First Annual

In Silico Drug Discovery Conference

December 3–4, 2014 | Durham, NC

HOSTED BY

North Carolina Biotechnology Center

Welcome to the first annual In Silico Drug Discovery Conference!

We are excited by your attendance and contribution at this first annual in silico conference.

It is widely recognized that traditional drug discovery and development are spiraling out of control with increased costs and lower success rates. It is also commonly believed that this effort will only be rescued by the increased and more effective application of IT to the drug discovery and development process. If the vision of precision medicine and eventually personalized medicine is to happen, it is important that new breakthroughs and better applications emerge for computational drug discovery.

In the past, in silico drug discovery had a bad reputation for its failure rate. However, this reputation was due primarily to inaccurate and low computational power algorithms (such as docking, low fidelity Monte Carlo models, literature mining and machine learning applied to limited data sets, etc). This approach has changed with the computational power of the cloud and more detailed biophysics models of molecular energy binding and behavior.

This conference promises to present cutting-edge methods that have been proven in practice. We have leading practitioners as keynotes and speakers. We have leaders in attendance and as sponsors. A growing, enthusiastic community of professionals are developing real drugs with in silico methods and will share what works and what doesn't to pave the way for even greater success.

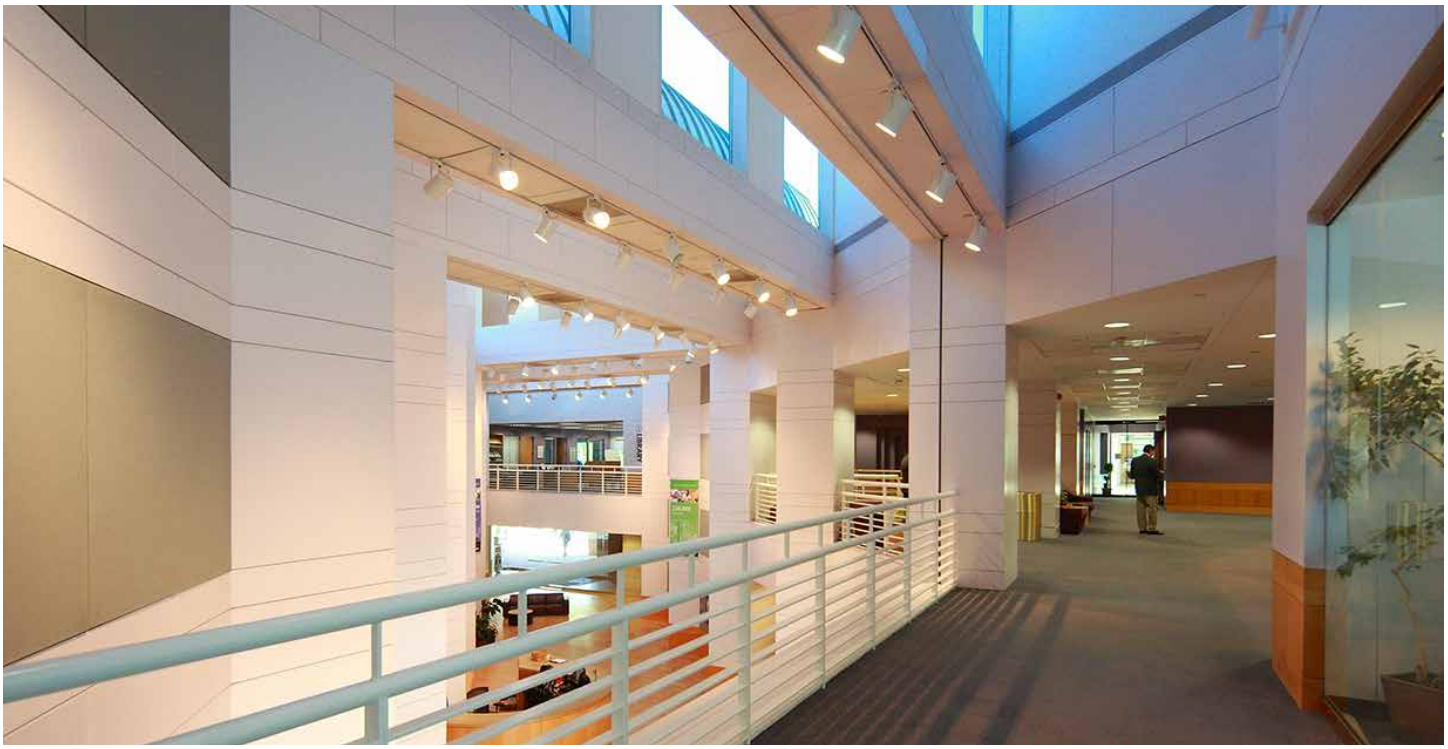
Thank you for attending!

Lawrence Husick
Founder and Chairman of the Board of Directors
Quantum Cures Foundation

Table of Contents

Agenda: Wednesday, December 3

8:45 a.m.–10:15 a.m.	Workshop: BioIT Solutions: 1Platform4 Discovery		Board of Directors Room
10:30 a.m.–noon	Workshop: ChemAxon: Chemistry Text Mining Suite		Board of Directors Room
1:30 p.m.–2:00 p.m.	Registration for In Silico Drug Discovery Conference		Glaxo Galleria
2:00 p.m.– 2:15 p.m.	Opening Remarks Doug Edgeton, President and CEO, North Carolina Biotechnology Center		Auditorium (location of all presentations)
2:15 p.m.–2:40 p.m.	Session 1: Intellectual Property *Chair: Dr. Bill Glauser, Schrödinger*	Medicinal Chemistry Due Diligence: Computational Predictions of an Expert's Evaluation of the NIH Chemical Probes Dr. Sean Ekins, Collaborations in Chemistry	
2:40 p.m.–3:05 p.m.		ChEMBL, SureChEMBL and UniChEM: Open Data for Drug Discovery Dr. John P. Overington, EMBL-EBI, Wellcome Trust Genome Campus	
3:05 p.m.–3:30 p.m.		*ChemCurator,* Computer-Assisted Patent Curation and Analysis Tool Dr. David Deng, ChemAxon LLC	
3:30 p.m.–4:15 p.m.		**KEYNOTE** From Computational Prediction Paper to Amusing, Wide Ranging, Slightly Unorthodox and Provocative Thoughts Dr. Chris Lipinski, Scientific Advisor, Melior Discovery	
4:15 p.m.–4:30 p.m.	Coffee Break		
4:30 p.m.–5:30 p.m.	Protecting the Products and Processes of In Silico Molecular Design: Patents, Copyrights, Open Source Licensing and the Public Domain Panel Discussion		
5:30 p.m.–7:30 p.m.	Networking and Poster Session		Ground Floor Atrium



Agenda: Thursday, December 4

7:50 a.m.–8:20 a.m.	Breakfast	
8:20 a.m. –8:30 a.m.		Opening Remarks Lawrence Husick, Founder and Chair, Quantum Cures Foundation
8:30 a.m.–8:55 a.m.	Session 2: Methods in Computational Drug Discovery *Chair: Dr. David Minh, Illinois Institute of Technology*	HALO/Gen: A Rapid Method for the Design and Development of Peptide Ligands as Target-Specific Bio-Markers and Drugs Dr. Harold Garner, Lynntech + Virginia Tech
8:55 a.m.–9:20 a.m.		ADME/Tox Predictions: Using Open-Source Software and Public Data to Build and Deploy Predictive Models for Activity Against Cytochrome P450 Enzymes Dr. Paul J Kowalczyk, Syngenta Biotechnology Inc.
9:20 a.m.–9:45 a.m.		Target-Specific Native/Decoy Pose Classifier to Boost Ligands' Ranking Accuracy for Virtual Screening Applications Dr. Regina Politi, UNC Eshelman School of Pharmacy
9:45 a.m.–10:30 a.m.		**KEYNOTE** Drug Discovery with Three Dimensional Models of Everything Dr. Ruben Abagyan, Professor, UC San Diego, Skaggs School of Pharmacy
10:30 a.m.–11:00 a.m.	Coffee Break	
11:00 a.m.–11:25 a.m.	Session 3: Applications of Computational Drug Discovery *Chair: Dr. Elizabeth Frush, Cloud Pharmaceuticals*	QM/MM and Inverse Design for Novel Therapeutics Targeting Drug-Resistant pfDHFR-TS Malaria Dr. Shahar Keinan, Cloud Pharmaceuticals Inc.
11:25 a.m.–11:50 a.m.		Ligand Engineering of PCSK9/LDL-R Protein-Protein Interaction Inhibitors Dr. John L. Kulp, BioLeap
11:50 a.m.–12:15 p.m.		Identification of Small Molecule Inhibitors of Type III Secretion System ATPase EscN from Enteropathogenic *E. coli* Dr. Wieslaw Swietnicki, Wroclaw Research Centre
12:15 p.m.–1:00 p.m.		**KEYNOTE** How to Increase the Impact of Structural Chemistry and Predictive Science for Drug Discovery Dr. Frank Brown, Associate VP and Head, Global Structural Chemistry, Merck Research Laboratories
1:00 p.m.–2:00 p.m.	Lunch	
2:00 p.m.–2:25 p.m.	Session 4: Optimizing Drug Discovery *Chair: Dr. Alex Tropsha, UNC Eshelman School of Pharmacy*	Reaction Driven Molecular Invention – Generating Synthetically Feasible Design Ideas Dr. Lei Wang, Certara
2:25 p.m.–2:50 p.m.		Multi-Dimensional Activity Cliff Analysis Dr. Robert Scoffin, Cresset Group
2:50 p.m.–3:15 p.m.		Presentation Dr. Aaron Virshup, Autodesk
3:15 p.m.–4:00 p.m.		**KEYNOTE** The Evolving Role of Modeling and Informatics in Drug Discovery Dr. Pat Walters, Principal Research Fellow, Vertex Pharmaceuticals
4:00 p.m.–4:20 p.m.	Coffee Break	
4:20 p.m.–5:20 p.m.	Market Trends for In Silico Drug Discovery Panel Discussion	
5:20 p.m.–5:30 p.m.		Concluding Remarks Rachelle J. Bienstock, PhD, Environmental Protection Agency

1Platform4 Discovery

Learn how 1Platform4 Discovery from BioIT Solutions supports sophisticated laboratory processes such as supernatant expression and antibody purification. The system combines supply-chain activities, sampling, bar coding, and a Biorepository in one intuitive, easy-to-use interface. Its highly adaptive nature makes it easy to pivot when new needs arise.

See how a distributed laboratory process can be supported using a single-interface. Each role has a customized portal, tuned for its needs. Since the system is available on the internet, this approach can be extended to suppliers, partners, and customers. Observe how this federated model provides secure, distributed, access which in-turn enhances collaboration and customer satisfaction.

Chemistry Text Mining Suite

A huge chemical space exists in chemical patents. However, unlocking this information can be labor-intensive, time-consuming, and in some cases, costly. ChemAxon has long been developing informatics tools for IP and life science researchers, including a Chemistry Text Mining Suite to address this issue in an automatic and cost-effective manner. This was developed based on ChemAxon's Naming and Markush Technology.

In this workshop, an overview of fundamental technology will be introduced, followed by demonstrations of recent developments in computer-assisted patent mining, particularly in extracting chemical information from patent documents.

Drug Discovery with Three Dimensional Models of Everything

Ruben Abagyan, PhD

UC San Diego, Skaggs School of Pharmacy. Email: ruben@ucsd.edu

The number of protein structures in the Protein Data Bank exceeds 100,000. Even the most recalcitrant to crystallization membrane proteins, G-protein coupled receptors (GPCRs) and channels, are no longer out of reach. These structures can be converted to specific three dimensional models, or fields, that help facilitate with the following major tasks:

- Building models by homology for the proteins not yet crystallized, subtypes, hetero-oligomers, and defining details not visible in crystal structures;
- In silico docking and screening for new modulators of a specific protein;
- Screening of a single compound against a panel of models to identify poly-pharmacology, adverse effects, or, in some cases identify a target of a phenotypic screen.

The types of three dimensional models and the methods involved continue to evolve, improve and organized into knowledge bases and automated workflows. Finally, we show how computational methods help to design crystallizeable constructs for membrane proteins and their complexes.

How to Increase the Impact of Structural Chemistry and Predictive Science for Drug Discovery

Frank Brown, PhD

Merck Research Laboratories, 770 Sumneytown Pike, West Point, PA 19486
frank.brown@merck.com

The way we have conducted structural chemistry over the last 30 years has been quite rewarding and there have been many advances and drugs that have come from those efforts. However, we cannot rest on these gains. It is time to think as boldly about structural chemistry and drug discovery as we did 30 years ago. We must find ways to have greater impact on the industry. The impact will come from better accuracy, which is hard to achieve and slow to realize. It will come from better utility, by infusing the methods into more scientific disciplines – which is scientifically easy but have large organizational barriers. It will also be realized by creating better networks for sharing data, methods, and models. We need to bring together science, organizational issues, and the use of cloud technologies to raise the tide of structural chemistry.

From a Computational Prediction Paper to Amusing, Wide Ranging, Slightly Unorthodox and Provocative Thoughts

Christopher A. Lipinski, PhD

Melior Discovery, 860 Springdale Drive, Exton, PA 19341 Email:chris44@aol.com

"Medicinal Chemistry Due Diligence: Computational Predictions of an Expert's Evaluation of the NIH Chemical Probes", a paper presented elsewhere at this meeting is the starting point of my story. Colleagues at Collaborative Drug Discovery were eager to see if they could mimic my medicinal chemistry evaluation of the NIH's 308 Molecular Library Probes. As background, I had been part of a crowd-sourcing evaluation of the probes in 1969 when there were just 64 probes. In initially compiling the data which would eventually reside in machine readable form I discovered a hidden NIH spreadsheet with about two thirds of some of the data I needed. In getting the remaining data I discovered the near non-existent communication between the public world of PubChem and the proprietary world of ACS's CAS SciFinder©. I rethought my ideas on what a drug discovery team should do when exploring freedom to operate on a new lead. I found a data dump of 5000 compounds into a US patent application from an HTS assay on the NIH's Molecular Library compounds. I am fairly sure this led to unexpected intellectual property results for a lot of academic labs. Working on our initial computational paper I had the opportunity to put some of my favorite ideas into a second commentary-like paper with idea input from new collaborators. Ideas (especially mine) can be dangerous. What really constitutes prior art? Can disclosing too much data be a problem? Why do medicinal chemists tend to repeat the same motifs in their syntheses? What really is medicinal chemistry due diligence and how does it differ from what a biologist would do? How does medicinal chemistry due diligence tie in with target and ligand network maps and with evolution?

The Evolving Role of Modeling and Informatics in Drug Discovery

Pat Walters, PhD

Vertex Pharmaceuticals. Email:Pat_walters@vrtx.com

Over the last few years, we have seen a dramatic increase in the amount of data generated as part of drug discovery programs. Increasing regulatory pressures and an awareness of the importance of drug-like properties have led to an effort to more rigorously characterize drug candidates. Discovery compounds are now routinely subjected to a battery of properties assays as well as in-vitro and in-vivo ADME evaluation. In addition to dealing with an increase in the number of assays, teams are also incorporating calculated properties, ADME models, and multiple ligand efficiency metrics as part of the optimization process. This data explosion is further compounded by large amounts of data coming from public sources through databases like ChEMBL, PubChem, RCSB, and DrugBank. As part of this new paradigm, the role of the modeler in drug discovery is changing. Modelers must be able to integrate this plethora of information and enable discovery teams to make effective decisions. This presentation will focus on some of the challenges facing the modeling community and provide a few suggestions for future directions.

ChemCurator, Computer-Assisted Patent Curation and Analysis Tool

David Deng[1], Árpád Figyelmesi[2] and Daniel Bonniot[3]

[1] *ChemAxon LLC, One Broadway, Cambridge, MA, 02142, USA, ddeng@chemaxon.com*
[2,3] *ChemAxon Kft., Záhony u. 7, Building HX, 1031 Budapest, Hungary*

Because of the significant investment associated with the development of a new drug, chemical patents are essential to pharmaceutical industry for intellectual property protection. For the same reason, patent analysis is an important but very labor intensive task. In addition to numerous examples and assay data scattered across the pages, complex Markush structures are often used to make the patent claim as broad as possible. Extracting useful information through hundreds of pages can be time consuming even for experienced information scientists.

ChemAxon has long been developing informatics tools for IP and life science researchers. *Document to Structure* can automatically extract chemical names from all types of documents; while *Markush Technology* can create, enumerate and search complex Markush structures. Now, by combining these two technologies, we present a new product, *ChemCurator*, to help expedite patent curation.

ChemCurator can open a patent document, and extract structures, including exemplified structures, R-group fragments, etc., from texts and images in a semi-automated way. The structures are displayed next to their original location in the document, which makes navigation easier and more intuitive. Users can drag-and-drop to populate and merge R-group fragments together, and quickly re-assemble the Markush structure according to the claim. Exemplified structure can also be searched against the Markush structure to see if they are within the chemical space or not. All these features will be introduced in this presentation.

We believe *ChemCurator* can be a very useful tool to help scientists analyze patents. Although it cannot completely eliminate human intervention, it can greatly reduce the processing time while confirming that the output Markush structure represents the chemical contents of the patent. A few new features are also in our develop plan, including automatic Markush structure creation from a structure library; overlap analysis of Markush databases; and assay data extraction.

Medicinal Chemistry Due Diligence: Computational Predictions of an Expert's Evaluation of the NIH Chemical Probes

Nadia Litterman[1], Christopher A. Lipinski[2], Barry A. Bunin[1] and Sean Ekins[1,3]

[1] *Collaborative Drug Discovery, Inc., 1633 Bayshore Hwy, Suite 342, Burlingame, CA, U.S.A*
[2] *Christopher A. Lipinski, Ph.D., LLC. 10 Connshire Drive, Waterford, CT 06385-4122, USA.*
[2]*Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, U.S.A.*
ekinssean@yahoo.com

In the past decade the National Institutes of Health (NIH) has funded extensive high throughput screening (HTS) efforts in both intra-mural and academic centers to identify small molecule chemical probes or tool compounds via the Molecular Libraries Screening Center Network (MLSCN) and the Molecular Library Probe Production Center Network (MLPCN). By 2009 it was estimated to have cost $385 million and by 2010 $576.6 million then funding was dramatically scaled back. Various definitions for compounds to become probes have been reported. To date the NIH-funded academic screening centers have discovered just over 300 chemical probes and some of the groups have demonstrated sophisticated drug discovery capabilities. We have collected the evaluations of a very experienced medicinal chemist on the likely chemistry quality of these probes based on a number of criteria including literature related to the probe and potential chemical reactivity. The presence of and pattern of literature references for each probe is a key part of the quality assessment. We were particularly interested whether the human evaluation aspect of medicinal chemistry due diligence could be computationally predicted. Analysis of the molecular properties of these compounds scored as desirable suggested higher pKa (for acidic compounds), molecular weight, heavy atom count and rotatable bond number. We also used a process of sequential Bayesian model building and testing as we included additional probes. Starting from 57 compounds, our final models included 307 compounds, which were classified as possessing or lacking undesirable features. Comparing different machine learning methods identified Bayesian models with optimal 5-fold cross validation (Receiver Operator Curve, ROC = 0.73). With an external test set of 15 further probes the model had an accuracy of 67% and ROC of 0.78. Our results suggest that computational machine learning methods can learn from the decisions of a medicinal chemist with a good degree of accuracy and can be used in addition to measures of drug-likeness and other filtering rules created to date. A comparison versus other molecule quality metrics or filters such as QED, BadApple and ligand efficiency indicates that a Bayesian model based on a single medicinal chemist's decisions for a small set of probes can make decisions that are preferable in classifying desirable compounds. We have made this data publically accessible in the CDD Vault and also recently implemented the model using CDD Models, a new approach for secure selective model sharing in the CDD Vault. CDD Models uses open source FCFP_6 fingerprints and a naïve Bayesian algorithm. Any medicinal chemist could potentially score the set of NIH probes based on their own criteria and build similar models so it would add efficiency to the workflow of scoring libraries computationally before screening. This work illustrates that the medicinal chemists insights can be captured by machine learning.

HALO/Gen: A Rapid Method for the Design and Development of Peptide Ligands as Target-Specific Bio-Markers and Drugs

Harold R. Garner[1], John E. Mueller[2], Christi L. Parham[2], Victor Palmer[2], Tony Ragucci[2], G. Duncan Hitchens[2], Sriram Shankar[2], Enusha Karunasena[1], Lovelace Soirez[2], Lauren J. McIver[1], Jasminkumar Bavarva[1]

[1] *Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 2406, skipgarner@gmail.com*
[2] *Lynntech, Inc., 2501 Earl Rudder Frwy S., College Station, TX 77845*

Computational approaches have been developed to design small molecules as binding agents/drugs. Another class of drugs is biologicals, including many that are peptide based. We have developed a computational approach to design and optimize peptide ligands as biomarkers and drugs. We have used this computational system in conjunction with experimental techniques to iteratively optimize and validate the peptide designs; a system we call HALO/Gen (High Affinity Ligand Optimization/Generation). Peptide-based drugs have certain advantages over other biologicals, specifically they offer advantages over antibody-based reagents which require select storage conditions and limited shelf-life.

The HALO/Gen system for peptide ligand design is based on the principal that 'like binds like'. Using a combination of algorithms designed to identify protein ligands similar to a query sequence, a proprietary database is mined for candidate ligands. These families of peptides are permutated at select, site-specific regions to create a library of tens of thousands of candidate daughter peptide sequences. Using peptide array technology, 30K-300K candidate peptide designs are screened for binding affinities. These data are used to further down-select candidate daughter peptides for more computational iterations and biological assays that are important to selection optimization and analysis. Ultimately, the resulting target-specific peptide ligands can be used for diagnostics and therapeutics.

QM/MM and Inverse Design for novel therapeutics targeting drug-resistant pfDHFR-TS Malaria

Shahar Keinan and Elizabeth Frush
Cloud Pharmaceuticals, Inc., 2 Davis Drive, Suite 219, Durham, NC 27709
Email: skeinan@cloudpharmaceuticals.com

Dihydrofolate reductase, or DHFR, is an enzyme that converts dihydrofolate into tetrahydrofolate, a methyl group shuttle required for the de novo synthesis of purines, thymidylic acid, and certain amino acids. Unlike the mammalian form, the parasitic form of DHFR is found as a bifunctional enzyme linked to Thymidylate Synthase (DHFR-TS). Due to the importance of DHFR in DNA precursor synthesis, as well as its' structural diversity from the human form, this has led to developing DHFR inhibitors as anti-parasitic drugs, such as Trimethoprim. However, resistance to trimethoprim and other drugs aimed at *pf*DHFR-TS can arise due to a variety of mechanisms (such as site-mutations and gene amplification), limiting the success of their therapeutic uses. There is a clear need for developing therapeutics that will efficiently target the wild-type as well as the mutated protein and will result in long-term treatment of patients.

We have studied the structure of 3 different *pf*DHFR-TS mutants (wild type, double mutant C59R/S108N and quadruple mutant N51I/C59R/S108N/I164L)[1]. A set of inhibitors that bind to the 3 known *pf*DHFR-TS mutants have been reported in the literature [2]. We use parallel QM/MM calculations with a GBSA algorithm to train and then predict the strength of ligand binding for the 3 protein structures. The Inverse Design algorithm searches a large virtual chemical space and uses the binding strength predictions to find inhibitors that will bind strongly to all 3 mutants.

[1] PDB names 1J3I, 1J3J and 1J3K; Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M., Yuthavong, Y., *Nature Structural & Molecular Biology*, **2003**, 10, 357-365
[2] Sirichaiwat, C., Intaraudom, C., Kamchonwongpaisan, S., Vanichtanankul, J., Thebtaranonth, Y., Yuthavong, Y., *J. Med. Chem.*, **2004**, 47, 345-354; Kamchonwongpaisan, S., Quarrell, R., Charoensetakul, N., Ponsinet, R., Vilaivan, T., Vanichtanankul, J., Tarnchompoo, B., Sirawarapor, W., Lowe, G., Yuthavong, Y., *J. Med. Chem.*, **2004**, 47, 673-680

ADME/Tox Predictions: Using Open-Source Software and Public Data to Build and Deploy Predictive Models for Activity Against Cytochrome P450 Enzymes

Paul J Kowalczyk, PhD[1]

[1] *Syngenta Biotechnology Inc*
3054 Cornwallis Road
Research Triangle Park, NC 27709
paul.kowalczyk@syngenta.com

We present our efforts at building machine learning models meant to predict activity against cytochrome P450 1A2, 2C19, 2C9, 2D6 and 3A4. These models have been constructed using open-source software and public data. The steps associated with the construction of machine learning models, *i.e.,* data retrieval, data curation, data exploration, and model construction, evaluation, interpretation and deployment have been executed using tools available in the R software environment for statistical computing and graphics[1]. We present results for eight machine learning methods, focusing on classification: recursive partitioning, random forests, neural networks, partial least squares, support vector machines, k-nearest neighbors, self-organizing maps and naïve bayes. Various tuning parameters for each method are analyzed, *e.g.* the dimensions and topology of a self-organizing map; the depth of trees for partitioning methods; the number of nodes / hidden nodes in a neural network; the choice of kernel. The choice of molecular descriptors has also been studied. Descriptors studied include topological descriptors (*e.g.,* atom pairs), circular fingerprints (*e.g.,* ECFP / FCFP), and constitutive fingerprints (*e.g.,* MDL keys). Metrics for model performance include AUC, sensitivity, specificity and Cohen's kappa. We show how these models might be deployed using RStudio's Shiny, a web application framework for R [2]. These predictive models become interactive web applications. We show how one might (1) download a compound dataset, (2) select one or more machine learning models, and (3) generate predictions for the compounds in the supplied dataset. All the R scripts developed for constructing these machine learning models will be made available.

[1] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

[2] Shiny, Easy web applications in R (2013). RStudio, Inc, URL: http://www.rstudio.com/shiny

Ligand Engineering of PCSK9/LDL-R Protein-Protein Interaction Inhibitors

John L. Kulp, Jr., Ian S. Cloudsdale, John L. Kulp, III, and Frank Guarnieri

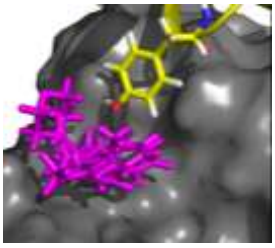
John L. Kulp, Jr., BioLeap, Inc., The Biotech Center, 3805 Old Easton Road, Doylestown, PA 18902, United States jlkjr@bioleap.com

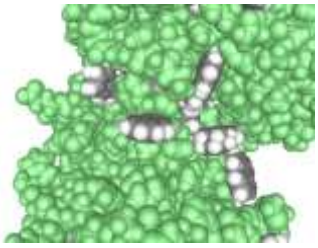
A novel ligand engineering process is described with application to the development of small molecule inhibitors of the PCSK9/LDL-R protein-protein interaction (ppi). Antibody inhibitors of PCSK9 have demonstrated dramatic clinical success in lowering cholesterol in treatment of hypercholesterolemia. However the development of orally-available small molecule inhibitors of PCSK9, as with inhibitors for other ppi targets, has generally not been successful using conventional drug discovery methodologies. This motivates the need for new approaches.

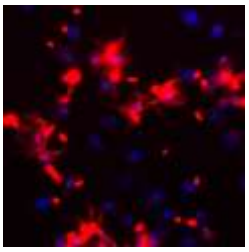
Our ligand engineering process involves three stages:

(1) **Formulate** a detailed hypothesis of the mechanisms required for inhibition, using information such as X-ray structures, mutation data, and hot-spot analysis[1];
(2) **Compose** custom compounds that meet these demanding requirements, using fragment-based design[2] and the evaluation of designed ligands with MD and QM simulations;
(3) **Validate** the inhibition hypothesis with experimental protocols that carefully minimize false negatives, including biophysical (SPR, ITC) assays for binding affinity, cell assays for biological function, and X-ray crystallography to validate the inhibition mechanism.

The computational tasks are highly automated, with easy-to-use tools employed by medicinal chemists in designing compounds that meet inhibition and pharmacological requirements. Example data from the PCSK9 program are shown below.


Hot-spot analysis


Fragment Maps


LDL Uptake Assay

The application of this process to the de novo design, implementation, and testing of four distinct chemical series for inhibiting PCSK9 will be presented. Compounds achieved micromolar affinities in surface plasmon resonance direct binding assays and LDL uptake in HepG2 cells. An initial assessment of the application to other ppi targets, such as PD1 immunotherapies in cancer, will be presented.

[1] J. Am. Chem. Soc. **2011**, 133, 10740.

[2] J. Comput.-Aided Mol. Des. **2012**, *DOI: 10.*1007-10822.

ChEMBL, SureChEMBL and UniChEM – Open Data for Drug Discovery

John P. Overington[1]

[1] *EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambs, CB10 1SD, UK. jpo@ebi.ac.uk*

The link between biological and chemical worlds is of critical importance in many fields, not least that of healthcare and chemical safety assessment. A major focus in the integrative understanding of biology and medicine in particular are genes/gene proteins and the networks and pathways describing their interactions and functions; similarly, within chemistry there is much interest in efficiently identifying drug-like, cell-penetrant compounds that specifically interact with and modulate these targets. The number of genes of interest is of the range of 10^5 to 10^6, which is modest with respect to plausible drug-like chemical space – 10^{20} to 10^{60}.

We have built a public database (ChEMBL http://www.ebi.ac.uk/chembl) linking chemical structures (~10^6) to molecular targets (~10^4) and phenotypic assays. The data covers molecular interactions and pharmacological activities and Absorption, Distribution, Metabolism and Excretion (ADME) properties. ChEMBL can be viewed as a map of the general features of molecular properties and features important for both small molecule and protein targets in drug discovery. We have then used this empirical data kernel to extend analysis across the human genome, and to large virtual databases of compound structures (for example GDB-13) – we have also integrated these data with genomics datasets, such as the GWAS catalogue.

The patent literature can be viewed as a noisier and less curated 'early view' of the published, peer-reviewed literature that ChEMBL is based on, and we have recently added SureChEMBL (http://www.surechembl.org) to our EMBL-EBI resources. This contains an automated pipeline for chemical structure extraction and annotation of the patent literature, and currently contains in the order of 10^7 molecules.

We have complemented ChEMBL and SureChEMBL with a large-scale InChI-based resolver – UniChem (http://www.ebi.ac.uk/unichem) which contains approaching 10^8 structures. UniChem offers a very flexible way of providing rapid cross-references to external data resources and has become our major route to chemical structure integration. Further annotation of UniChem content sources according to availability of compounds, likely synthesizable and virtual allows novel workflows in knowledge-based compound design and selection.

Applications of ChEMBL, SureChEMBL and UniChem to key data challenges in *in silico* design will be discussed, alongside our technical strategy and resource development roadmap.

Target-Specific Native/Decoy Pose Classifier to Boost Ligands' Ranking Accuracy for Virtual Screening Applications.

Regina Politi[1,2], Denis Fourches[1] and Alexander Tropsha[1]

[1] *Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman school of pharmacy, University of North Carolina, Chapel Hill, NC, 27599, United States; reginap@email.unc.edu*
[2] *Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC, 27599, United States;*

We have implemented a hybrid docking and scoring workflow that combines traditional docking with pose filtering in an effort *(i)* to improve the identification of native-like poses for any ligand and *(ii)* to boost the accuracy of ligand ranking for virtual screening. As a unique component of our approach, a special pose classifier has been developed on the basis of binary classification models trained to discriminate native-like from decoy poses using a single known x-ray structure for a protein-ligand complex. To build the classifier, a single cognate ligand with known native pose is docked multiple times into the respective protein and the generated poses are divided into two classes (native-like and non-native) using an RMSD threshold of 2Å. Different poses are characterized by MCT-Tess descriptors of the protein-ligand interface developed in our group previously [1] and the random forest approach is used to discriminate the two pose classes based on pose descriptors. Schrodinger's Glide docking software is used to generate poses for each ligand and rank them according to both SP and XP scoring functions. The pose classifier is then applied to the Glide-generated ligand poses to filter out those predicted as decoys following by re-ranking of the remaining poses. When applied to a congeneric series of ten steroid ligands as part of the CSAR2013 benchmark, the ranking accuracy for these ligands evaluated by the Spearman correlation coefficient was 0.64 for SP and 0.52 for XP, but reached 0.75 for SP/MCT-Tess consensus scoring [2]. As part of the CSAR2014-phase 1 benchmark, our hybrid method correctly identified for three different targets (SYK, FXA, and TRMD) the *one and only* pose (out of 200 pre-generated decoys) that was within 1Å RMSD of the actual native pose. In Phase 2, the native-like pose was predicted for several congeneric molecules for all three targets. Overall, these studies reconfirm that target-specific pose scoring models are capable of enhancing the reliability of structure-based molecular docking by discarding irrelevant ligand poses prior to applying the conventional scoring approaches.

[1] Hsieh, J.-H.; Yin, S.; Wang, X. S.; Liu, S.; Dokholyan, N. V.; Tropsha, A. Cheminformatics meets molecular mechanics: a combined application of knowledge-based pose scoring and physical force field-based hit scoring functions improves the accuracy of structure-based virtual screening. J. Chem. Inf. Model. 2012, 52, 16–28

[2] Fourches, D.; Politi, R.; Tropsha, A. J. Chem. Inf. Model. Target-Specific Native/Decoy Pose Classifier Improves the Accuracy of Ligand Ranking in the CSAR2013 Benchmark. *Submitted*.

Multi-dimensional activity cliff analysis

M Mackey, T Cheeseright, R Lawrence, M Slater
Cresset, Cambridge, UK. tim@cresset-group.com

During lead optimization the stepwise progression of compound activity is often disrupted by compounds that cause a disproportionately large (positive or negative) change in the biological response. These activity cliffs have long been recognized as an important source of information about the requirements of the protein for the series of interest [1]. However detection and analysis of these critical regions of SAR has been generally limited to the use of 2D similarity methods such as fingerprints.

Previously [2] we have presented methods to extend the detection of activity cliffs to include 3D similarity. The use of 3D similarity also often provides intuitive explanation of the fundamental reasons for these cliffs, leading to a much better understanding of why the structural change(s) led to such an unexpected change in binding. In particular, examining the differences in electrostatic potential between a pair of molecules can illuminate why structurally small changes (such as different aromatic substitution patterns) can lead to significant activity cliffs. Doing this requires accurate representations of molecular electrostatic potentials (Fig 1).

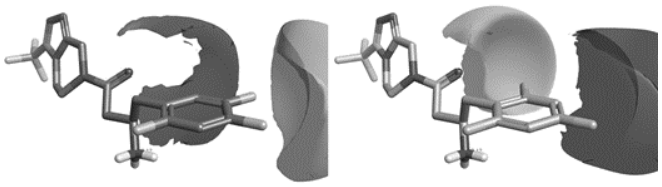


Figure 1 – difference maps showing the effect of moving a fluorine atom on electrostatic potential for two DPP IV inhibitors

The detection of activity cliffs for the primary activity end point is a valuable addition to the arsenal of drug discovery scientists. However, modern drug discovery rarely proceeds through the optimization of a single end point. More often project teams are tasked with optimizing the primary activity while minimizing the effect on a secondary, selectivity target or on a critical ADMET parameter. We have therefore studied the application of the 3D activity cliff analysis to multiple activity endpoints. These "selectivity cliffs" highlight where molecular changes have a large effect on the activity against one target but not another. Visualization of this data is a challenging task. If there are 500 molecules in the dataset then there around 250,000 data points to be analyzed for two activity end points. We will discuss the challenges and present some novel visualization techniques to deal with this data (Fig 2).

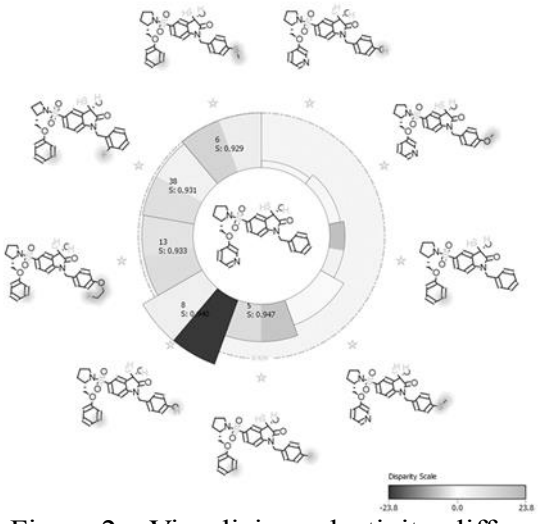


Figure 2 – Visualizing selectivity cliffs

1. Stumpfe , D.; Hu , Y.; Dimova , D.; Bajorath , J. Recent Progress in Understanding Activity Cliffs and Their Utility in Medicinal Chemistry. *J. Med. Chem.*, **2014**, 57 (1), 18–28.
2. Mackey, M; Presentation at ACS National Meeting, Indianapolis, **2013.**

Identification of small molecule inhibitors of type III secretion system ATPase EscN from enteropathogenic *E. coli*

Hanna Krezel[1], Karol Wrzeszcz[1#], Lukasz Bzdzion[1], Irmina Grzegorek[1], Katarzyna Nowinska[1], Grzegorz Chodaczek[1] and Wieslaw Swietnicki[1,2*]

[1]*Wroclaw Research Centre EIT+ Sp. z o.o., Dept. of Biotechnology, ul. Stablowicka 147, 54-066 Wroclaw, PL*

[2]*Institute of Immunology and Experimental Therapy of Polish Academy of Sciences, ul.Weigla 12, 53-114 Wroclaw, PL*

[*]Correspondence should be addressed to: wieslaw.swietnicki@eitplus.pl

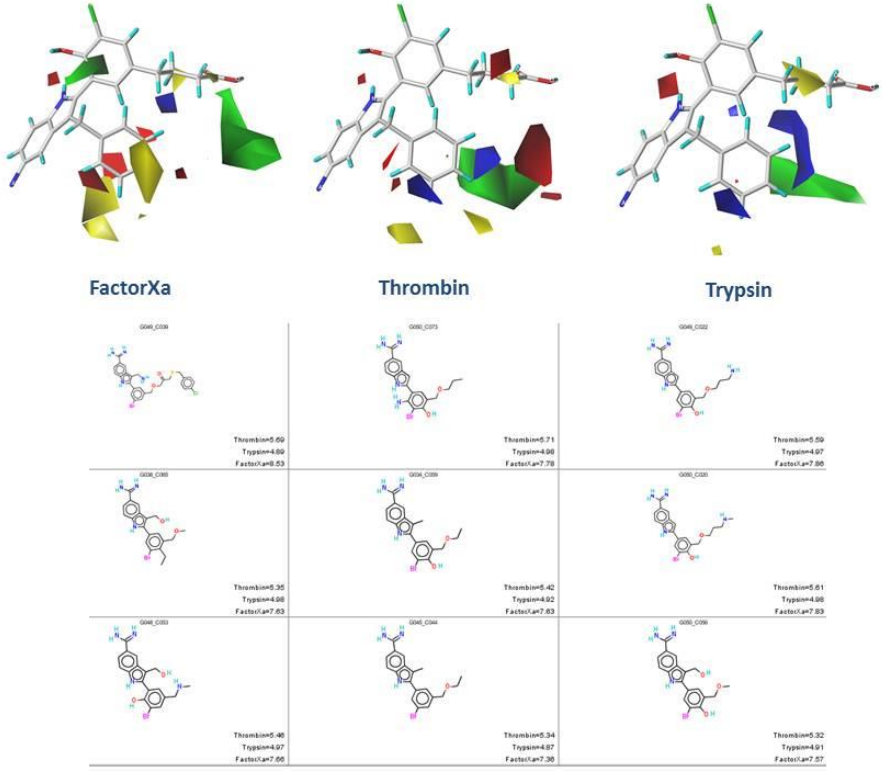
Abstract. Enteropathogenic *E. coli* (EPEC) is a human pathogen using type III secretion system for delivery of proteins directly into the human host. The system contains a single ATPase, EscN, which is essential for uncoupling of proteins from their complexes with chaperones before the delivery. The structure of EscN ATPase (PDB code: 2obm) was used to construct a search template to screen computationally for small molecule inhibitors blocking its active site. A total of twelve candidates out of seventy tested showed inhibitory activity against EscN ATPase and only five were non-toxic to human cells at 100 μM concentration. Two lead candidates were examined but only one, Compound 54, was selected for further optimization based on the experimental data. After one stage of QSAR optimization, at least five derivatives were found to be competitive inhibitors of EscN capable of blocking virulence factor secretion as determined by Western blotting with antibodies developed to an effector. One candidate, Compound 54-6, had IC_{50} of 58 μM but a very high (3.5 mM) inhibition constant K_i. The discrepancy was explained by proposed binding mode in the active site of enzyme. The compound was also minimally toxic to mammalian cells as determined by cell viability, apoptosis, cell cycle and mitochondrial potential interference assays. In the cell infection model of HeLa cells with EPEC, Compound 54-6 blocked effector secretion at least by 60% at 100 μM concentration. The compound decreased also pedestal formation and cell entry by bacteria, when analyzed by confocal microscopy. The second best inhibitor active in blocking effector secretion and HeLa cell infection by EPEC was Compound 54-9.

Multi-Criteria Drug Discovery: using CoMFA models to drive target specificity

Lei Wang[1], Brian Masek[1], Fabian Boes[1], Bernd Wendt[1] and Stephan Nagy[1]
[1] *210 N Tucker, Suite 350, St. Louis, MO, 63101, United States , 3149513430, lei.wang@certara.com*

A successful drug candidate will not only need to overcome ADME, physical and safety properties, but also often need to achieve a selectivity profile against related targets.

The approach we present here combines a de novo design approach with powerful ligand-based scoring function that consists of three different 3D QSAR models for different targets to generate selective inhibitors. A training set was developed with literature and public information where 50 molecules with activities for FactorXa, Trypsin and Thrombin were selected. Three different Topomer CoMFA models were built using the training set. By combining predictions of the activity profile with different weights and penalty score, a scoring function was created that can drive the invention of new and selective compounds.



Market Trends for In Silico Drug Discovery

In the past, in silico drug discovery was unsuccessful because of inaccurate and incomplete computational models (docking had low accuracy, selectivity was impossible, compute power was not there, and chemical property modeling was early). In the past 18 months, there have been significant breakthroughs with in silico drug discovery that may overcome many of these barriers. This panel will address the market dynamics for these new and better methods in the face of skepticism from the past, emerging proliferation of genomics, and increasing university involvement in therapeutics research. The impact of the digital health revolution will be addressed as well as the high cost of ignoring this trend.

Moderator: Ed Addison, CEO, Cloud Pharmaceuticals

Panelists:
Tom Caruso, Founder of Preclinix
John McGrath, Managing Partner of Infinity Venture Group
Taylor Milsal, Marketing at Autodesk
Ken Sorensen, Managing Director of Array Capital

Protecting the Products and Processes of In Silico Molecular Design: Patents, Copyrights, Open Source Licensing and the Public Domain

The protection of tools and methods used for in silico drug discovery has been made far more uncertain by recent events. Supreme Court decisions in cases such as *Mayo v. Prometheus* and *CLS Bank v. Alice* have called into question the basis of most patents on computer-implemented methods dealing with scientific principles, while guidelines now in use at the Patent Office are causing just about every claim mentioning a computer to be rejected out of hand. This complicates the already-problematic task of claiming libraries of compounds that, in many cases, are related through their properties rather than their chemical structures. Our panelists will address several aspects of this issue that is so critical to insuring that new leads and the methods used to design them may be protected to insure that innovators may continue to fund their development efforts.

Moderator: Lawrence Husick, Partner, Lipton, Weinberger & Husick

Panelists:
Gregory Aharonian, Patent Consultant, Editor of Internet Patent News Service, and Director, Center for Global Innovation/Patent Metrics
Larry A. Weinberger, Esquire, Lipton, Weinberger & Husick

Computational Drug Design Targeting Protein-Protein Interactions

Rachelle J. Bienstock, PhD

EPA. Email: rachelleb1@gmail.com

Novel discoveries in molecular disease pathways within the cell, combined with increasing information regarding protein binding partners has lead to a new approach in drug discovery. There is interest in designing drugs to modulate protein-protein interactions as opposed to solely targeting the catalytic active site within a single enzyme or protein. There are many challenges in this new approach to drug discovery, particularly since the protein-protein interface has a larger surface area, can comprise a discontinuous epitope, and is more amorphous and less well defined than the typical drug design target, a small contained enzyme-binding pocket. Computational methods to predict modes of protein-protein interaction, as well as protein interface hot spots, have garnered significant interest, in order to facilitate the development of drugs to successfully disrupt and inhibit protein-protein interactions. This review summarizes some current methods available for computational protein-protein docking, as well as tabulating some examples of the successful design of antagonists and small molecule inhibitors for protein-protein interactions. Several of these drugs are now beginning to appear in the clinic.

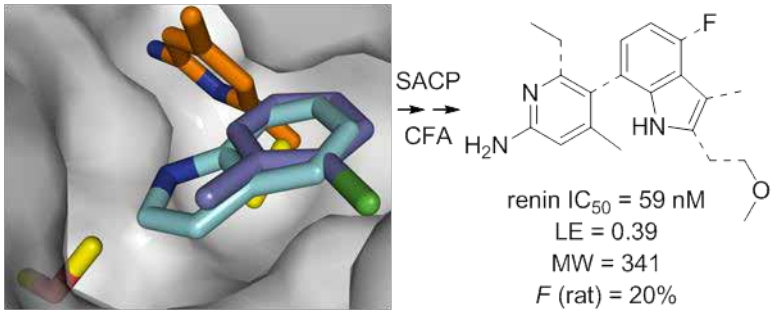
A Fragment Approach to Scaffold Hopping Using Simulated Annealing of Chemical Potential in Lead Discovery and Optimization - Renin: A Case Study

Ian S. Cloudsdale, John K. Dickson, Jr, Thomas E. Barta, Brian S. Grella, Emilie D. Smith, and Frank Guarnieri

Ian Cloudsdale, BioLeap, Inc., 6350 Quadrangle Drive, Suite 110, Chapel Hill, North Carolina 27517, United States isc@bioleap.com

ABSTRACT

We have applied Simulated Annealing of Chemical Potential (SACP) to a diverse set of ~150 very small molecules to provide insights into new interactions in the binding pocket of human renin, a historically difficult target for which to find low MW inhibitors with good bioavailability. SACP provides an efficient, thermodynamically principled method of ranking chemotype replacements for scaffold hopping and manipulating physicochemical characteristics for drug development. We introduce the use of Constrained Fragment Analysis (CFA) to construct and analyze ligands composed of linking those fragments with predicted high affinity. This technique addresses the issue of effectively linking fragments together and provides a predictive mechanism to rank order prospective inhibitors for synthesis. The application of these techniques to the identification of novel inhibitors of human renin is described. Synthesis of a limited set of designed compounds provided potent, low MW analogs (IC_{50}s < 100 nM) with good oral bioavailability (*F* >20-58%).



Putting Cheminformatics Mobile Apps and Collaborative Tools into Action for Tuberculosis

Sean Ekins[1,2], Alex M. Clark[3] and Joel S. Freundlich[4,5]

[1] *Collaborative Drug Discovery, Inc., 1633Bayshore Hwy, Suite 342, Burlingame, CA, U.S.A*
[2]*Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, U.S.A. ekinssean@yahoo.com*
[3] *Molecular Materials Informatics, 1900 St. Jacques #302, Montreal, Quebec, Canada H3J 2S1*
[4]*Department of Medicine, Center for Emerging and Reemerging Pathogens, Rutgers University – New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07103, USA.*

[5]*Department of Pharmacology & Physiology, Rutgers University – New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07103, USA.*

In order to develop potential therapeutics for *Mycobacterium tuberculosis* (*Mtb*) (which kills over 1.4M people per year) we have explored the use of different informatics approaches and databases. The proliferation of mobile devices has seen the development of apps that can be used for sophisticated chemistry workflows and we have taken advantage of these. These apps can offer capabilities to the practicing medicinal chemist, that are approaching those of conventional desktop-based software, whereby each app focuses on a relatively small range of tasks. Apps for drug discovery are already evolving rapidly and are able to communicate with each other to create composite workflows of increasing complexity, enabling informatics aspects of drug discovery (i.e. accessing data, modeling and visualization) to be done anywhere by potentially anyone. We will describe some of our *Mtb* workflows as well as our use of collaborative tools such as CDD Vault.

We recently developed a freely available mobile app (*TB Mobile*) for both iOS and Android platforms that displays *Mtb* active molecule structures and their targets with links to associated data. The app was developed to make *Mtb* target information for approximately 800 molecules available to as large an audience as possible. We have recently updated the iOS version of the app to include an implementation of ECFP_6 fingerprints that we have made open source. Using these fingerprints, the user can propose compounds with possible anti-TB activity, and view the compounds within a cluster landscape. Proposed compounds can also be compared to existing target data, using a näive Bayesian scoring system to rank probable targets. We have also curated 20 further compounds to evaluate this version of the app and associated targets. *TB Mobile* can now manage a small collection of compounds that can be imported from external sources, or exported by various means such as email or app-to-app inter-process communication. *TB Mobile* represents a valuable dataset, data-visualization aid and prediction tool which we will demonstrate.

We are using a combination of mobile apps and commercial desktop software to further our goals in working with collaborators on *Mtb*. We will illustrate how we use collaborative software and have recently implemented machine learning models into this tool that have been used for *Mtb* drug discovery. The approaches we illustrate could also be used outside of the neglected disease space for drug discovery.

Eliminating the Dependence of QSAR Models on Tautomeric Representation

Marvin Waldman, Robert Fraczkiewicz, and Robert D. Clark

Simulations Plus, Inc.42505 10th Street West, Lancaster, CA 93534 robert@simulations-plus.com

The quality and predictivity of most QSAR models used in drug design and development often depend on the particular tautomeric and valence structures used to represent the molecules of interest. This is because the location of hydrogens, bond orders, and formal charges affect the values of atomic and molecular descriptors upon which the models are based. For descriptors such as partial charges, this dependence is typically due to basing atom typing on hybridization: e.g., sp^3-hybridized oxygen in -OH is treated differently than its sp^2-hybridized counterpart in the =O group. That same oxygen may well occur in both forms in different tautomers of the same molecule, however. In the aqueous environment of interest in drug design applications, representing a compound as any one of its tautomers is likely to distort the QSAR model obtained. Even worse is the danger of choosing a tautomer present at low abundance, which will bias the model building process or the reliability of predictions or both. We have developed a descriptor generation method which is independent of tautomer and valence structure representation to address this problem and will illustrate its application to the atomic descriptors used in S+pKa, which is our global model of protic ionization constants. Preliminary results will be shown and compared with the "traditional" approach along with a discussion of advantages and potential pitfalls of the method.

5

QM/MM Study of Mg(II)/Mn(II) Divalent Metal Dependence on Structure and Reactivity of DNA Polymerase Lambda (λ)

Sangwook Wu[1] and Shubin Liu[2]

[1]Department of Physics, Pukyong National University, Busan, 608-737, Korea
[2]Research Computing Center, University of North Carolina, Chapel Hill, North Carolina 27599-3420, United States

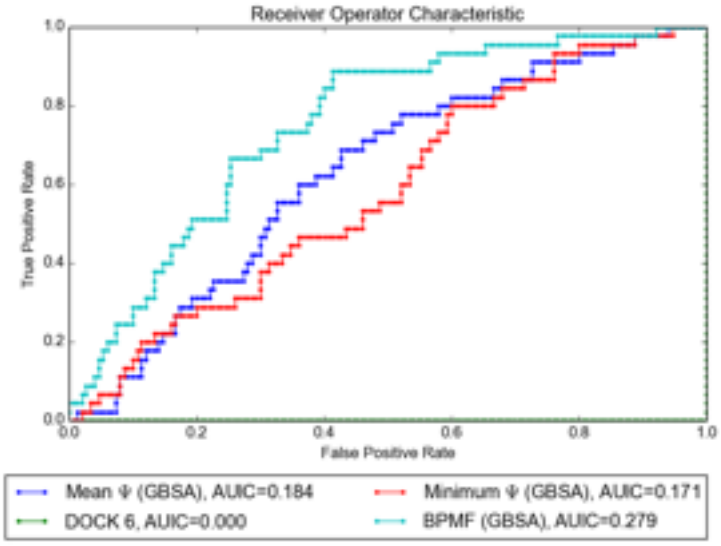
The DNA polymerase lambda (λ), belonging to family X, performs the repair of the damaged DNA. It is composed of the DNA binding domain, nucleotide binding domain and catalytic domain. In the catalytic domain, the DNA polymerase λ coordinates with two divalent metals (Mg(II)/Mn(II)). Interestingly, DNA polymerase λ has a strong preference of Mn(II) over Mg(II) in nucleotide incorporation. Molecular dynamics simulation (MD) shows no significant difference in RMSD values for the catalytic domain of solvation–equilibrated structures for both Mg(II)/Mn(II) ions (during 10 ns simulations). This motivates us to investigate the catalytic domain of DNA polymerase λ for the impact of Mg(II)/Mn(II) on its structure and reactivity by employing the QM/MM method. A broken-symmetry density functional theory study verifies an antiferromagnetically coupled state between the Mn(II)–Mn(II) pair, unlike the Mg(II)-Mg(II) pair, where no such coupling mechanism is in place. From the energy difference between the high spin state (HS) and the broken-symmetry state (BS) of the Mn(II)–Mn(II) pair, we observe the significant difference of local environment in the catalytic domain, which leads to the weakly antiferromagnetic interaction of d^5–d^5 electrons through the superexchange coupling mechanism. The antiferromagnetic property of Mn(II)–Mn(II) pair in the BS state may play a critical role in nucleotide incorporation into the catalytic domain of DNA polymerase λ.

Implicit Ligand Theory: A New Paradigm for Virtual Screening

David D.L. Minh[1]

[1] *Department of Biological and Chemical Sciences, Illinois Institute of Technology, 3101 S Dearborn Av., Chicago IL 60616, dminh@iit.edu*

[**Talk preferred**] Implicit ligand theory is a new framework for standard binding free energy calculations that is based on multiple rigid configurations of the receptor [1]. Due to receptor rigidity, computational methods based on implicit ligand theory are much faster than alchemical free energy methods that require full system flexibility. Nevertheless, the calculations rigorously account for binding entropy as well as enthalpy.



A binding free energy between a ligand and *rigid* receptor configuration, or binding potential of mean force (BPMF), is an intermediate quantity towards estimating standard binding free energies. BPMFs may be estimated using simulation methods common to other alchemical calculations, including Hamiltonian replica exchange. BPMFs were calculated for receptor-ligand pairs from the directory of useful decoys, enhanced (DUD-E) [2], for ampicillin C beta-lactamase. For the 47 actives and the top 150 scoring decoys selected by UCSF DOCK 6, BPMFs exhibited a superior receiver operating characteristic (ROC) and area under semi-log curve (AUlC) over the minimum and mean interaction energy (ψ).

In addition to ROC curves based on BPMFs for other systems, I plan to present standard binding free energy calculations; these will be assessed by comparison to known experimental values. I hope to make the case that *in silico* drug discovery can soon move beyond the paradigm of minimum interaction energies and carefully consider entropic contributions to ligand binding.

[1] David D. L. Minh. "Implicit ligand theory: Rigorous binding free energies and thermodynamic expectations from molecular docking". Journal of Chemical Physics 137 (2012): 104106.
[2] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet. "Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking." Journal of Medicinal Chemistry 55 (2012): 6582–6594.

Algorithm to Search Diverse Optimum Molecules from the Small Molecule Universe

Chetan Rupakheti[1], Aaron Virshup[2], David Beratan[2], Weitao Yang[2]

[1] Program in Computational Biology and Bioinformatics, Duke University + *crr22@duke.edu*
[2] Department of Chemistry, Duke University + *aaron.virshup@duke.edu*
[2] Department of Chemistry, Duke University + *david.beratan@duke.edu*
[2] Department of Chemistry, Duke University + *weitao.yang@duke.edu*

The small molecule universe (SMU) is defined as a set of over 10^{60} synthetically feasible molecules with molecular weight less than 500 Da. Exhaustive enumerations and testing of all SMU molecules for the purpose of designing a library of useful drugs or materials is impossible. We describe an extension of ACSESS framework[1], that can generate a representative library of small molecules belonging to the SMU targeted towards a physical property (example: protein-ligand docking score). We show that the method is efficient compared to exhaustive enumeration and existing evolutionary algorithms for generating libraries of useful molecules by testing in an enumerated GDB9 chemical universe containing ~300,000 molecules.

[1] A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang, and D. N. Beratan, "Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds.," *J. Am. Chem. Soc.*, vol. 135, no. 19, pp. 7296–303, May 2013.

8

Photo-damaged DNA Repair Under Extreme Conditions

Shoresh Shafei, Agostino Migliore and David Beratan[1]

[1] *Department of Chemistry, Duke University, shoresh.shafei@duke.edu*

Ultraviolet light can damage DNA by forming cyclobutane pyrimidine dimer (CPD) lesions. The repair strategy catalyzed by photolyases involves photo-excitation of the flavin adenine dinucleotide (FAD) cofactor and transfer of an electron to the DNA lesion, which leads to dimer cleavage. This repair mechanism has not yet been explored under extreme conditions, such as very high or low temperatures.

We theoretically investigate the electron transfer near the freezing or boiling point of the solvent by studying two extremophile forms of the ancient DNA-repair enzyme DNA photolyase. Our simulations of the docked photolyase and DNA aim to determine how DNA fluctuations influence the propensity for its photochemical damage and how modest thermal fluctuations in a docked pose and larger-amplitude structural fluctuations among multiple poses influence DNA repair. We examine the dependence of damage and repair propensities on the temperature and solvent conditions by combining conformational sampling with quantum calculations of the electron tunneling interactions that affect the ET rates.

Analyzing High-Throughput Screening Data Using Freely-Accessible Navigator Software

Denis Fourches and Alexander Tropsha

Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. Email: fourches@email.unc.edu

Many drugs are characterized by polypharmacological mechanisms of action. Thus, prospective drug discovery studies often start by testing large compound libraries in multiple and diverse High-Throughput Screening (HTS) assays. These large heterogeneous data collections pose numerous computational challenges concerning processing, curation, and analysis of the output files generated by plate readers. We have developed the freely-accessible HTS Navigator software [1] to enable and facilitate the processing and analysis of polypharmacological HTS data. We report on the capabilities of the Navigator (Figure 1) and present several case studies where we employed cheminformatics approaches embedded within the Navigator to curate [2] and analyze large datasets of compounds tested against different panels of targets. Examples include libraries of compounds tested for their inhibition potencies across several CYP450s; or for their inhibition of multiple protein kinases; or with binding profiles against multiple GPCRs. We show that Navigator can quickly identify and flag compounds with unique mono- and dual-selectivity for certain targets in the curated HTS matrix. We discuss the problem of experimental variability in HTS data and its consequences for molecular modeling. We emphasize the synergistic potential of different cheminformatics approaches to detect both false-positive and false-negative compounds using neighborhood analysis and target baseline correction factors. Finally, we describe the Chemical−Biological Read-Across (CBRA) approach [3] also implemented in the Navigator program to infer the activity of external compounds from both chemical (defined by chemical similarity) and biological (defined by the similarity of HTS profiles) analogues.

[1] Fourches D, Sassano MF, Roth BL, Tropsha A. HTS navigator: freely accessible cheminformatics software for analyzing high-throughput screening data. *Bioinformatics*, **2014**, 30, 588-589.

[2] Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J Chem Inf Model,* **2010**, 50, 1189-1204.

[3] Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, Tropsha A. Integrative Chemical-Biological Read-Across Approach for Chemical Hazard Classification. *Chem Res Toxicol*, **2013**, 26, 1199-208.

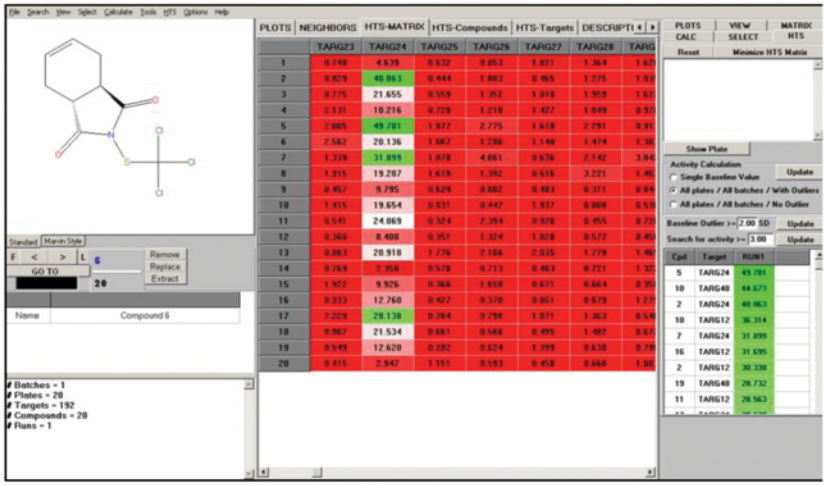


Figure 1. HTS Navigator program.

A Novel Treatment for Liver Injury in Western Diet Mouse Models

Zijun Yu,[1] Cen Xie,[2] Frank J. Gonzalez[2]

[1]31 Center Dr, Bethesda, MD 20892 + zijunyu3@gmail.com
[2]31 Center Dr, Bethesda, MD 20892

It has long been known that the actions of farnesoid X receptor (FXR) and G protein-coupled receptor TGR5 are key in preventing liver damage such as non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH), and inflammation [1-3]. Recently, 6α-ethyl-3α,7α-dihydroxy-24-nor-5β-cholan-23-sulfate (INT-767) has been identified as an agonist for both FXR and TGR5, the first described to potently and selectively activate both receptors [8]. Preclinical results indicate that INT-767 is a safe and effective modulator of FXR and TGR5-dependent pathways, suggesting potential clinical applications in the treatment of liver and metabolic diseases [8-9]. However, much of the mechanism in which INT-767 treats liver disease is still unclear. The aim of the current work is to characterize metabolic profiles in the Western diet induced-NAFLD mouse model using metabolomics and lipidomics to provide insight into the therapeutic mechanisms of dual FXR/TGR5 agonist INT-767 on liver injury. Hepatic mRNA levels of genes associated with bile acid synthesis, lipid metabolism, inflammation, and fibrosis are compared before and after treatment of INT-767. This study elucidates the process by which a new potent and versatile drug combats liver disease, paving the way for safer treatment of liver injury, the most prevalent chronic disease in the world.

BCL-2 and BAX, similar folding with antagonistic role: Molecular Dynamics Study

J.L. Rosas-Trigueros [2*], I. Ilizaliturri-Flores [3*], B. Zamora-López [4], J. Correa-Basurto [5], C.G. Benítez-Cardoza[1], A. Zamorano-Carrillo[1]

[1]Laboratorio de Investigación Bioquímica y Biofísica Computacional, ENMH-IPN, [2]Laboratorio de Graficación e Inteligencia Artificial, ESCOM-IPN, [3]INR-SSA, [4]Departamento de Salud Mental, Facultad de Medicina, FACMED-UNAM, [5]Laboratorio de Bioinformática, Modelado Molecular y Diseño de Drogas, ESM-IPN. Guillermo Massieu H. 239, Gustavo A. Madero, Residencial La Escalera, 07320 Ciudad de México, Distrito Federal.

**Both authors participated equally*

Members of the Bcl-2 family proteins have been reported as regulators of the cell death and survival. By sequence analysis of these proteins, four BH (BH: Bcl-2 homology) domains BH1 (19 AA), BH2 (15 AA), BH3 (14 AA), and BH4 (20 AA) have been characterized. Also, this family includes 25 proteins divided into antiapoptotic (Bcl-2, Bcl-w, Bcl-xL, A1, and Mcl-1), proapoptotic (Bax, Bak y Bok). Both subsets of proteins regulate the apoptosis by homo- and heterodimerization among them. Structural studies of Bcl-2 family proteins show that they exhibit a similar folding to forming-pores toxins. Bcl-2 is a regulator of apoptosis and overexpressed in several cancers. Consequently, this protein is attractive as a target of drugs, antisense oligonucleotides, or inhibitory small molecules, to explore new therapeutical options against cancer. In turn, Bax (BH3-BH1-BH2) is an essential protein in the apoptosis, actively promoting changes into mitochondrial potential membrane and activating caspases. In this work, we explore by molecular dynamics the conformational spaces of the Bcl-2 and Bax proteins, describing their similarities and more stable domains. Also, we propose possible mechanisms and structures that participate in the cell death regulation.

References:

1.Rosas Trigueros JL, Correa Basurto José, Benítez Cardoza CG and Zamorano Carrillo A. Insights into the structural stability of Bax from molecular dynamics simulation at high temperatures. *Protein Science,* 20(12)2035-2046, 2011. ISSN: 0961-8368.

2.Ilizaliturri-Flores I, Correa Basurto J, Benítez Cardoza CG and Zamorano Carrillo A. A study of the structural properties and thermal stability of human Bcl-2 by Molecular Dynamics Simulation. *Journal of Biomolecular Structure and Dynamics*, ISSN: 0739-1102 2013. DOI:10.1080/07391102.2013.833858.

Platinum Sponsors

Autodesk www.autodesk.com
Autodesk helps people imagine, design and create a better world. Everyone—from design professionals, engineers and architects to digital artists, students and hobbyists—uses Autodesk software to unlock their creativity and solve important challenges.

Cloud Pharmaceuticals www.cloudpharmaceuticals.com
Cloud Pharmaceuticals designs new drugs and drug candidates and accelerates the development process by implementing new, innovative cloud computing technology in place of traditional methods. We manage development of multiple assets through an extensive network of partners. This reduces cost, decreases development time, and creates novel molecules with therapeutic value.

Premium Sponsors

Quantum Cures Foundation www.quantumcures.org
Quantum Cures Foundation is dedicated to solving rare, neglected, and orphan diseases worldwide through the collaborative effort of public and private organizations using advanced technology to discover new drugs and repurpose old ones for these diseases.

Schrödinger www.schrödinger.com
Schrödinger is a scientific leader in computational chemistry, providing software solutions and services for life sciences and materials research.

Gold Sponsors

BioIT Solutions www.bioit.com
Protect and manage your data and workflows with 1Platform from BioIT Solutions. Science moves fast and you need an IT system that can pivot quickly. Streamline your lab work with 1Platform4 Discovery. Power your Diagnostics business with 1Platform4 Diagnostics. Manage your company's clinical and pre-clinical testing activities with 1Platform4 Drug Development.

ChemAxon www.chemaxon.com
ChemAxon is a leader in providing chemical software development platforms and desktop applications for the biotechnology and pharmaceutical industries. By focusing upon active interaction with users and core portability, ChemAxon creates leading edge cross platform solutions to power modern cheminformatics and chemical communication.

PreClinix www.preclinix.com
Based in the Research Triangle Park of North Carolina, Preclinix incubates drug assets in partnership with the owners of those assets, taking them from laboratory discovery to the clinic. Preclinix is transforming the preclinical drug development process by combining management expertise, predictive filtering, just-in-time services, and capital.