Workshop

Optimizing Drug Design

Lorentz Center, 20-23 July 2009

Abstracts

http://www.lorentzcenter.nl/lc/web/2009/359/info.php3?wsid=359
www.pharma-it.net

Keynote Presentations

Matthew Segall

Enabling Interactive Multi-Objective Optimization for Drug Discovery Scientists

Biofocus DPI, Oxford, UK

The key decisions in drug discovery involve the design and selection of compounds with an appropriate balance of many properties and hence a high chance of achieving the therapeutic goals of a project. To guide this multi-objective optimization, many forms of data, *in silico*, *in vitro* and *in vivo*, must be integrated and assessed against a project's requirements. The impact of this complex analysis is greatest if it is interactive and intuitive to key decision makers, allowing new ideas to be explored in real time.

We will describe an approach that starts by clearly defining the property requirements and their relative importance. All of the available data are then integrated and assessed against these combined criteria to give a score for each compound, an estimate of the likelihood success of each compound against the property criteria. Within this assessment, the uncertainty in the data is taken into account; most sources of drug discovery data have significant statistical or experimental uncertainties and, when these are combined, we should consider the impact on our ability to confidently choose between different compounds or chemistries. Furthermore, when selecting compounds, it may not be appropriate to choose only the 'best' compounds; it is often possible to explore greater chemical diversity, to spread risk and gather additional information, with little loss in terms of the overall quality of the selected compounds.

The impact of this analysis on the direction of the chemistry must be clearly visualised to rapidly explore alternatives and guide decisions. We will discuss approaches to intuitively present the outcomes and underlying data.

Finally, we will consider how we can challenge the decision-making process itself. Criteria in drug discovery are often subjective and we should consider the impact these have on the decisions we take before they are finally fixed.

Carlo Poloni* and Danilo Di Stefano**

Multiobjective Robust Design Optimization of docked ligands

University of Trieste, Via A.Valerio 10, 34142 Trieste, Italy
**ESTECO srl, AREA Science Park, Padriciano 99, 34012 Trieste, Italy
poloni@units.it, distefano@esteco.com

Protein – ligand docking is one of the key steps in developing new drugs. Potential new candidate drugs are matched to the receptor and the best is selected for further steps. Computational approaches to simulate docking problems are becoming more and more popular, capturing the interest of scientific and industrial communities. But many problems arise in computational approaches: efficient definition of force fields that capture main chemical and physical interactions, efficient search algorithms in the conformational space, molecular representation. All those aspects put severe limitations on a docking analysis, the main of which is considering the receptor as a rigid body, allowing conformational motions only for the small molecule.
This work proposes a new methodology to overcome this main limitation. The approach is based on a multiobjective robust design optimization of candidate docked conformations. Robust design is an engineering approach to the design that tries to develop products that are not over-optimized with respect to the design variables but, vice versa, that are tolerant to some uncertainties. Uncertainties can arise from difficult parameter control due to instrumental lack of sensitivity, or they can arise from stochastic oscillations of environmental variables (for example temperature, pressure).
The application of a robust design methodology to a docking problem is based on interpreting the candidate drug as a design to be placed in the environment formed by the receptor molecule. Small conformational motions of side chains of the receptor binding site are considered as uncertainties of design parameters. This way, the docking is interpreted as a design of the ligand in a small-perturbed environment.
A multiobjective formulation of a robust docking was performed using the M.O.R.D.O. module in modeFRONTIER™ optimization tool, which integrates key concepts from sensitivity analysis, multiobjective optimization and evolutionary algorithms, allowing to explore the conformational search space in a robust multiobjective way.
Preliminary results on standard benchmark sets seem to confirm the effectiveness of the methodology. Further test are currently under development.

Nathan Brown

Multiobjective Optimisation in Drug Discovery: Where Are We Now?

Ian Parmee

Evolutionary and Agent-based Search and Exploration in Combinatorial Chemical Library Design and in the De Novo Design of Novel Molecules

---

Christos Nicolaou

Focusing Multi-Objective Optimization Search: Knowledge-driven De Novo Drug Design

---

Regular Session Papers

John Holliday, Aysha Al Khalifa, Shereena M. Arif, Maciej Haranczyk, Peter Willett

Recent Studies for Optimising Similarity-based Virtual Screening

Molecular similarity methods, including similarity search, database clustering and compound selection, rely on three key components: a set of attributes or features which are used to characterize the compounds, a weighting scheme which might be applied to those features, and a measure which will quantify the degree of similarity between the respective sets of feature. Recent studies in the area of virtual screening are described which seek to compare the effects of variations in all three of these components. Alternative sets of coefficient have been applied, either individually or in combination using data fusion techniques, in order to identify those which are most applicable to the similarity method chosen and the characterizations used. Novel methods of data fusion have been investigated, based on alternative similarity coefficients and characterizations which, though simple in operation, support the use of data fusion as a means to optimize similarity search. In addition, alternative weighting schemes have been applied to commonly-used compound characterizations and their effects on virtual screening performance investigated. Results show that considerable variation in performance can be obtained when variations in these three components are introduced. The correct choice of components is, therefore, an important consideration when designing a virtual screening procedure. However, the variable effect of these components is shown to have a positive effect when consensus methods are applied using data fusion techniques.

Pascal Bonnet, Ortho Biotech Oncology Research & Development, Johnson & Johnson, Beerse, Belgium

An efficient conformational search tool for drug design

As computational drug design becomes increasingly reliant on virtual screening and high-throughput 3D modeling workflows, the need for fast, robust and

reliable methods for sampling molecular conformations has become greater than ever. Furthermore, chemical novelty is at a premium, forcing medicinal chemists to explore more complex structural motifs and unusual topologies. This necessitates the use of conformational sampling techniques that work well in all cases. Here, we compare the performance of several popular conformational search algorithms on three broad classes of macrocyclic molecules. These methods include Catalyst, Caesar, Macromodel, MOE, Omega, Rubicon and two newer self-organizing algorithms developed at Johnson & Johnson known as stochastic proximity embedding (SPE) and self-organizing superimposition (SOS). Our results show a dramatic advantage for the three distance geometry methods (SOS, SPE and Rubicon) followed to a lesser extent by Macromodel. The remaining techniques, particularly those based on systematic search like Omega and Caesar, often failed to identify any low energy conformations and are unsuitable for this class of molecules. Taken together with our previous study on drug-like molecules, these results suggest that SPE and SOS are the most robust and universally applicable methods invented to date, with the latter being preferred because of its superior speed.

Michael Emmerich and Viktoria B.V. Lee, Alexander Aleman

Landscape Analysis in Optimizing Molecular Drug Design

Recent Tools for the analysis of large discrete search spaces in optimization are discussed for their application in Molecular Drug Design.

Lisa Michielan

Investigating Potency and Selectivity of Human Adenosine Receptor Antagonists

Adenosine receptors (ARs) are widely considered interesting and promising therapeutic targets in the family of G-protein coupled receptors (GPCRs). The growing interest on the distinct AR subtypes (A1, A2A, A2B and A3) has inspired the development of potent and selective ligands, demonstrating the potential role of the AR in several physiopathological processes. Generally, the therapeutic application of compounds not differentiating between receptor subtypes might be accompanied by efficacy problems or limiting side-effects. Therefore, the design of drug candidates with improving potency, but with minimum side-effects, is the key to therapeutic success. To date, very few valuable in silico tools are available for the prediction of receptor subtype selectivity. We will present a novel application of the multi-label classification by combining the autocorrelated molecular descriptors encoding for the Molecular Electrostatic Potential (autoMEP) with cross-training Support Vector Machine (ct-SVM) analysis to select AR antagonists with high potency and

subtype selectivity. AR antagonists may present multiple molecular properties of multiple AR subtypes. The alternative ct-SVM method represents thus the appropriate strategy for multiple and overlapping classes tasks. Three statistically valuable autoMEP/ct-SVM models, based on decreasing thresholds of receptor binding affinity (500 nM, 250 nM and 100 nM), have been generated to simultaneously describe human A1R, A2AR, A2BR and A3R subtypes potency profile and the potential selectivity of a large collection of known xanthine and pyrazolo-triazolo-pyrimidine analogs. Our autoMEP/ct-SVM models have been applied in series as useful screening approach for new xanthine and pyrazolo-triazolo-pyrimidine derivatives to filter potent and selective AR antagonists. We have further validate our strategy by predicting 13 new synthetized pyrazolo-triazolo-pyrimidine analogs, inferring the full AR potency spectrum and selectivity profile.

Eelke v. der Horst

Multi-Objective Evolutionary Design of Selective Adenosine receptor ligands

LACDR, The Netherlands

M.S.Baig, N. Manickam

Homology modeling and docking studies of Comomonas testosterone B-356 biphenyl-2, 3-dioxygenase involved in degradation of polychlorinated biphenyls.

Environmental Biotechnology Division, Indian Institute of Toxicology Research Lucknow-226001, India

Biphenyl dioxygenase is a microbial enzyme which catalyzes the stereospecific dioxygenation of aromatic rings of biphenyl congeners leading to their degradation. Hence, it has attracted the attention of researchers due to its ability to oxidize chlorinated biphenyls, which are one of the serious environmental contaminants. In the present study, the three-dimensional model of á-subunit of Biphenyl dioxygenase (BphA) from Comamonas testosteroni B-356 has been constructed. The resulting model was further validated and used for docking studies with a class of chlorinated biphenyls where the kinetic parameters of these biphenyl compounds were well matched with the docking results in terms of conformational and distance constraints. The binding properties of these biphenyl compounds along with identification of critical active site residues could be used for further site-directed mutagenesis experiments in order to identify their role in activity and substrate specificity,

ultimately leading to improved mutants for degradation of these toxic compounds.

Divita Garg*, Koen Nauwelaerts[#], Michael Sattler[#], Rebecca Wade*

Thymidylate Synthase protein-mRNA interactions

* EML-Research gGmbH, Schloss Wolfsbrunnenweg 33, 69118 Heidelberg, Germany
# Chair of Biomolecular NMR Spectroscopy, Lichtenbergstrasse 4, 85747 Garching, Germany

The protein thymidylate synthase is an important enzyme in the thymidine biosynthesis pathway and an established anti-cancer target. In addition to playing a role in thymidine synthesis, it interacts with its own mRNA to autoregulate its translation. Interaction of the protein with drug molecules interrupts this autoregulation, leading to development of resistance. A small fragment from the complete mRNA is reported to be the most important for interacting with the protein. We aim to obtain insights into this interaction using computational techniques such as PIPSA, structure prediction and simulations, as well as experimental techniques such as NMR. Additionally, we are studying the binding of small molecules to the RNA fragment which could also provide a means to prevent the development of resistance. Preliminary results will be presented in this poster.

Thorsten Meinl

Multiobjective Diversity Selection for HTS

Diversity Selection is a common task in early drug discovery, be it for removing redundant molecules prior to HTS or reducing the number of molecules to synthesize from scratch. One drawback of the current approach, especially with regard to HTS, is, however, that only the structural diversity is taken into account. The fact that a molecule may be highly active or completely inactive is usually ignored. This is especially remarkable, as quite a lot of research is involved in improving virtual screening methods in order to forecast activity. We therefore present a modified version of diversity selection, which additionally takes the predicted activities of the molecules into account. Not very surprisingly both objectives - maximizing activity whilst also maximizing diversity in the selected subset - conflict. As a result, we end up with a multiobjective optimization problem for which we present several solutions, ranging from traditional evolutionary algorithms and graph theoretic approaches to a specialized novel algorithm known as Score Erosion.

Amiram Goldblum, David Marcus, Anwar Rayan

Forming focused libraries and discovering active molecules with Iterative Stochastic Elimination algorithm

Dept. Of Medicinal Chemistry, School of Pharmacy, TheHebrew University of Jerusalem, Israel 91120
amiram@vms.huji.ac.il, http://www.md.huji.ac.il/models

Our algorithm, Iterative Stochastic Elimination (ISE) enables to find the best set of descriptors and descriptor ranges to model the difference between active and inactive molecules on a specific target. By using molecular properties as our descriptors and applying ISE to optimize descriptors and their ranges for a particular bioactivity, we obtain a model composed of sets of filters, each consisting of a group of property ranges. The set of filters (post-clustering) serves to assign a Molecular Bioactivity Index (MBI) of that activity to any molecule (due to its properties), thus having the ability to grade molecules and form "focused libraries" of any desired size. Molecules with top MBI values may be focused further for other required properties such as Drug Likeness, Selectivity, BBB crossing, solubility etc. We have shown extremely high enrichments with MBI, as well novelty in molecular structures due to low similarity by Tanimoto criteria. Top molecules may be further focused by docking with ISE-dock (Gorelik and Goldblum, Proteins 71, 1373-86, 2008) in order to combine 2D with 3D properties, in those cases that the structure of the target is available.
Recently, we produced a focused library of ~800 molecules from the ZINC database as potential inhibitors of acetylcholinesterase (AChE), following the production of a highly efficient model. These molecules were docked to mouse AChE and the top 10 molecules were purchased. Only 9 arrived, 1 of which was in minute amounts. Out of the 8 molecules tested for inhibiting Erythrocyte AChE, 5 were soluble and 3 of them had IC50 values in the low mircomolar range (3.25, 35. and 3.75 microMolar). These molecules have now been purchased in larger amounts to perform small animal test of cholinergic effects (with Marta Weisntock,-Rosin, Pharmacology Dept. at HUJI) and for co-crystallization (with Joel Sussman, Weizmann Inst.).

Alexander Aleman, Michael Emmerich, Johannes Kruisselbrink

Enhancing Search Space Diversity in Evolutionary Multiobjective Optimization of Molecules using Niching

Leiden University, LIACS/LACDR

Evolutionary Multiobjective Optimization (EMO) is often used as a tool for generating new ideas for molecular designs. Since it is difficult to consider all

constraints on drug like molecules a-priori often a large fraction of the molecules found by EMO tools is not feasible. It would thus be desirable to present a diverse population of high quality molecules (for the stated objectives) such that the chemist can select those molecules from that set that seem most suitable. However, EMO techniques have the tendency to focus the search in a particular region of the search space, e.g. a local Pareto front, thereby possibly failing to detect other high performance regions. Niching in the search space has been suggested as an approach to enhance diversity of solutions in evolutionary algorithms. To incorporate it into algorithms for searching the chemical space, a suitable distance measure is needed. We discuss experiences with niching in chemical space and show that the diversity of populations can be strongly enhanced without loosing quality of the results.

Heike Trautmann,

Multiobjective Optimization with Desirability Functions and Desirability Indices

Department of Statistics, TU Dortmund

The Concept of desirability, introduced by Harrington (1965), is a method for multiobjective optimization in industrial quality management. Via desirability functions (DFs), which allow for comparing different scales of the quality criteria by mapping them to [0, 1], and the desirability index (DI) the multivariate optimization problem is converted into a univariate one. Based on design of experiment methods and nonlinear optimization techniques levels of process influencing factors can be determined that maximize the DI and therefore the overall process quality.
Modifications came up either in terms of more flexible DFs (e.g. Derringer and Suich (1980)) or in terms of different DIs. The desirability concept became widely accepted for practical applications. The main problem is that in the classical approach the uncertainty of the models linking quality criteria to process influencing factors is neglected. Thus mostly the optimization result will be highly uncertain, and the optimal value of the DI will not be guaranteed in the course of the process. Recently density and distribution functions of the DFs and the DI were derived (Trautmann and Weihs (2006), Steuer (2005)) which allow for a robust optimization of the DI and for setting up control charts in order to monitor the optimized DI over time (Trautmann, 2004).
An overview about the desirability concept will be given including different types of applications. In addition a method for combining the desirability concept with multiobjective evolutionary algorithms in order to focus on relevant parts of the Pareto-front by integrating a-priori-expert-knowledge in the multiobjective optimization process
(Mehnen et al., 2007; Trautmann and Mehnen, 2009) will be presented.
G. C. Derringer and R. Suich. Simultaneous Optimization of Several Response Variables. Journal of Quality Technology,
12:214 – 219, 1980.

J. Harrington. The desirability function. Industrial Quality Control, 21:494 – 498, 1965.
J. Mehnen, H. Trautmann, and A. Tiwari. Introducing User Preference Using Desirability Functions in Multi-
Objective Evolutionary Optimisation of Noisy Processes. In Kay Chen Tan and Jian-Xin Xu, editors, CEC 2007,
IEEE Congress on Evolutionary Computation, pages 2687–2694, Swissotel The Stamford, Singapore, 2007.
D. St
euer. Statistische Eigenschaften der Multikriteriellen Optimierung mittels Wünschbarkeiten. PhD thesis, Technische
Universität Dortmund, 2005. URL http://hdl.handle.net/2003/20171.
H. Trautmann. Qualitätskontrolle in der Industrie anhand von Kontrollkarten für Wünschbarkeitsindizes – Anwendungsfeld
Lagerverwaltung. PhD thesis, Technische Universität Dortmund, 2004. URL http://hdl.handle.net/2003/2794.
H. Trautmann and J. Mehnen. Preference-based pareto-optimization in certain and noisy environments. Engineering
Optimization, 41:23–28, 2009.
H. Trautmann and C. Weihs. On the distribution of the desirability index using Harrington's desirability function.
Metrika, 63(2):207–213, 2006.

Johannes Bader and Eckart Zitzler

Integrating the Issue of Solution Robustness Into Hypervolume-based Multiobjective Search

In engineering design, usually complex optimization problems arise that involve multiple objective functions. In this context, robustness – here considered to be the sensitivity with respect to slight decision variable variations - is a crucial issue that needs to be taken into account when developing corresponding optimization methods.
In this talk, we discuss different concepts for integrating robustness into multiobjective search using evolutionary algorithms and similar techniques. The focus is on a recent approach that employs the hypervolume indicator for fitness assignment. We present ways to handle robustness, e.g., by additional objectives or constraints, and compare selected methods on different system design problems.

Ofer M. Shir

Experimental Optimization in Quantum Control

*Rabitz Group, Princeton University

Quantum Control (QC), sometimes referred to as Optimal Control or Coherent Control, aims at altering the course of quantum dynamics phenomena for specific target realizations, typically by means of closed-loop, adaptively shaped laser pulses. This field has experienced a rapid increase of interest during recent years, in parallel
to the technological developments of ultrafast laser pulse shaping capabilities, that made it possible to turn this early-days dream into reality. Quantum Control Experiments (QCE), the topic of this talk, consider the realization of QC in the laboratory, where the objective function evaluation cannot be done through a computer simulation, but rather requires the execution of a real-world experiment. The yield, or success, thus corresponds to a physical measurement. The optimization task of QC systems typically poses many algorithmic challenges, such as high-dimensionality, noise, constraints handling, etc. Its attractive feature, nevertheless, is the extremely short duration of an experiment, in comparison to other real-world experimental systems: A typical QC measurement is carried out in the kHz regime, allowing a wellaveraged single experiment to be recorded in the order of a single second. This talk will discuss the main characteristics of experimental optimization, particularly QCE systems, in comparison to optimization of simulated systems, and will also provide practical guidelines for real-world experiments with Evolutionary Algorithms.

Heike Trautmann, Boris Naujoks

Online Convergence Detection in Evolutionary Multiobjective Algorithms

The decision at which point in time an optimisation run should be terminated is a critical issue in global optimisation. Exactly the number of function evaluations is desired which is needed to reach the best possible solutions with the current setting/parametrisation while not wasting resources with any function evaluations thereafter. This in particular holds in industries where one function evaluation is extremely expensive in terms of time or other resources. In real-world optimisation nearly all applications turn out to be of multi-objective nature, and most of the time the objectives are conicting. In addition to extreme solutions favouring only one objective, good alternative solutions are highly appreciated. This is one reason why evolutionary multi-objective optimisation algorithms (EMOA) became a state-of-the-art method for solving multi-objective optimisation problems. Based on the nature of evolutionary algorithms (EA) featuring a set of solutions, EMOA are able to approximate the whole Pareto front in only one optimisation run. However, these optimisation runs are very costly and normally
terminated after a _xed, prede_ned number of function evaluations.
Online convergence detection (OCD) enables the termination of an EMOA based on two basic properties that indicate convergence based on quality indicators of the received solutions at each generation of the optimisation run (e.g.

hypervolume indicator). Iteratively for a time window of a given number of generations backward from the current one, statistical tests are performed on the variance of the quality indicator values (_2-variance-test) and the slope of the trend of these values (t-test on the regression coe_cient). OCD terminates the optimisation run if either the variance or the regression test indicates convergence for all quality indicator values. We tested OCD on mathematical test cases ([2]) and real-world applications from aerodynamics ([1]). OCD was able to save a signi_cant number of function evaluations while preventing a critical deterioration with respect to all quality indicators.

References

1. B. Naujoks and H. Trautmann. Online convergence detection for aerodynamic applications. In A. Tyrrell et al., editors, IEEE Congress on Evolutionary Computation 2009, CEC 2009, pages 332{339. IEEE press, US, May 2009.
2. Tobias Wagner, Heike Trautmann, and Boris Naujoks. Ocd: Online convergence detection for evolutionary multi-objective algorithms based on statistical testing. In Matthias Ehrgott et al., editor, Evolutionary Multi-Criterion Optimization, 5th International Conference, EMO 2009, Nantes, France, April 7-10, 2009. Proceedings, volume 5467 of Lecture Notes in Computer Science, pages 198{215. Springer, 2009.

* * * * *