD1 complex demethylates a repressive histone mark (H3K9) and then promote genes activation.<sup>3</sup> Experimental data show, also, that LSD1 is strongly expressed in prostate cancers with high a Gleason score.<sup>4</sup> For these reasons, specific modulation of LSD1 activity might be a promising therapeutic strategy in tissues where AR has a key physiological role.

LSD1 is a flavin-dependent amine oxidase which shares sequence identity with other flavin-dependent amine oxidases like monoamine oxidase (MAO), and polyamine oxidase (PAO). At the first stage of our work we analyzed differences and similarities of the active site among these enzymes. Next, several docking studies were evaluated using the available crystal structures of LSD1 and the related oxidases to discover a potential lead compound. For the evaluation studies we selected different ligand data sets, well known as inhibitors of MAO and PAO. The docking setup which showed the highest accuracy and enrichment factors are now selected for virtual screening of LSD1 inhibitors. Preliminary biological data are available and will be discussed in the context of the target structure.

### References:

<sup>1</sup> K. Luger, AW. Maeder, RK. Richmond, DF. Sargent, TJ. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, *Nature*, **389**, 251-260 (1997).

<sup>2</sup> J. Woyscka, A. T.M. Milne, C.D. Allis, Taking LSD1 to a New High, *Cell*, **122**, 654-658 (2005).

<sup>3</sup> E. Metzger, M. Wissman, N. Yin, J.M. Müller, R. Schneider, A.H.F.M. Peters, T. Günther, R. Buettner, R. Schüle, LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription, *Nature*, **437**, 436-439 (2005).

<sup>4</sup> P. Khal, L. Gullotti, L.C. Heukamp, S. Wolf, N. Friedrichs, R. Vorreuther, G. Solleder, P.J. Bastian, J. Ellinger, E. Metzger, R. Schüle, R. Buettner, Androgen Receptor Coactivators Lysine-Specific Histone Demethylase 1 and Four and Half LIM Domain Protein 2 Predict Risk of Prostate Cancer Recurrence, *Cancer Research*, **66**, 11341-11347 (2006).

## A Structure-Based Model for the Prediction of Protein Druggability

**Daniel Muthas, Agata Krasowski, Ruth Brenk**

University of Dundee  
College of Life Science, Dow Street, Dundee, DD1 5EH, Scotland, UK  
E-mail: [d.muthas@dundee.ac.uk](mailto:d.muthas@dundee.ac.uk)

Methods that can differentiate between promising and problematic targets are of significant importance in the target selection phase of a drug discovery programme. This has led to that the concept of a target's druggability has become a commonly used term. However, this metric is complex and encapsulates many different aspects of drug development, e.g. therapeutic relevance. Another key criterion for druggability is that the target protein should be possible to modulate with a small, preferably orally available, compound. Here we present a multivariate approach to predict if a binding site is probable to bind small, orally available molecules. To achieve this, a high-quality data set was compiled to facilitate model building and validation.

Our data set is an extension of the data set described by Cheng et al.<sup>1</sup>, and comprises of 162 structures of which 110 are considered druggable and 52 non-druggable. The druggability of each protein was manually assigned to ensure a consistent assessment. A protein was considered druggable if there is evidence of orally available drugs, or if there is literature support of the binding of Lipinski like<sup>2</sup> compounds to the protein. On the contrary, a protein was deemed non-druggable if all known modulators were prodrugs, required active transport or no reported Lipinski like binders were identified in the literature.

The binding site of each protein was characterized using a set of descriptors to capture the information about the size, polarity and structural composition of the binding site. The data set was then split into a training set of 113 structures and a validation set of 49 structures. A PCA of the data set indicated that the binding sites of the druggable proteins were more homogenous than those of the non-druggable proteins. In order to derive a druggability score, an OPLS-DA was performed (Figure 1). Our results demonstrated that it was possible to assign cut-offs that identified both druggable and non-druggable proteins with high precision.

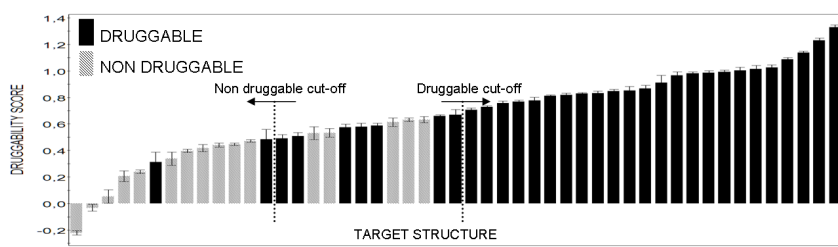


Figure 1: Druggability prediction of the 49-protein validation set.

In conclusion, a high-quality protein druggability benchmark set has been assembled and a protein druggability model has been derived that shows stable performance in discriminating between druggable and non-druggable protein targets.

**References:**

- <sup>1</sup> A. C. Cheng, R. G. Coleman, K. T. Smyth, Q. Cao, P. Souldard, et al. *Nat. Biotech.*, **25**, 71 (2007).
- <sup>2</sup> C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney. *Adv. Drug Deliv. Rev.*, **46**, 3 (2001).

*In Silico* Discovery of  
2-amino-4-(2,4-dihydroxyphenyl)thiazoles as Novel  
Inhibitors of DNA Gyrase B

Andrej Perdih<sup>1</sup>, Matjaž Brvar<sup>1</sup>, Marko Oblak<sup>2</sup>, Lucija Peterlin Mašič<sup>3</sup>, Tom Solmajer<sup>1</sup>

<sup>1</sup> National Institute of Chemistry, Hajdrihova 19, 1001 Ljubljana, Slovenia

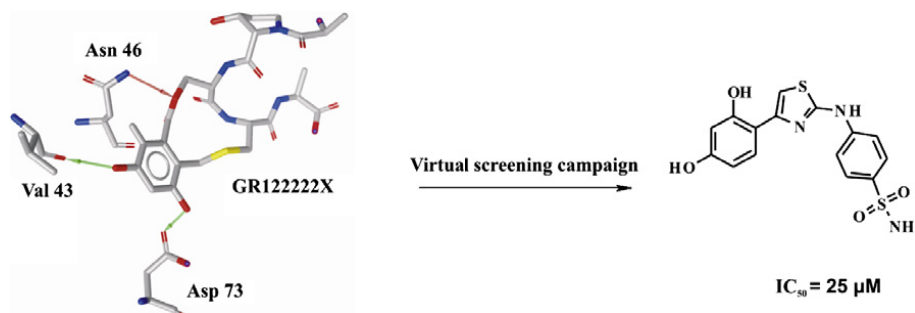
<sup>2</sup> Lek Pharmaceuticals, d.d., Verovškova 57, 1526 Ljubljana, Slovenia

<sup>3</sup> Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, 1001 Ljubljana, Slovenia

E-mail: [andrej.perdih@ki.si](mailto:andrej.perdih@ki.si)

The emergence of bacterial resistance to most of the clinically used antibiotics is driving an urgent need for the development of novel and effective antibacterial agents.<sup>1</sup> The main challenge remains the discovery of highly potent antibacterials with broad spectrum of efficacy and improved safety profile.<sup>2,3</sup>

Cyclothialidines<sup>4</sup> are a class of bacterial DNA gyrase B (GyrB) subunit inhibitors, targeting its ATP-binding site. Starting from the available structural information on cyclothialidine GR122222X (2),<sup>5</sup> an *in silico* virtual screening campaign was designed combining molecular docking calculations with three-dimensional structure-based pharmacophore information. A novel class of 2-amino-4-(2,4-dihydroxyphenyl)thiazole based inhibitors with low micromolar antigyrase activity was discovered.<sup>6</sup>



**References:**

- <sup>1</sup> E. D. Brown, G. D. Wright, *Chem. Rev.* **105**, 759 (2005).
- <sup>2</sup> L. L. Silver, *Biochem. Pharm.* **71**, 996 (2006).
- <sup>3</sup> A. Perdih et al., *Bioorg. Med. Chem. Lett.* **19**, 2668 (2009).
- <sup>4</sup> P. Anghern et al., *J. Med. Chem.* **47**, 1487 (2004).
- <sup>5</sup> R. J. Lewis et al., *EMBO J.* **15**, 1412, (1996).
- <sup>6</sup> M. Brvar, A. Perdih, et al., *Bioorg. Med. Chem. Lett.* **20**, 958 (2010).

## ParaDockS – A Framework for Molecular Docking

Martin Pippel<sup>1</sup>, Renè Meier<sup>2</sup>, Wolfgang Sippl<sup>1</sup>

<sup>1</sup> Institut für Pharmazeutische Chemie, Martin-Luther-Universität  
Halle-Wittenberg Wolfgang-Langenbeck-Str.4, 06120 Halle (Saale)

<sup>2</sup> RCPE GmbH, Inffeldgasse 21a/II, 8010 Graz, Austria E-mail:  
`martin.pippel@pharmazie.uni-halle.de`

The prediction of possible binding geometries as well as a ranking of putative protein-ligand complexes according to their binding affinities are the intention of so called molecular docking approaches. To evaluate these complexes against each other, scoring functions are required. In the recent years knowledge-based scoring functions have been evolved. They exploit the vast amount of experimentally determined structures to derive statistical atom pair potentials. In this work, the focus has been set on the implementation and the validation of a fast and robust knowledge-based objective function PMF04 into the molecular docking program PARADOCKS (Parallel Docking Suite)<sup>1</sup>. PARADOCKS is a flexible, easily extensible and open source docking program that was developed in our research group.

For the implementation of the PMF04 scoring function we extracted the atom pair potentials from the publicly available potential of mean force (PMF)<sup>2</sup>. To avoid unfavourable docking conformations an additional vdW-term is added.

To make it more easy for developers to incorporate their own or adapted objective functions a PARADOCKS Subgraph Search Description (PSSD) is implemented. It is a line notation to describe subgraph patterns similar to SMARTS. The primary difference is the lack of implicit hydrogen treatment and implicit bonds.

In the following, the performance of PARADOCKS for virtual screening was compared with the commercially available docking program GOLD. Therefore 13 targets were selected from the directory of useful decoys (DUD) which contains a set of unbiased actives and decoys. Especially in terms of early enrichment PARADOCKS outperforms the three available scoring functions implemented in GOLD.

### References:

<sup>1</sup> <http://paradocks.org/>

<sup>2</sup> I. Muegge, *J. Med. Chem.*, **49**, 5895 (2006).

## Impact of the Number of Inactive Samples on a Large Scale Learning Approach

L. Rosenbaum, A. Jahn, G. Hinselmann, N. Fechner and A. Zell

Center for Bioinformatics, University of Tübingen

Sand 1, D-72076 Tübingen

E-mail: [lars.rosenbaum@uni-tuebingen.de](mailto:lars.rosenbaum@uni-tuebingen.de)

Large scale approaches play an increasingly role in virtual screening because of improved high-throughput screening technology. The resulting data sets comprise up to a several hundred thousand chemical compounds and are highly unbalanced. Because of memory and computation time constraints, state-of-the-art nonlinear kernel machines are not suited for these large scale data sets. A convenient way to circumvent this problem is to reduce the number of inactive samples. Another common approach in virtual screening is similarity-fusion methods. Most of these discard all inactive samples<sup>1</sup>. It is, however, desirable to use as much information as possible for constructing a model.

We investigate the impact of the number of inactive samples on the performance of the linear support vector machine LIBLINEAR<sup>2</sup>. Linear support vector machines can train models extremely fast if the input comprises sparse input vectors. This is achieved by a Molprint-like fingerprint algorithm<sup>3</sup>. We trained models on six publicly available data sets with a size of 175,000 compounds<sup>1</sup>. The classifiers were trained with a decreasing number of inactive samples. The performance of the classifiers was measured by the area under the receiver operator characteristics (AUC) and the early recognition metric BEDROC<sup>4</sup>. The experiments show, that while the number of inactive samples have no impact on the AUC, the early recognition performance lowers significantly.

In real world screening applications only the top-ranked structures are further analyzed, thus the early recognition performance is often more important than the overall performance of a classifier. Thus, the results highlight the gain of information by using the inactive samples for building models for virtual screening. This is consistent with results from a comparison of similarity-fusion approaches, where a method, that also considered the inactive samples performed significantly best<sup>1</sup>. We extended the LIBLINEAR Java library<sup>5</sup> with state-of-the-art high-throughput screening metrics and threaded cross-validation. The software is freely available from the authors.

### References:

<sup>1</sup> R.J. Nasr, S.J. Swamidass, P.F. Baldi, *Cheminformatics*, **1**, 7 (2009).

<sup>2</sup> R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, *J. Mach. Learn. Res.*, **9**, 1871 (2008).

<sup>3</sup> A. Bender, H.Y. Mussa, R.C. Glen, *J. Chem. Inf. Comput. Sci.*, **44**, 170 (2009).

<sup>4</sup> J.-F. Truchon, C.I. Bayly, *J. Chem. Inf. Model.*, **47**, 488 (2007)

<sup>5</sup> <http://www.bwaldvogel.de/liblinear-java/> (2010)

## Billions of Virtual Small Molecules for Drug Discovery

**Lars Ruddigkeit, Ruud van Deursen, Lorenz Blum and Jean-Louis Reymond\***

Department of Chemistry and Biochemistry, University of Berne,  
Freiestrasse 3, 3012 Berne, Switzerland  
\*E-mail: [jean-louis.reymond@ioc.unibe.ch](mailto:jean-louis.reymond@ioc.unibe.ch)

As a rather small number of drugs have entered the market in the last 20 years, the need for new chemical entities is apparent.<sup>1</sup> In silico methods like molecular scaffolds analysis, breeding of molecules by genetic algorithms and exhaustive enumeration of chemical space<sup>2,3</sup> can assist the search for novel molecules. Recently our group published GDB-13.<sup>4</sup> With nearly a billion molecules it currently is the largest publicly available small organic molecule database.

Here we report the extension of the approach up to 15 atoms. The previously used algorithm for GDB-13 was nearly at the technical limit of computation. For the enumeration of larger molecules the algorithm had to be modified. We introduced an entirely new and faster algorithm, including substantial extensions within JChem.<sup>5</sup> Furthermore a set of restricted filters for topology and functional groups was defined, based on the statistical occurrence of these structural elements in ZINC.<sup>6</sup> The total number of molecules was thus reduced substantially while focussing on a more relevant chemical space.

We will show the results of the enumeration of chemical space up to 15 heavy atoms. This results into billions of molecules. We then analyse the molecular diversity and distribution of descriptor values. These results show which part of these molecules fulfil drug-like and fragment-like criteria.

GDB-15 contains a vast quantity of structures not present in known databases such as ACX, Zinc or PubChem. As a result of this it can certainly contribute to the search for new bioactive molecules in chemical biology and medicinal chemistry.<sup>7</sup>

### References:

- <sup>1</sup> I. Kola, J. Landis, *Nat. Rev. Drug. Discovery.*, **8**, 711 (2004).
- <sup>2</sup> T. Fink, H. Bruggesser, J.-L. Reymond, *Angew. Chem. Int. Ed.*, **44**, 1504 (2005).
- <sup>3</sup> T. Fink, J.-L. Reymond, *J. Chem. Inf. Model.*, **47**, 342 (2007).
- <sup>4</sup> L. C. Blum, J.-L. Reymond, *J. Am. Chem. Soc.*, **131**, 8732 (2009).
- <sup>5</sup> Chemaxon Ltd. <http://www.chemaxon.com/>
- <sup>6</sup> J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.*, **45**, 177 (2005).
- <sup>7</sup> K. T. Nguyen et al., *ChemMedChem.*, **3**, 1520 (2008).

## Open Source Large Scale Structure Based Prediction of Druggability of Protein Cavities

Peter Schmidtke, Xavier Barril

Departament de Físicoquímica, Facultat de Farmàcia, Universitat de Barcelona  
Av. Joan XXIII, s/n; 08028 Barcelona, Spain  
E-mail: [pschmidtke@mmb.pcb.ub.es](mailto:pschmidtke@mmb.pcb.ub.es)

During the last decade, the concept of druggability of a given target has seen the day. Within this context, several groups proposed computational methods for predicting druggability of binding sites.<sup>1,2,3</sup> Here a new approach is presented towards druggability prediction on small molecule binding pockets that can be applied in an automated way for analysis of large structural databases.

Fpocket, an open source cavity detection software suite<sup>4</sup>, was used to retrieve cavities and extract descriptors. A logistic model was trained and validated on a larger dataset than used by precedent contributions. State of the art prediction performance was obtained on discriminating druggable cavities from decoys and known non druggable cavities.

Cavity detection and drug score calculations are very fast and can be used to analyse large structural databases as well as molecular dynamics.

As open source / open data project the drug score was implemented into the fpocket cavity detection software and will be freely available for download under <http://www.sourceforge.net/projects/fpocket>.

Due to the ambiguity in defining a druggable binding site, a web based collaborative data creation platform was developed in order to stimulate the community to contribute together to efficient data set maintenance for training and validation of druggability prediction methods. This tool will be available at : <http://fpocket.sourceforge.net/dcd>.

### References:

- <sup>1</sup> P.J. Hajduk, J.R. Huth, S.W. Fesik, *J. Med. Chem.*, **48**, 2518 (2005).
- <sup>2</sup> A.C. Cheng, R.G. Coleman, *Nat. Biotech.*, **25**, 71 (2007).
- <sup>3</sup> T. Halgren, *JCIM.*, **49**, 377 (2009).
- <sup>4</sup> V. Le Guilloux, P. Schmidtke, P. Tufféry, *BMC Bioinformatics.*, **10**, 168 (2009).

## Elucidating the Molecular Mechanism of Activation in PKC-theta

Jesús Seco

Physico-chemistry Department  
Av. Joan XXIII, s/n, 08028 Barcelona  
E-mail: [jseco@mmb.pcb.ub.es](mailto:jseco@mmb.pcb.ub.es)

The AGC family of protein kinases constitutes one of the widest kinase families, representing around 12% of the human kinome. Many of its members play critical roles in cellular activation and differentiation, and are involved in numerous pathophysiological processes. This family shares a special structural feature not present in any other kinase family, a C terminal tail that wraps around the whole of the kinase domain and bears one or two functionally important extra phosphorylation sites crucial for the correct functioning of most AGC kinases. Here, we take PKC-theta as a prototypical member of the family, using a series of molecular dynamics simulations to study different activation states. We first show how the apo-unphosphorylated kinase is populating a conformational space in which basic structural elements hinder both ATP binding and activation loop phosphorylation, both of which are necessary for catalytic competence. Our simulations then show how this structure can be activated by two alternative sequences of events. If phosphorylation on the activation loop takes place first, a transition is seen where the kinase opens the ATP binding site to more readily accept substrate binding. But the molecular mechanism of signal transmission works both ways, and substrate binding to the unphosphorylated form de-protects the activation loop, exposing it for phosphorylation. This is consistent with recent evidence of paradoxical activation of PKCs by inhibitors. Furthermore, the simulations also show how the additional phosphorylation of the C-tail results in a more active form both in the apo state, where opening of the binding site is more pronounced, and in the ATP-bound state, where the C-tail induces correct arrangement of the structural elements necessary for catalysis, giving a hint on the functions this C tail might have in regulation. Taken together, the results presented here suggest that alternative sequences of events can lead to competent enzyme-cofactor complex formation, and demonstrate the profound impact phosphorylation exerts on the structure and flexibility of the kinase fold.

## Docking and Binary Classification Model of Isothiazolones as Irreversible Inhibitors of Histone Acetyltransferase PCAF

**Suhaib Shekfeh<sup>1</sup>, Silviya D. Furdas<sup>2</sup>, Manfred Jung<sup>2</sup>, Wolfgang Sippl<sup>1</sup>**

<sup>1</sup> Department of Pharmaceutical Chemistry, Martin-Luther-Universität, Halle/Saale

<sup>2</sup> Institute of Pharmaceutical Sciences, Albert-Ludwigs Universität, Freiburg.

E-mail: [suhaib.shekfeh@student.uni-halle.de](mailto:suhaib.shekfeh@student.uni-halle.de)

Aryl and alkyl N-substituted isothiazolones have been shown to inhibit histone H3 and H4 acetylation by PCAF and p300. Aryl and alkyl N-substituted isothiazolones are known PCAF inhibitors and the inhibition is due to an irreversible interaction with thiol groups. PCAF inhibition of isothiazolones is abolished in the presence of thiol-reducing agents like dithiothreitol (DTT) or glutathione. Furthermore, the activity was not restored in experiments involving the incubation of PCAF with two isothiazolones followed by dialysis. According to the supposed mechanism of disulfide bridge formation, the isothiazolones could form a covalent bond with the side chain of Cys574 located in the active site of PCAF. Using the available PCAF crystal structure and different side-chain conformations for Cys574 we carried out covalent and non-covalent docking studies using the GOLD software. Based on the docking results novel pyridoisothiazolone and benzoisothiazolone compounds were synthesized and tested in vitro and in vivo. Novel compounds active in the low micromolar range could be obtained. We analyzed whether the biological activity of the compounds is also dependent on the reactivity of the electrophilic system (isothiazolones). This reactivity is determined by the difference between the HOMO energy of the nucleophile Cys574 and the LUMO energy of the electrophile. We calculated a series of QM descriptors like HOMO and LUMO energy and related descriptors to classify the isothiazolones into actives ( $IC_{50} = 3.40 - 25 \mu M$ ) and into inactives ( $IC_{50} > 50 \mu M$ ). Using the QuaSar Module in MOE.2008.10, a binary classification model has been established which is able to classify the isothiazolones depending on the LUMO energy of the isothiazolone ring system and the HOMO-LUMO gap of the system.

## Chemogenomic Profiling in Chemical Oncology Research

Haridas Rode, Martin Sos, Roman Thomas and Daniel Rauh

Chemical Genomics Centre of the Max Planck Society  
Otto-Hahn-Strasse 15, D-44227 Dortmund  
E-mail: [daniel.rauh@cgc.mpg.de](mailto:daniel.rauh@cgc.mpg.de)

Reversible inhibitors of the epidermal growth factor receptor kinase (EGFR) are the first class of small molecules to improve progression-free survival of patients with EGFR-mutated lung cancers.

Irreversible EGFR inhibitors introduced as second generation drugs to overcome acquired resistance by the T790M<sup>1</sup> resistance mutation of EGFR have so far demonstrated limited clinical activity in patients with T790M-mutant tumors. In this study, we systematically analyzed the determinants of the activity and selectivity of the second generation EGFR inhibitors.

Focused libraries of irreversible as well as structurally corresponding reversible EGFR-inhibitors of the quinazolines and quinoline scaffold were synthesized for chemogenomic profiling involving over 98 genetically defined non-small cell lung cancer cell lines.<sup>2</sup> Overall, our results show that the growth inhibitory potency of all irreversible inhibitors against the EGFR-T790M resistance mutation was limited by reduced target inhibition, linked to decreased binding velocity to the mutant kinase. Combined treatment of T790M-mutant tumor cells with BIBW-2992 and the phosphoinositide 3-kinase /mTOR inhibitor PI-103 led to synergistic induction of apoptosis. These findings offer a mechanistic explanation for the limited efficacy of irreversible EGFR inhibitors in EGFR-T790M gatekeeper-mutant tumors, and they prompt combination treatment strategies involving inhibitors that target signaling downstream of the EGFR.

### References:

- <sup>1</sup> A. Michalczyk, S. Klüter, H.B. Rode, J.R. Simard, C. Grütter, M. Rabiller, D. Rauh, *Bioorg Med Chem*, **16**, (2008).
- <sup>2</sup> M. Sos, H. Rode, S. Heynck, M. Peifer, F. Fischer, S. Klüter, V. Pawar, C. Reuter, J. Heuckmann, J. Weiss, L. Ruddigkeit, M. Rabiller, M. Koker, J. Simard, M. Getlik, R. Thomas, D. Rauh, *Cancer Res*, **70**, (2010).

## Identification and Characterization of Stabilizer of Inactive Kinase Conformations with FLiK

Jeffrey Simard, Christian Grütter, Matthäus Getlik and Daniel Rauh

Chemical Genomics Centre of the Max Planck Society

Otto-Hahn-Strasse 15, D-44227 Dortmund

E-mail: [daniel.rauh@cgc.mpg.de](mailto:daniel.rauh@cgc.mpg.de)

Kinase dysregulation disrupts the intricate network of intracellular signaling pathways and contributes to the onset of diseases such as cancer. Although several kinase inhibitors are on the market, inhibitor selectivity and drug resistance mutations persist as fundamental challenges in the development of effective long-term treatments.

Chemical entities binding to less conserved allosteric sites would be expected to offer new opportunities for scaffold development. Because no high-throughput method was previously available, we developed a fluorescence-based kinase binding assay for identifying and characterizing ligands which stabilize inactive kinase conformations.<sup>1,2</sup>

Here, we present a description of the development and validation of this assay using the serine/threonine kinase p38 $\alpha$  and the tyrosine kinase cSrc. By covalently attaching fluorophores to the activation loop of the kinases, we were able to detect conformational changes and measure the  $K_d$ ,  $k_{on}$ , and  $k_{off}$  associated with the binding and dissociation of ligands to the allosteric pocket. Additionally, we used protein X-ray crystallography together with our assay to examine the binding and dissociation kinetics to characterize identified hits and to guide further compound development.<sup>3,4</sup>

### References:

<sup>1</sup> J. R. Simard, S. Klüter, C. Grütter, M. Rabiller, H.B. Rode, D. Rauh, *Nat Chem Biol*, **5**, 394, (2009)

<sup>2</sup> J. R. Simard, M. Getlik, C. Grütter, V. Pawar, S. Wulfert, M. Rabiller, D. Rauh, *J Am Chem Soc*, **131**, 13286, (2009)

<sup>3</sup> J. R. Simard, C. Grütter, V. Pawar, B. Aust, A. Wolf, M. Rabiller, S. Wulfert, S. Klüter, C. Ottmann, D. Rauh, *J Am Chem Soc*, **131**, 18478, (2010)

<sup>4</sup> M. Getlik, C. Grütter, J. Simard, S. Klüter, M. Rabiller, H.B. Rode, A. Robubi, D. Rauh, *J Med Chem*, **52**, 3915, (2009)

## Open Access HTS Platform Providing WDI Derived Screening Collections

Edgar Specker<sup>1</sup>, Martin Neuenschwander<sup>1</sup>, Michael Lisurek<sup>1</sup>, Bernd Rupp<sup>1</sup>,  
Ronald Kühne<sup>1</sup>, Jens Peter von Kries<sup>1</sup> & Ronald Frank<sup>2</sup>

<sup>1</sup> Leibniz Institute for Molecular Pharmacology, Robert-Rössle-Str. 10, 13125 Berlin

<sup>2</sup> Helmholtz Centre for Infection Research, Inhoffenstr.7 , 38124 Braunschweig

E-mail: specker@fmp-berlin.de; www.chembionet.info

ChemBioNet was initiated by biologists and chemists from academia to support systematic usage of small molecules to explore biological systems. Chemists and biologists benefit by receiving bioactivity profiles for their unique molecules and by open access to high throughput technologies enabling for usage of compounds as unique tools for dosage dependent controlled interference with biological functions. The mission of ChemBioNet is to provide a “real” link between biologists, chemists and other scientific disciplines and sufficient resources for Chemical Biology. For this purpose ChemBioNet has defined professional standards for compound library management and formats for efficient exchange and documentation of chemical and screening data in the shared database maintained at the FMP. Four partner institutes (Helmholtz Centre for Infection Research, the Max-Delbrück-Center for Molecular Medicine, the University of Oslo and the Leibniz Institute for Molecular Pharmacology, FMP) have co-financed a shared central compound collection. The screening collection was enriched with bioactive and chemical diverse drug-like compounds according to the following strategy: The WDI (World Drug Index) was searched for bioactive compounds with a maximum common substructure and the most diverse compounds were extracted from each set<sup>1,2</sup>. Compound selection was scrutinized in accordance with the “Lipinski rules”<sup>3</sup>. Reactive and unstable compounds were removed by a self-compiled reactivity filter from a vendor library of 12 million compounds. Hitherto, about 40000 compounds have been deposited in the ChemBioNet screening library from national and international groups. The open access screening platform of the ChemBioNet supports HTS with all kind of reader-based technologies, “Alphascreen” technology<sup>4</sup>, on chip capillary electrophoresis<sup>5</sup>, impedance measurements<sup>6</sup> and high content screen<sup>7</sup> conducted in 384 well plate format. ChemBioNet will focus on establishing and maintaining professional infrastructures necessary to support academic screening projects in a European context. This repository of actually more than 40.000 compounds will be continuously complemented by compound collections donated by chemists or partnering screening platforms of the network.

### References:

<sup>1</sup> M. Lisurek, B. Rupp, J. Wichard, M. Neuenschwander, J. P. von Kries, R. Frank, J. Rademann, R. Kühne, *Molecular Diversity* 2009, **13**, doi:10.1007/s11030-009-9187-z.

<sup>2</sup> WDI, Derwent World Drug Index, Release 2005, Derwent Information Ltd., London.

<sup>3</sup> C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* 1997, **23**, 3-25.

<sup>4</sup> R. Bossé, C. Illy, J. Elands, D. Chelsky, *Drug Discovery Today* HTS supplement 2000, 42-47.

<sup>5</sup> J. Dunne, H. Reardon, V. Trinh, E. Li, J. Farinas, *Assay und Drug Development Technologies* 2004, **2**, 121-129.

<sup>6</sup> S. L. Kirstein, J. M. Atienza, B. Xi, J. Zhu, N. Yu, X. Wang, X. Xu, Y. A. Abassi, *Assay Drug Dev. Technol.* 2006, **4**, 545-553.

<sup>7</sup> K. Hellmuth, S. Grosskopf, C. T. Lum, M. Würtele, N. Röder, J. P. von Kries, M. Rosario, J. Rademann, W. Birchmeier, *Proc. Nat. Acad. Sci. USA* 2008, **105**, 7275-7280.

## Ligand Binding to Human DPP III and its Mutant H568N - Computational Analyses

Antonija Tomić, Jasminka Špoljarić, Marija Abramić, Sanja Tomić\*

Ruđer Bošković Institute  
Bijenička 54, HR-10002 Zagreb, Croatia  
\*E-mail: [sanja.tomic@irb.hr](mailto:sanja.tomic@irb.hr)

Human dipeptidyl-peptidase III (DPP III) hydrolyses distinctive synthetic substrate Arg-Arg-2-naphthylamide and a number of biologically active peptides (*e.g.* endomorphins) by cleaving dipeptides from their N-termini. Like the other members of recently recognized metallopeptidase family M49<sup>1</sup>, it has *HEXXXH* structural motif important for zinc ligation and catalytic activity. Besides its contribution in normal intracellular protein catabolism, there are evidences for its regulatory and pathophysiological role as well<sup>2</sup>. Human DPP III was recognized by the group in Zagreb as a biochemical indicator of endometrial and ovarian malignancies<sup>3</sup>. Others indicated the role for DPP III in the endogenous pain-modulatory system, defense against oxidative stress and in cataractogenesis<sup>4,5,6</sup>.

We used the crystal structure of ligand-free human DPP III (PDB ID:3FVY) and by combining docking with molecular dynamics (MD) simulations we determined binding modes for two synthetic ligands, inhibitor Tyr-Phe-hydroxamate (YFNHOH) and substrate Arg-Arg-2-naphthylamide (RRNA). In order to study changes in the protein active site and protein conformation induced by ligands binding, we performed series of 10 ns long MD simulations for the wild type (WT) enzyme, its H568N mutant and their complexes.

In the crystal structure of DPP III the central zinc ion is coordinated by two histidines (His-450, His-455), one glutamate (Glu-508) and one water molecule. MD simulations revealed the octahedral coordination of Zn<sup>2+</sup>, for example in the complexes with ligands (RRNA and YFNHOH) it is additionally coordinated by Glu-451 and by carbonyl oxygen belonging to the second peptide bond from the substrate N terminus. RRNA binds similarly to H565N as well. However, inhibitor in the complex with H568N moves away from Zn<sup>2+</sup> and after about 8 ns of MD simulations it stabilizes in the new position closer to the 'lower' domain. According to Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) calculations, and in agreement with the experimental data, the binding affinity of YFNHOH is significantly lower for the mutant than for the WT DPP III.

Based on the results of the simulations we proposed two possible reaction mechanisms for RRNA hydrolysis. In one Glu-508 behaves as general base and His-450 as the proton carrier, and in the other Glu-451 plays dual role of both the water activator and the proton carrier.

### References:

- <sup>1</sup> M. Abramić. et al., *Period. Biol.* **106**, 161 (2004).
- <sup>2</sup> J.-M. Chen, et al., (Eds.), *Handbook of Proteolytic Enzymes.*, vol. 1, *Elsevier*, Amsterdam, 809 (2004).
- <sup>3</sup> Š. Šimaga, et al., *Gynecol. Oncol.* **91**, 194 (2003).
- <sup>4</sup> T. Chiba, et al., *Peptides*, **24**, 773 (2003).
- <sup>5</sup> Y. Liu, et al., *Proc. Natl. Acad. Sci. USA*, **104**, 5205 (2007).
- <sup>6</sup> H. Zhang, et. al, *Histochem. J.* **33**, 511 (2001)

## Towards a Reproducible Generation of Test Sets for the Evaluation of *in silico* Screening Methods by Immediate Neighborhood Classification

Simon Vogel, Matthias Bauer, Frank Boeckler

Department of Pharmacy, Center for Drug Research,  
Ludwig-Maximilians University, Butenandtstr. 5-13, 81377 Munich, Germany  
Medicinal Chemistry & Drug Design, Eberhard-Karls University Tuebingen,  
Auf der Morgenstelle 8, 72076 Tuebingen, Germany  
E-mail: [frank.boeckler@uni-tuebingen.de](mailto:frank.boeckler@uni-tuebingen.de)

Starting from a database of compounds, *in silico* screenings are designed to stepwise reduce the quantity of compounds, ideally starting with computationally less demanding filter steps such as physicochemical properties. Subsequently, the remaining structures are further narrowed down by more sophisticated 1D, 2D, or 3D filters such as fingerprint, substructure or pharmacophore matching, and docking, respectively.

In case biologically active compounds are known beforehand, a reasonable way to validate an *in silico* screening approach is to design a suitable test set. This is done by mixing active compounds with an excessive number of inactive compounds (decoys) to assess the quality of the particular 1D, 2D or 3D filter step in terms of differentiation between active and decoy compounds.

In this work we present a fully automated protocol aiming to generate challenging and suitable test sets for the comparison of docking algorithms and scoring functions. To be useful a properly selected test set should meet the following criteria:

- i Actives and decoys should not be easily separable by low dimensional filter
- ii False negatives (actives hidden in the decoy set) should be avoided whenever possible
- iii To keep the separation challenging, the structural diversity of decoys should be maximized
- iv The decoy selection should be tailor-made for each single active to ensure a good and equal representation of the physicochemical property space of each active by the same number of decoys

This protocol will help to standardize and facilitate the decoy set selection of new benchmarking sets for docking.



## 4 List of Participants