INTRODUCTION

The Chemistry-Information-Computer (CIC) division of the German Chemical Society (GDCh) [1] invited the chemoinformatics and molecular modelling community to the 9th German Conference on Chemoinformatics (GCC2013) to discuss recent developments in the field of Computational Chemistry. The conference was held from the 10th to the 12th November 2013 in Fulda, Germany.

It was the first time for the current CIC board, who took office in the beginning of 2013, to organize the German Conference on Chemoinformatics. The board decided to adapt the conference to the recent developments in Computational Chemistry and the novel challenges the field is faced with. The session Chemical Information, Patents and Databases made room for the session Hot Topics and Developments, the scientific focus of other sessions was shifted, and a novel session, Research Telegrams, provided PhD Students an opportunity to present the current status of their research in oral presentations of 15 minutes each. Furthermore, new members were welcomed to the Scientific Advisory Board and the conference venue was moved from Goslar, Germany, where the conference has been held for the last eight years, to Fulda, Germany.

The 26 oral presentations covered Molecular Modelling and Drug Design, Chemoinformatics, and Materials Science. 62 posters were presented in two poster sessions. The more than 140 attendees from 20 nations demonstrated that the German Conference on Chemoinformatics is an internationally well-established event in the global Chemoinformatics and Modelling community.

The CIC-Award for Computational Chemistry 2013 for the best PhD thesis in the realm of Computational Chemistry was granted to Dr. Anne Mai Wassermann for her excellent PhD thesis “Computational Analysis of Structure-Activity Relationships – From Prediction to Visualization Methods”, which she worked on under the supervision of Prof. Dr. Jürgen Bajorath at the Rheinische Friedrich-Wilhelms-Universität Bonn, Germany [2].

ORAL PRESENTATIONS

O1
Putting together the pieces: building a reaction-centric electronic lab notebook for mobile devices
Alex Clark
1900 St. Jacques 302, Montreal H3J2S1, Canada
E-mail: aclark.xyz@gmail.com
Journal of Cheminformatics 2014, 6(Suppl 1):O1

The presentation will describe 4 years of work creating chemical structure based user interfaces for mobile devices, combined with creation of cloud-hosted webservises for supporting functionality. A number of products have been created along the way, for drawing structures and reactions, managing collections of data, searching databases, creating publication quality graphics, sharing and collaborating, calculating properties, and model building for drug discovery research. The focus will be on the assembly of these core technologies to create a fully featured electronic lab notebook (ELN) product for capturing chemical reactions.

While a number of web or mobile products already exist for laboratory notetaking, none of them provide the high level chemistry understanding and sophisticated cheminformatics functionality that will be described. The ability to combine all of the necessary capabilities onto a mobile device such as an iPhone or iPad demonstrates that this new generation of computing platforms is ready to play an important role in the realm of cheminformatics.

O2
Can quantum-chemical NMR chemical shifts be used as criterion for force-field development
Thomas E Exner1,2,*, Andrea Frank1, Heiko M Möller2, Martin Dračinský1
1Institute of Pharmacy, University of Tübingen, 72076 Tübingen, Germany;
2Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany;
3Institute of Organic Chemistry and Biochemistry, Academy of Sciences, 166 10 Prague, Czech Republic
E-mail: Thomas.Exner@uni-tuebingen.de
Journal of Cheminformatics 2014, 6(Suppl 1):O2

Fragment-based quantum chemical calculations based on our adjustable density matrix assembler (ADMA) are able to calculate NMR chemical shifts even for proteins and protein-protein complexes [1,2]. The agreement between the calculated and experimental chemical shifts in these calculations is, however, highly dependent on including conformational sampling and explicit solvent molecules. On the one hand, ensembles

© 2014 various authors, licensee Chemistry Central Ltd. All articles published in this supplement are distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
taken from classical MD simulations are suited for $^{13}$C and $^1$H chemical shift calculations if polar protons are neglected [3]. On the other hand, input structures from a Car–Parrinello MD resulted in landmark improvements over calculations based on classical MD especially for amide protons, which are predicted too high-field shifted based on the latter ensembles [4]. The better results are caused by the solute–solvents interactions forming shorter hydrogen bonds as well as by the internal degrees of freedom of the solute. With the obtained accuracy and the possibility of identifying the structural reasons for discrepancies between the experimental and calculated data, NMR chemical-shift calculations are now a perfect tool for e.g. the validation of new, improved force fields.

References

O3
Facing the challenges of computational target prediction
Karen T Schomburg*, Matthias Rarey
Center for Bioinformatics, University of Hamburg, Hamburg, Germany
E-mail: schomburg@zbh.uni-hamburg.de
Journal of Cheminformatics 2014, 6(Suppl 1):O3

To which proteins does a compound bind? Is it selective or promiscuous? These are questions which can be answered by inverse virtual screening, which identifies potential targets for a molecule of interest. Experimental screening of a molecule on thousands of targets is costly and elaborate. Contrarily, structure-based computational methods are only limited by the availability of 3D structures, rendering them an important complement to experimental methods.

Our inverse screening method XxirT combines triangle descriptor matching [1] with a new ranking approach, considering a reference score for each pocket. A precalculated bitmap encoding of the descriptors and an efficient design of a database for 3D protein structures allows a rapid screening of thousands of protein-ligand complexes with a query compound.

Classical protein-ligand scoring functions are not capable of inter-target score comparison, since the absolute values are target-dependent. Therefore, in inverse virtual screening, a major problem is the design of a ranking scheme allowing the comparison of target scores with respect to one query compound [2]. A lack of available data for statistical evaluation of the ranking capability further complicates the task. Data sets mostly contain positive hits, e.g. binding affinities for one molecule to several proteins while lacking negative data points, such as ‘the molecule does not bind to this protein’. As a basis for statistical evaluation of our new inverse screening concept, we introduce a dataset consisting of a ligand set of

Figure 1(abstract I1) Participants of the 9th German Conference on Chemoinformatics (GCC2013), November 10–12, 2013 in Fulda, Germany.

Figure 2(abstract I1) The winners of the poster award and the CIC-Award for Computational Chemistry 2013: from left to right, Jens Kunze (poster award, Swiss Federal Institute of Technology, Zurich, Switzerland), Laura J. Kingsley (poster award, Purdue University, West Lafayette/IN, USA), Michael Reutlinger (poster award, Swiss Federal Institute of Technology, Zurich, Switzerland), Anne Mai Wassermann (CIC-Award for Computational Chemistry, dissertation prize; Novartis Institutes for Biomedical Research, Cambridge/MA, USA) and Thomas Engel (Chair of the GDCh CIC division).
approved drugs and the scPDB target database [3]. Drugs are well qualified for use in a method evaluation dataset, as they are well tested and had to pass selectivity standard tests to get approved. Therefore, their targets are comparably well-characterized, allowing a classification into true positives and true negatives. This approach is the first that evaluates a structure-based inverse screening method on a systematic statistical test.

References

O5
Visualization of chemical space for medicinal chemists
Peter Ertl
Novartis Institutes for BioMedical Research, CH-4056 Basel, Switzerland
E-mail: peter.ertl@novartis.com
Journal of Cheminformatics 2014, 6(Suppl 1):O5

One of the most common tasks that cheminformatics experts in pharmaceutical industry are facing practically daily is analysis and visualization of large collections of molecules. Typical areas, where this is needed are analysis and enhancement of company compound archive, analysis of high-throughput screening data, design of combinatorial libraries, chemogenomics analyses and many others. But also researchers in academia are facing similar challenges when analyzing large public molecular databases that become available recently or even structures generated in silico. This presentation will provide overview of various methods used to analyze and visualize chemical space with particular focus on needs of medicinal chemists. When displaying results, for chemists it is of great importance that the molecules are represented by their actual structures, or at least by their scaffolds and not only by points as it is common in other scientific fields. This particular requirement makes chemistry visualizations challenging because of necessity to squeeze a lot of information on rather limited computer screen real estate. In the presentation various chemistry visualization techniques will be discussed, starting from classical display of molecules as tables and grids, through visualization based on analysis of scaffold, up to advanced cheminformatics visualizations techniques recently developed at Novartis, such as a method for natural ordering or scaffolds or Molecule Cloud diagrams.

O6
Limits to molecular matched-pair analysis: the experimental uncertainty case
Christian Kramer1, Klaus Liedl2
Department of General, Inorganic and Theoretical Chemistry, University of Innsbruck, 6020 Innsbruck, Austria
E-mail: Christian.Kramer@uibk.ac.at
Journal of Cheminformatics 2014, 6(Suppl 1):O6

Matched-Molecular Pair (MMP) analysis has recently emerged as a data analysis technique in medicinal chemistry. It quickly gained scientific momentum because it tackles key questions in lead optimization. In contrast to classical global QSAR models that attempt to predict the absolute numbers of ADME (absorption, distribution, metabolism, excretion) and toxicological properties, MMP analyses predict the difference in (bio-) chemical properties that can be expected due to small chemical modifications to lead structures, with a much smaller and well-controlled error than global QSAR models.

The power of MMP analysis depends on the number of previously documented similar molecular transformations, whereas the definition of chemical similarity plays a key role: the more generous the definition of similarity, the higher the variance, and thus the clearer the effect on ADME-Tox parameters, but also the less data pairs will be available [1].

The (bio-) chemical effect and the significance of the results depends on the experimental uncertainty (=noise) in the data. There is a clear mathematical association between the noise level and the minimum activity difference necessary for statistical significance. Here we demonstrate how the experimental uncertainty and variability [2,3] affect Matched Molecular Pair Analysis. It can be shown that for small sample sizes (Context-specific MMPs), the activity differences have to be very large in order to be statistically significant. A full equation for the estimation of minimum significant activity difference, depending on the number of samples, standard deviation of the measurements and the true variance of the biochemical effect is developed. The influence of consistency of assay setups can directly be quantified via the variability and practical consequences for MMP analysis will be presented.

References
Entropy in specificity and thermodynamics of binding
Klaus R. Liedl
Leopold-Franzens-University Innsbruck, Faculty of Chemistry and Pharmacy, Centre for Chemistry and Biomedicine (CCB), Innrain 82, 6020 Innsbruck, Austria
E-mail: Klaus.Liedl@uibk.ac.at
Journal of Cheminformatics 2014, 6(Suppl 1):C08

Entropy is an elusive and somehow non-intuitive concept. Nevertheless, entropy governs spontaneous thermodynamic processes as important contribution to Gibbs Free Energy. Information theory defines Shannon entropy as a measure for uncertainty. In the context of protein binding the inherent link between flexibility, thus conformational entropy, and substrate specificity is discussed. Substrate promiscuity of proteases is quantified as cleavage entropy correlating local binding site flexibility directly with substrate readout. Caspases are examined as example protease family, where active site dynamics play a major role in mediating substrate specificity. Direct comparison of entropy in substrate data allows highlighting previously unexpected similarities in substrate recognition in proteases. Promiscuous binding to several protease targets demonstrates the emerging importance of quantitative studies on binding specificity. Shannon entropy applied to probability densities is used to rationalize ordering or disordering by binding processes. We have developed a data-driven method to reconstruct probability densities from discrete sampling by computer simulations. Application to solvent degrees of freedom leads to excellent correlation with experimental data.

Quantum-mechanics-based molecular interaction fields for 3D-QSAR
Ahmed Elkerdawy1,*, Stefan Gussregn1, Hans Matter1, Matthias Hennemann1,2, Timothy Clark1,3,4
1Computer-Chemie-Centrum, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Nagelsbachstrasse 25, 91052 Erlangen, Germany; 2Sanofi-Aventis Deutschland GmbH (R&D), LGCR, Structure, Design and Informatics, Building G 878, 65526 Frankfurt am Main, Germany; 3Interdisciplinary Center for Medical Materials, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Nagelsbachstrasse 49, 91052 Erlangen, Germany; 4Centre for Molecular Design, University of Portsmouth, King Henry Building, Portsmouth PO1 2DY, UK
E-mail: Ahmed.Elkerdawy@fa.de
Journal of Cheminformatics 2014, 6(Suppl 1):O10

Computer-aided drug design (CADD) shift toward using quantum-mechanics (QM)-based approaches is not only the result of the ever growing computational power but also due to the need for more accurate and more informative approaches to describe molecular properties and binding characteristics than the currently available ones. QM-approaches do not suffer from the limitations inherent to the ball-and-spring description and the fixed atom-centered charge approximation in the classical force fields mostly used by CADD methods. [1] We introduce a protocol for shifting 3D-QSAR, one of the most widely used ligand-based drug design approaches, through using QM-based molecular interaction fields (MIFs) which are the electron density (p),

References
hydrogen bond donor field (HDF), hydrogen bond acceptor field (HAF) and molecular lipophilicity potential (MLP) to overcome the limitations of the current force-field-based MIFs (FF-MIFs). The average performance of the QM-MIFs (QMFA) models for nine data sets was found to be better than that of the conventional FF-MIFs models. In the individual data sets, the QMFA models always perform better than, or as well as, the conventional approaches. It is particularly encouraging that the relative performance of the QMFA models improves in the external validation (Figure 1).

Reference

GO with the flow and accessorize your drugs
Gisbert Schneider
Department of Chemistry and Applied Biosciences, Eidgenössische Technische Hochschule (ETH), Wolfgang-Pauli-Str. 10, 8093 Zürich, Switzerland
E-mail: gisbert.schneider@pharma.ethz.ch
Journal of Cheminformatics 2014, 6(Suppl 1):O11

The fast pace of drug discovery programs is supported by high-throughput screening campaigns to identify new chemical entities, where the underlying screening compound collections benefit from combinatorial libraries with lead- and drug-like properties. We will discuss recent developments in this field from two perspectives by addressing how to (i) assemble focused combinatorial libraries with desired properties (search problem), and (ii) identify targets for the designed compounds and, vice versa, compounds with desired target activities (scoring problem). Developments from our laboratory will be highlighted, specifically the MAntA (Molecular Ant Algorithm) approach for molecular building block prioritization, the target prediction method SPIDER (SOM-based Prediction of Drug Equivalence Relationships), and a quantitative Gaussian process model covering several hundred macromolecular targets for compound scoring. Using these algorithms we successfully generated activity-focused combinatorial small molecule and peptide libraries, de-orphaned de novo designed molecules, and correctly predicted off-target liabilities of known drugs. Selected examples of these practical applications will be presented, including an integrative application of combinatorial design, microfluidics-assisted synthesis, and activity testing.

Identification of host interactions for phenotypic antimalarial hits
Andreas Spitzmüller*, Jordi Mestres
Chemotargets SL and Systems Pharmacology Group, Research Programme on Biomedical Informatics (GRIB), IMIM Hospital del Mar Research Institute and Universitat Pompeu Fabra, Parc de Recerca Biomédica, Doctor Aiguader 88, 08003 Barcelona, Catalonia, Spain
E-mail: aspitzmuller@imim.es
Journal of Cheminformatics 2014, 6(Suppl 1):O12

Malaria is one of the most epidemic infectious diseases in the world affecting millions of patients and causing more than 500,000 deaths each year. Although there are several established antimalarials in clinical use, there is an urgent need for new drugs due to rapid resistance development. In recent years, more than 20,000 hits phenotypically active against P. falciparum, one of the major malaria causing agents, were disclosed from three independent HTS campaigns [1-3]. In order to make these hit libraries accessible to as many biological laboratories as possible, the MMV compiled and distributed the Open Access Malaria Box, a set of 400 chemically diverse active compounds [4]. One important task is now to elucidate the mode of action of those compounds. However, beside the parasite targets it is also necessary to identify potential host interactions in order to anticipate the risk of undesired side effects of those chemotypes at the earliest possible stage of development.

To this end, we applied a ligand-based virtual target profiling approach to predict possible interactions with human targets [5]. Amongst others, kinases and GPCRs were identified as the most important target classes. Subsequently, several hundred predicted interactions were selected for prospective experimental testing. Results showed that a substantial part of the Malaria Box exhibits the potential of interacting with human GPCRs.
To this extent, this was unexpected beforehand since the pathogenic agent does not contain any GPCRs. In this respect, particular attention was given to 5-HT2B receptor agonism, an effect associated to cardiac valvulopathy [6].

References


O13
Looking over the rim: algorithms for cheminformatics from computer scientists
Thorsten Meinl1, Bernd Wiswedel1, Michael R Berthold1,2
1KNIME.com AG, Zurich, 8005, Switzerland; 2University of Konstanz, Konstanz, 78457, Germany
E-mail: thorsten.meinl@knime.com
Journal of Cheminformatics 2014, 6(Suppl 1):O13

In recent years a number of methods were invented in the data mining/machine learning field that have received little attention in the cheminformatics world even though they offer interesting properties for these types of applications as well - even compared to some similar algorithms published primarily in the cheminformatics space. In this talk we want to highlight three of these algorithms/approaches. The first is MoSS [1], a frequent subgraph miner that can not only be used to find common substructures in a set of molecules but is also able to compute the MCSS very fast and has some extension especially suited for molecules. The second presented approach deals with the problem of finding diverse subsets of molecules [2]. Quite interestingly, not only finding a diverse subset can be a challenging task but already the definition of diversity is not as straightforward as it seems at the first glance. The third algorithm goes along the same lines but tries to find similar molecules by looking at their properties from so-called parallel universes[3]. Each universe contains a set of related properties and partial predictive models are built in each universe separately. Through interactive model construction, e.g. by so-called Neighbourgrams, the models from one universe can aid the construction of a models in other universes.

References


O14
Structure-activity relationship analysis on the basis of matched molecular pairs
Anne Mai Wassermann
Novartis Institutes for Biomedical Research, Cambridge, MA, 02139, USA
E-mail: anne_mai.wassermann@novartis.com
Journal of Cheminformatics 2014, 6(Suppl 1):O14

Matched molecular pairs (MMPs), i.e., pairs of compounds that are related to each other by a specific molecular transformation, have become an integral tool of drug discovery [1,2]. Generally spoken, matched molecular pair analysis (MMPA) aims at the extraction of all MMPs from a set of compounds and their association with calculated or measured property changes. Using public bioactivity data, we have used MMPs as a consistent reference framework to identify sets of chemical replacements that either have the propensity to induce large-magnitude potency changes or tend to retain compound potency across diverse targets [3,4]. Furthermore, we have extended the concept of MMPs to matched molecular series, i.e., analog series with different molecular core structures but corresponding substitution patterns [5,6]. The identification of series with alternative core structures but similar SAR trends is highly relevant for lead optimization where SAR information from one series that has been explored historically is ideally used to guide compound design efforts for a new chemotype [6].
References

O15
Combining pharmacophore- and MD-based modelling for phase II metabolism prediction
Christin Rakers1, Gerhard Wolber2
1Computer-Aided Drug Design, Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, Freie Universitat Berlin, 14195 Berlin, Germany
E-mail: christin.rakers@fu-berlin.de
Journal of Cheminformatics 2014, 6(Suppl 1):O15

As metabolism is considered a main cause for adverse drug reactions and failures of new drug candidates, our goal is to establish an in silico method to efficiently predict phase II metabolism – in particular sulfotransferase (SULT) activity. Since sulfotransferases exhibit low substrate specificities caused by their high degree of conformational freedom [1], activity prediction is a challenging task.

We therefore established a workflow based on molecular dynamics (MD) simulations to cover the whole spectrum of structural flexibility and incorporated it into multiple pharmacophores that represent specific modes of action. Using an ensemble of pharmacophores for virtual screening ensures accurate categorization of potential SULT ligands (e.g. substrates, inhibitors). Recent advances in MD technology [2] allowed for refinement of these pharmacophores by high-throughput MD simulations of ligand-target complexes. In addition, the initial binding process of a soluble ligand to the substrate-binding site of SULT was captured in unbiased 100 ns simulations using the software Desmond [3].

References

O16
Analysis and visual summarization of molecular dynamics simulation
Fredrick Robin Devadoss1, Victor Paul Raj2
1Department of Chemistry, University of Konstanz, 78457 Konstanz, Germany
E-mail: fredrick.devadoss@uni-konstanz.de
Journal of Cheminformatics 2014, 6(Suppl 1):O16

Molecular dynamics (MD) simulation, a standard technique used to study the dynamical properties of biomolecules, is very useful in collecting the trajectories, a series of snapshots – the coordinates of the system - of larger systems for longer simulation times. These MD generated trajectories are huge in size (many gigabytes) and the data analysis may take much longer time than the data generation. Managing the large amount of data and presenting them in a flexible and comprehensible manner are the major challenges. Analyzing these trajectories with standard parameter like root-mean square deviation (RMSD) may not reveal the most interesting properties of the dynamics.

To overcome these challenges, Cα torsion angles [1] - torsion angles build by four consecutive Cα atoms - are highly valuable as similarity measure on a substructure scale and to find major events - the information on the time at which a transition occurs (temporal domain) and the local structural changes (spatial domain) of it is combinedly called as “event”- occurring in the course of the MD simulation. By calculating the time series of the Cα torsion angles and their clustering it is possible to determine the mechanistic details on a residual length scale and find major events occurring in the simulation of large proteins or protein complexes. The main advantage of the Cα torsion angle criterion is that it does not depend on a previous alignment of the structures, and that the direction of the change is also defined. Heat maps of Cα torsion angle give nice graphical representations of processes described by the MD simulations. Clustering of snapshots according to the specific Cα torsion angles is used to automatically find the spatial domains of the structural changes. If all the snapshots are assigned to a single cluster, then those residues are considered as rigid core and the remaining residues are considered as flexible parts. The temporal domain can be characterized in more detail by finding continuous time intervals assigned to a single cluster as (meta) stable structures and time intervals where the assignment jumps between two clusters as transitional periods. Since the outliers can be removed from the fuzzy clusters, starts and ends of time patches now qualify as important events for the underlying substructure and structural changes of larger regions are caused by an accumulation of such substructure events.

DNA polymerase I – the open ternary complex of the large fragment of Thermus aquaticus DNA polymerase I (Klenaq1), which is used here as a practical application for Cα torsion angle based analysis, shows a hand-like arrangement, including a thumb, a palm and a finger domain [2]. The catalytic cycle leading to nucleotide insertion comprises several steps including a large structural rearrangement in the form of a movement of the finger domain towards the thumb domain, i.e. the transition from the open to the closed form. Molecular dynamics simulation were performed using the AMBER 10 suite of programs [3]. To get the visual picture of the ongoing processes, the Cα torsion angles with the differences to the crystal structure of the open form were plotted as heat map. The rigid and the flexible parts were clearly seen with no or a large number of significant changes, respectively, from the heat map. Once the Cα torsion angles corresponding to the rigid parts are removed, the remaining regions change only in a specific time interval of the simulation. The spatial and temporal domains of the structural changes were identified automatically by clustering of snapshots (using KNIME [4]) and finding the continuous time intervals, respectively.

References
4. KNIME: The Konstanz Information Miner. [http://www.knime.org/].
halogenation leads to a favorable halogen bonding interaction with the binding site. A scoring function, derived from a myriad of QM-calculations on the MP2 TZVPP-level of theory, evaluates the quality of the assumed interaction. To assess the prediction accuracy of our scoring function, we initially chose 50 examples from an exhaustive PDB scan randomly, representing the full range of equally distributed halogen bonding scores. For each file, the geometry was then recreated using iodobenzene representing the ligand system and N-methylacetamide as the backbone structure. Thus, the halogen bonding geometry of the database example was transferred as accurately as possible to our small model system and the interaction was calculated as an MP2 single point. The resulting energy was normalized and plotted against the predicted halogen bonding score. With only a few exceptions, most deviations were below 10% leading to an overall $r^2$ of 0.87. This highlights that our present scoring function is a blueprint for integration into general empirical scoring functions, which at present ignore halogen bonding interactions. Hence, recognition of halogen bonding will be implemented soon into protein-ligand docking and scoring.

References

O18
inSARA: intuitive single-target (large-scale) SAR interpretation and multi-target cross-reactivity analysis
Sabrina Wollenhaupt*, Knut Baumann
Institut für Medizinische und Pharmazeutische Chemie, Technische Universität Braunschweig, Beethovenstr. 55, 38106 Braunschweig, Germany
E-mail: s.wollenhaupt@tu-bs.de
Journal of Cheminformatics 2014, 6(Suppl 1):O18

inSARA (Intuitive networks for Structure-Activity-Relationships analysis) was primarily developed with the objective to support the medicinal chemist in tackling SAR analysis and visualization of large data sets in a more intuitive way than fingerprint-based approaches [1]. The method takes advantage of the synergic combination of the reduced graph (RG) and the maximum common substructure (MCS) concept [2]. The main feature of the inSARA concept is a hierarchical network structure of clearly-defined substructure relationships based on common pharmacophoric features. Thus, straightforward SAR interpretation is possible by interactive network navigation. When focusing on a set of active molecules at one single target, the resulting inSARA networks were shown to be valuable for various essential tasks in SAR analysis, such as the identification of activity cliffs or activity switches, biospecific replacements or SAR hotspots. Based on the identification of nearest neighbours in the networks, the prediction of bioactivities is also possible. Furthermore, inSARA can be used to investigate similarities between different targets. Targets are compared based on the overlap of common pharmacophoric pattern (RG MCSs) of the corresponding inSARA networks. According to the similar property principle, similar ligands are expected to bind to similar targets [3]. Therefore, this ligand-based analysis not only revealed meaningful similarity relationships between the analysed targets but is also beneficial for the detection of potential off-target relationships and cross-reactivities. Especially when investigating targets where no structural information is available but a set of active ligands is known (e.g. GPCRs), this complementary approach can provide important knowledge for drug design.

References

O19
Simulating “soft” electronics
Tim Clark
Computer-Chemie-Centrum, Department Chemie und Pharmazie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nagelsbachstraße 25, 91052 Erlangen, Germany
E-mail: tim.clark@fau.de
Journal of Cheminformatics 2014, 6(Suppl 1):O19

Simulating electronic devices built from flexible organic molecules requires both adequate conformational sampling and a reliable quantum mechanical description of the electronics of the system. The former is best achieved using classical (force field) molecular-dynamics simulations, from which individual geometries (“snapshots”) can be used for subsequent quantum mechanical calculations. As the repeating unit in the classical simulations typically involves thousands of atoms, the quantum mechanical techniques must also be able to handle many thousands of atoms quickly and effectively on modern parallel hardware. The newly developed EMPIRE program has been used in such a scheme to perform simulations on self-assembled-monolayer field-effect transistors (SAMFETs). Calculations of the size needed require novel interpretation and post-processing techniques based on local properties calculated on grids, rather than the more traditional population analyses. The results of simulations using the techniques described above will be presented and the algorithms and parallelization strategies implemented to be able to calculate as many as 100,000 atoms on 1,024 cores.

O20
New insights on the interface between metal oxide and biosystem
Walter Langels,*, Susan Köppen,*, Wenke Friedrichs,*, Armin Marx,*, Bastian Öhler†
†Institut für Biochemie, Universität Greifswald, 17487 Greifswald, Germany,
‡Hybrid Materials Interfaces Group, Faculty of Production Engineering University of Bremen, 28359 Bremen, Germany
E-mail: langel@uni-greifswald.de
Journal of Cheminformatics 2014, 6(Suppl 1):O20
The interface between titanium dioxide and biological solution is relevant for the bioactivity of titanium implants. Adhesion phenomena between inorganic solids are of general interest for mineralization and bone formation processes. Such interfaces are complex systems with many constituents including the hydroxylated metal oxide support, hydrocarbon contamination, ionic water solution and a variety of biopolymers. Our computational methods comprise classical molecular dynamics with around 10^6 atoms in the 10-100 ns range and ab initio molecular dynamics wit 100-200 atoms for some ps.

- Electronic structure calculations show that even very thin layers of TiO_2 on Ti metal may be commensurate and crystalline [1]. A variety of small- and rough-surface are implemented in force field simulations.
- The oxide is hydroxylated and its surface charge density correlates with the pH-value. The highly hydrophilic TiO_2 is in practice screened by hydrocarbons, which may enhance inflammatory complications of implants [2]. Simulations on the nature of this contamination are presented.
- Sequence specific protein adsorption on inorganic surfaces was found in several experiments on inorganic surfaces, even though no key-lock mechanism is conceivable. We propose two effects [3]: (i) Contacts of single amino acid side chains to local charges in the surface have rupture energies, which sensitively depend on the electrostatic. The adhesion of appropriate double contacts is very strong exceeding simple hydrogen-bonding. (ii) Soft motifs of proteins easier attach to the surface than rigid helices or strands.
- Close to solids water has an ordered structure, which only slightly depends on the surface charge density. These layers hinder protein adsorption. A major difference to the bulk is the reduced water mobility there.

References

O22
Balancing selectivity vs stability using molecular dynamics and umbrella sampling
Jeremie Mortier1,2, Elisabeth K Nyakatura1, Markus Miittinen2, Carsten Baldauf2, Gerhard Wolfer1, Beate Kocks1
1Department of Biology, Chemistry and Pharmacy, Freie Universität Berlin, Takustrasse 3, 14195 Berlin, Germany; 2Department of Physics, Freie Universität Berlin, Takustrasse 3, 14195 Berlin, Germany; 5Fritz Haber Institute, Faradayweg 4-6, 14195 Berlin, Germany E-mail: jeremie.mortier@fu-berlin.de
Journal of Chemical Informatics 2014, 6(Suppl 1):O22

Coiled coils are highly represented in biologically relevant macromolecules involved in important biological functions, such as gene expression regulation. The coiled coil environment has the great advantage to provide very well defined intermolecular recognition surfaces. The peptide system VPE-VPK is a rationally designed heterodimeric coiled coil structure [1,2]. The characteristic structure of the α-helical coiled coil allows randomizing the interaction partners of this dimeric system. Using a pool of VPE mutants that contains every possible combinations of the 20 canonical amino acids, specific binders could be searched empirically [1,3]. In this work, three key positions in the hydrophobic core were randomized in a VPE phage displayed library (Figure 1). This screen led to the identification of a novel core packing between VPE and VPK. One single consensus sequence was selected by the system, bearing a tyrosine in the hydrophobic core. Surprisingly, the dimer selected by phage display has a lower stability compared to the mother system. This important result raises the central question of selectivity vs stability. In order to address both aspects, theoretical investigations were conducted using molecular dynamics within the Gromacs suite. Pulling apart the two helices up to 3.00 nm from each other, potentials of mean force were calculated by umbrella sampling with a view to compare the energy barriers of the mother dimer to the phage display variant.

References

O23
Membrane simulation analysis using Voronoi tessellation
Gunther Lukat1,2, Jörn Sommer3, Jens Kugler4
1Bio-/Medical Informatics Department, Bielefeld University, Bielefeld, NRW, 33615, Germany; 2Theoretical Astrophysics Group, Hamburg Observatory, Hamburg, HH, 21029, Germany; 3Applied Bioinformatics, University of Tübingen, Tübingen, WB, 72076, Germany E-mail: gunther@CELmicrocosmos.org
Journal of Chemical Informatics 2014, 6(Suppl 1):O23

The study of membranes and embedded proteins represents an advanced task in the field of molecular simulation. While nowadays a profound selection of sampling techniques, molecular topologies and theoretical approaches is available, the analysis of actual simulation data remains a difficult endeavour. For homogeneous lipid bilayer simulations, the calculation of the bilayer thickness or area per lipid is directly accessible. For lipid mixtures, e.g. with cholesterol or embedded proteins, this is no longer the case. To face this challenge, APL@Voro has been developed [1]. The open-source and freely available graphical application is able to handle united-atom and coarse-grained trajectories generated with GROMACS [2]. Instructive, two-dimensional geometric representations of the lipid bilayer can easily be created based on Voronoi diagrams and Delaunay triangulations. The values, calculated on the geometric structures can be visualized in an interactive environment, plotted and exported to different file types (see Figure 1).
Even phase transitions within a bilayer can be tracked and visualised in an instructive and convenient way. APL@Voro represents a major improvement for the analysis of complex membrane simulations. The application is available at http://www.aplvoro.org.

References

O24
Use of DEKOIS 2.0 to gain insights for virtual screening
Frank M Boeckler, Matthias R Bauer, Tamer M Ibrahim, Simon M Vogel
Department of Pharmacy & Biochemistry, Eberhard Karls University, Tuebingen, 72076, Germany
E-mail: frank.boeckler@uni-tuebingen.de
Journal of Cheminformatics 2014, 6(Suppl 1):O24

With DEKOIS we have created an automated workflow to efficiently generate decoy sets based on a certain number of actives for any target [1]. Physicochemical similarity should be maximized between decoys and actives in order to yield challenging sets for benchmarking, while exact mimicking of potentially active substructures should be avoided to omit latent actives in the decoy set (LADS). Overall, the diversity of actives and decoys should be maximized to avoid artifacts based on clusters. Applying this philosophy, we have added more details to describe the physicochemical space and applied this protocol to generate sets for targets which had not been accessible before. These DEKOIS 2.0 sets [2] are available online (http://www.dekois.com) for benchmarking and development of new tailor-made scoring functions. Further extension toward additional targets can facilitate a systematic comparison of the virtual screening performance of docking tools and scoring functions in a target dependent way. Based on DEKOIS 2.0, we have investigated strengths and weaknesses of popular docking tools and scoring functions. In addition, we assessed how differences in the setup and preparation can affect the pROC profiles and early enrichment, in particular [3]. For this analysis we have developed an automated protocol to identify and highlight chemotype-specific differences in ligand recognition.

Figure 1(abstract O22) VPE-VPK represented as (a) a helical wheel diagram and (b) a ribbon diagram, with randomized positions in yellow, and the directly interacting position in green.

Figure 1(abstract O23) Pentamere of Vpu from HIV-1 in a POPC Bilayer. The Voronoi cells for the monomeres are colored, the cells for the lipids are left in white.
References

O25
In silico polypharmacology: retrospective recognition vs. rational design
Ewgenij Proschak
Institute of Pharmaceutical Chemistry, Goethe-University of Frankfurt, Frankfurt, D-60438, , Germany
E-mail: proschak@pharmchem.uni-frankfurt.de
Journal of Cheminformatics 2014, 6(Suppl 1):O25

The “one drug – one target – one disease” paradigm in drug discovery has been reconsidered during the last decade. This paradigm change was mainly caused by high attrition rates in drug approvals due to toxicity and lack of efficacy. Computational techniques play an important role in prediction and recognition of novel targets for approved drugs. We will discuss two machine learning approaches – self organizing maps and inverse distance weighting – for polypharmacological profiling of bioactive compounds, exemplified by two prospective studies [1,2]. While the recognition of occasional polypharmacological behavior is an established task, the rational design of multitarget ligands remains challenging. Dual or multi-target ligands have several advantages compared with selective compounds, including improved efficacy and more simple pharmacokinetic and pharmacodynamic properties in comparison to the combination of several drugs. In this context we present two in silico approaches to design dual inhibitors of 5-lipoxygenase (5-LO) and soluble epoxide hydrolase (sEH). The first study contains the discovery of a benzimidazolo-based dual 5-LO/sEH inhibitor by means of in silico screening [3]. The strategy of the virtual screening protocol was an exhaustive pairwise evaluation of pharmacophore models for both targets to obtain a dual-target pharmacophore model. Our second study deals with the development of a fragment based strategy for dual-target drug discovery. Here, we applied a modified self-organizing map algorithm for in silico recognition of molecular fragments binding both targets. The predicted properties were confirmed by complementary screening techniques: STD-NMR and enzyme assay. The enlargement of the fragment hit led to submicromolar dual target inhibitor of sEH and 5-LO[4].

References

O26
KRIPo – a structure-based pharmacophores approach explains polypharmacological effects
Tina Ritschel1*, Tom Ji Schirr2, Frans GM Russel3,4
1Computational Discovery and Design (CDD): Group at the Centre for Molecular and Biomolecular Informatics (CMBI), UMC, Nijmegen, The Netherlands; 3Department of Pharmacology and Toxicology, UMC, Nijmegen, the Netherlands; 4Centre for Systems Biology and Bioenergetics, UMC, Nijmegen, the Netherlands
E-mail: t.ritschel@umcn.ru.nl
Journal of Cheminformatics 2014, 6(Suppl 1):O26

“The most fruitful basis for the discovery of a new drug is to start with an old one” is a citation from Sir James Black’s Nobel laurate (1988). The background of this statement lies in the fact that most drugs are able to bind to multiple protein targets in the human body, this is known as polypharmacology. This behaviour can lead to unwanted side effects, and innovative research to avoid such adverse properties is of great importance. Paradoxically, polypharmacology can also be used to create new therapeutic approaches, as the protein to which a drug binds causing a side effect in one case, can be the main target for another treatment. Many cases report about the problems and opportunities of polypharmacology.

Aims: In order for a drug to bind to multiple targets, the interaction sites of these targets must be similar on a molecular level. Using KRIPo (Key Representation of Interaction in POckets) [1] with specially developed pharmacophore fingerprints, we provide an objective method to accurately describe protein interactions.

Results: KRIPo was used to explain the molecular mechanism of adverse drug effects of HMG-CoA reductase inhibitors, better known as statins. A previously unknown binding site for statins in cytochrome b, the major subunit of mitochondrial complex III of the oxidative phosphorylation system, was predicted by KRIPo.

Conclusion: Combining docking studies with KRIPo and experimental data on complex III inhibition enabled us to explain the molecular details of statin binding to the predicted binding site.

Reference

POSTER PRESENTATIONS

P1 Determination of selected cetyltrimethylammonium halide parameters by molecular modeling. Study of their adsorption on montmorillonite Bekhelifa Leila, Zizi Zahia1, Hallouch Mustapha, BenAhmed Abdelatif2 Laboratory of Materials and Catalysis, Chemistry Department, Faculty of Sciences, Hai larbi Ben M'hidi BP 89 University of Sidi- Bel – Abbès; SIDI BEL ABBES, 22000, Algeria
E-mail: z_zahia@yahoo.com
Journal of Cheminformatics 2014, 6(Suppl 1):P1

There are new chemical substances that increase the strength of washing products such as surfactants. However, these products are not biodegradable and can all be found in surface water, groundwater. Cationic surfactants are one of the pollutants of greatest danger to man is his environment. Therefore, we have determined some parameters of three cationic surfactants by molecular modeling, using the CHEM 3D bio. We, therefore, optimized geometries and minimized the energy of cetyltrimethylammonium chloride (CTAC), the cetyltrimethylammonium bromide (CTAB) and cetyltrimethylammonium iodide (CTAI) by molecular mechanics. We calculated the size of these molecules and their molecular partition coefficients and their polarities using molecular dynamics. The results showed that the three pollutants are adsorbed on Mt (natural), Mt (Mg) and Mt (Ca) and are not adsorbed on Mt (Na). Mt (K) can adsorb only CTAC. The values of the partition coefficient of the three surfactants are equal to unity. The cationic surfactants studied are highly hydrophilic and spend very little through a membrane. Al so, these values showed that these surfactants are much more soluble in water than in octanol. Both surfactants, CTAB and CTAI have the highest polarity values. They can therefore more easily approach a spherical or cylindrical form. CTAC has the lowest polarity, it presents more difficulty to approach a spherical or cylindrical shape.

P2 Theoretical studies on cycloaddition reactions
Lydia Rhyma1, Ponnadurai Ramasami1, Luis R Domingo2
1Computational Chemistry Group, Department of Chemistry, Faculty of Science, University of Mauritius, Réduit, Mauritius; 2Departamento de Quimica Orgánica, Universidad de Valencia, Dr. Moliner 50, 46100 Burjassot, Valencia, Spain
E-mail: lyd.rhyman@gmail.com
Journal of Cheminformatics 2014, 6(Suppl 1):P2

Cycloaddition reactions represent one of the most powerful processes in organic chemistry. The most common types of cycloaddition reactions are the Diels-Alder (DA) and the 1,3-dipolar cycloaddition reactions (1,3-DCs) which lead to five and six membered rings, respectively.

In our ongoing efforts contribute to the understanding of DA and 1,3-DCs; we studied the following using the B3LYP/6-31G(d) level of theory:

1. The 1,3-DCs of the pyridinium-3-olates and pyrazinium-3-olates with methyl acrylate and methyl methacrylate [1,2].
2. The competitive hetero-DA and 1,3-DCs of methyl glyoxylate oxime and its tautomeric nitrone with cyclopentadiene in the absence and in the presence of BF₃ as a Lewis acid catalyst [3].
3. A systematic study on the 1,3-DCs of C≡C with substituted nitrile oxides (RCNO; R = F, Cl, Br, NC, CN and NO₂) [4].

Among the outcomes of our investigations is the successful use of theoretical methods to understand the regio- and stereoselectivity of the reactions considered. It is expected that experimentalists find the results useful for synthesis involving these moieties and cycloaddition reactions. This presentation will overview our ongoing research program to have more understanding of these cycloaddition reactions.

References

The MOGADOC database (Molecular Gas-Phase Documentation) has grown up to 11,500 inorganic, organic, and organometallic compounds, which were studied in the gas-phase mainly by microwave spectroscopy, radio-astro-nomy, and electron diffraction. The database contains about 9,000 numerical datasets with internuclear distances, bond angles and dihedral angles. Most of the corresponding molecular structures are also given as 3D presentations (ball-stick-models). The retrieval features of the HTML-based database have been described elsewhere in the literature [1,2]. Some years ago a Java-applet has been developed, which enables the 3D-visualization of the molecular structures. The user can interactively rotate, shift and scale the 3D-models and can “measure” bond lengths as well bond, dihedral and elevation angles [3].

Recently new “measurement” features have been supplemented (such as for distances between centroids, angles between ring planes, etc.). The project has been supported by the Dr. Barbara Mez-Starck Foundation.

References
areas of chemistry. This poster presentation will describe the current technical state of the InChI algorithm and how the InChI Trust is working to assure the continued support and delivery of the InChI algorithm.

P5

Compound optimization through data set-dependent chemical transformations
Antonio de la Vega de León, Jürgen Bajorath
Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dahlmannstrasse 2, D-53113 Bonn, Germany
Journal of Cheminformatics 2014, 6(Suppl 1) P5

Matched molecular pairs (MMPs) have previously been used to extract chemical transformations and study their effect on molecular properties such as activity [1,2]. Chemical transformations have been used to direct compound optimization efforts towards defined activity profiles [3]. Here we introduce a methodology to assess effects of chemical transformations based on MMPs 2014, of compounds active against specific targets. The effects of selected chemical transformations on drug design-relevant molecular properties were analyzed. For different data sets, transformations that were frequently found and induced favorable property changes were identified. These transformations were then iteratively applied to modify active compounds and move them into favorable regions of ADME-relevant property space. Activity of newly designed compounds was tracked using nearest-neighbor search in ChemBL. The results of our study indicate that activity-conservative data-set dependent transformation can aid in the design of new active compounds with favorable ADME characteristics [4].

References

P6

Impact of binding site waters on inhibitor design: contemplating a novel inverse binding mode of indirubin derivatives in DYRK kinases
Daniel Cappel1, Vassilios Myrianthopoulos2, Emmanuel Mikros2, Woody Sherman3
1Schrödinger GmbH, Dynamostraße 13, 68165 Mannheim, Germany;
2University of Athens, Panepistimiopolis Zografou, 15771 Athens, Greece;
3Schrödinger Inc., 120 West 45th Street, 17th Floor, New York NY 10036, USA
E-mail: daniel.cappel@schrödinger.com
Journal of Cheminformatics 2014, 6(Suppl 1) P6

DYRK kinases are involved in alternative pre-mRNA splicing as well as in neuropathological states such as Alzheimer’s disease and Down syndrome. In this study, we present the design, synthesis, and biological evaluation of indirubins as DYRK inhibitors with enhanced selectivity. Modifications of the bis-indole included polar or acidic functionalities at positions 5 and a bromine or a trifluoromethyl group at position 7, affording analogues that possess high activity and pronounced specificity. Compound 6i carrying a 5-carboxylate moiety demonstrated the best inhibitory profile. A novel inverse binding mode, which forms the basis for the improved selectivity, was suggested by molecular modeling and confirmed by determining the crystal structure of DYRK2 in complex with 6i. Structure–activity relationships were further established, including a thermodynamic analysis of binding site water molecules, offering a structural explanation for the selective DYRK inhibition [1].

Reference

P7

Charge-related topological index – various chemoinformatics applications
Nikolay T Kochev1, Ivan Bangov2, Emil Petrov3, Marina Moskovkina4, Borislav Stoyanov5
1University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry, University of Plovdiv, 24 Tsar Asen St., Plovdiv, Bulgaria; 2Faculty of Natural Sciences, Konstantin Preslavski Shumen University, 115 Universitetska Str., Shumen, Bulgaria; 3Department of Computer Informatics Faculty of Mathematics and Informatics, Konstantin Preslavski University of Shumen, 115 Universitetska Str., Shumen, Bulgaria
E-mail: nick@uni-plovdiv.net
Journal of Cheminformatics 2014, 6(Suppl 1) P7

We present several useful applications of the CTI index in the context of various chemoinformatics tasks. Charge-related Topological Index (CTI) was introduced initially by Bangov for solving the problem of 2D structure isomorphism within the computer-assisted structure generation from a gross formula [1]. CTI is a real number defined as a sum over all atom pairs: 

\[ CTI = \sum_{i,j} L_i J_i D_{ij}^{-1} \]

where \( D_{ij} \) is the inter-atomic topological distance and \( L_i = L_{0,i} - N_{H,i} + q_i \) is a local atom index (a characteristics of atom \( i \)) calculated from the atom valence, \( L_{0,i} \), the number of hydrogen atoms attached to atom, \( N_{H,i} \), and \( q_i \), which is the corresponding charge density. The partial charges are computed by the topological empirical method of Gasteiger-Marsili [2] calculated with a fixed number of iterations. The CTI index could be used for 3D structure and conformation perception [3] in the following form:

\[ CGI = \sum_{i,j} L_i L_j R_{ij}^{-1} \]

where \( R_{ij} \) is the geometrical distance (in the latter case the charges could be also obtained from quantum chemistry programs on a semi-empirical level). Respectively, \( L \) values could be employed to the perception of the structure symmetry and topological equivalence. In our latest study we proved the capabilities of CTI index for perception of duplicated structures in large structure collections [4]. It is shown that the CTI index can be safely used for a quick perception of the duplicated structures in large databases in a very fast identity (full structure) search. CTI index with a precision of 10 digits after decimal point can be used in databases with millions of compounds. It has been also shown that CTI can be used as a useful descriptor well describing both the structure branching and some electronic properties [3].

References

P8

The integration of Open3DTOOLS into the RDKit and KNIME
Paolo Tosco1, Nikolaus Stefl2, Greg Landrum3
1Department of Drug Science and Technology, University of Turin, Torino, I-10125, Italy; 2Novartis Institutes for Biomedical Research, Basel, CH-4002, Switzerland
E-mail: paolotosco@unito.it
Journal of Cheminformatics 2014, 6(Suppl 1) P8
Open-source software is especially appealing to academics, since it permits implementing novel methods into existing code with little effort, while allowing full disclosure of the underlying science; the lack of license fees and the ease of dissemination via public repositories are additional plusses.

However, open-source has also been growing popular within large companies, which have recognized the value of code sharing to increase the pool of users (and therefore testers and reviewers), as well as to gather new ideas and contributions. In the pharmaceutical field, an outstanding example is represented by the RDKit [1], a BSD-licensed C++ cheminformatics toolkit with Python, Java, and C# bindings, originally developed at Rational Discovery and currently used and being actively developed within the Novartis Institutes for BioMedical Research, which contributes in-house code enhancements back to the open-source version.

To enable effective and widespread adoption and usage of open-source tools, it is helpful to make them available as plugins to well-recognized platforms. Herein we describe the integration of Open3DTOOLS (namely, Open3DQSAR [2] and Open3DALIGN [3]) into the RDKit and KNIME [4]. This task required the preliminary implementation and validation of the MMFF94 (s) force field, upon which the Open3DTOOLS are based, in the C++ layer of the RDKit, followed by the extension of the RDKit API to enable molecular alignment, MIF computation and 3D-QSAR model building. Additionally, KNIME nodes were set up to allow access to Open3DTOOLS functionality within KNIME workflows.

We also present some test cases which illustrate the potential of these RDKit and KNIME extensions in the context of virtual screening and CADD.

References

P9
Molecular fragment dynamics study on the water-air interface behavior of non-ionic polyoxyethylene ether surfactants
Andreas Truszkowski1*, Annamaria Fiethen2, Hubert Kühn2, Thomas Wiebringhaus2, Achim Zielensky1, Matthias Epple2
1Inorganic Chemistry and Center for Nanointegration, University of Duisburg-Essen, Essen, 45141, Germany; 2CAM-D Technologies, Essen, 45117, Germany; 3Institute for Bioinformatics and Chemoinformatics, Westphalian University of Applied Sciences, Recklinghausen, 45665, Germany
Journal of Cheminformatics 2014, 6(Suppl 1) P9

Molecular Fragment Dynamics (MFD) is a mesoscopic simulation technique based on Dissipative Particle Dynamics (DPD). Whereas DPD beads in general may not necessarily be identified with chemical compounds at all, the MFD variant uses specific molecules or molecular fragments as its basic coarse-grained interacting entities (rather than the fine-grained atom types of Molecular Mechanics). MFD can be used to study formulations of drugs and active agents in oil, water and emulsions.

MFD simulations of the nonionic polyoxyethylene ether surfactants C6E6, C10E6, C12E6 and C16E6 at the water-air interface are performed to study their nanoscale structures and surface properties. The simulations of the self-aggregation of the polyoxyethylene ether surfactants lead to equilibrium nanoscale structures and computationally determined surface tensions which are in agreement with experimental data for different surfactant concentrations [1].

Reference

P10
Kernel density estimation of CSD distributions - an application to knowledge based molecular optimisation
Patrick McCabe1*, Oliver Korb1, Jason Cole1, Robin Taylor2
1CCDC, 12 Union Road, Cambridge, CB2 1EZ, UK; 2Taylor Cheminformatics Software, 54 Sherfield Avenue, Rickmansworth, Herts, WD3 1NL, UK
E-mail: mccabe@ccdc.cam.ac.uk
Journal of Cheminformatics 2014, 6(Suppl 1) P10

The Cambridge Structural Database (CSD) contains a large amount of molecular structure data (bond length, bond angle and torsion angle data.) Much of this data has previously been extracted in histogram form and provided in the Mogul program. Histograms however have several disadvantages e.g. they are not smooth, they depend on bin widths and bin end points.

Kernel density estimators do not bin data and have no end points but centre a kernel function at each data point and smooth kernel functions will generate smooth density estimates [1]. A difficulty of the approach though is how wide to make the kernel functions.

In this work kernel density estimation is used to generate probability density functions (pdfs) for bond length, bond angle and torsion angle histograms derived from the CSD. Gaussian kernels are used for bond length and bond angle data and a von Mises kernel is used for the torsion angle data [2]. The

Figure 1 (abstract P9)
resulting pdfs are smooth and are suitable for application to molecular geometry optimisation.

References

P11 A new software for fragment-based QSAR and its applications
Sergey B Sosnin, Vladimir A Palyulin, Nikolay S Zefirov
Department of Chemistry, Lomonosov Moscow State University, Moscow, 119991, Russia
E-mail: serg.sosnin@gmail.com
Journal of Cheminformatics 2014, 6(Suppl 1)P11

Fragment-based methods are quite popular in 2D QSAR/QSPR studies. In the advanced versions of these approaches for developing highly predictive models, one have to generate a huge set of descriptors that in turn requires well-designed algorithms and high-quality parallelism. To overcome these problems we developed the software for tagged generation of fragmental descriptors.

One of the most perspective programming paradigms is functional programming. Programs in pure functional languages are easy to parallelize, tend to be faultless, more clear, easy extensible [1]. Despite its tangible complexity for learning, functional programming is becoming rather popular, recently several cheminformatics frameworks: OUCH and chemf were presented. We developed the program QLab for the fragmentation of molecular graphs in the imperative programming language (Java, using CDK), and ported it under the name FragmentT into the functional language (Haskell). Aware of the fact that graph operations is a core of cheminformatics, FragmentT represents chemical structures by means of Functional Graph Library – a powerful tool for graph operations [2]. Using this software we processed several structure-activity/property databases containing hundreds of compounds and generated the sets of more than 500 000 fragmental descriptors in each case. Then for each database 100 most correlated with activity/property descriptors were selected using stagewise MLR. Based on selected descriptors the predictive models were developed using artificial neural networks (ANN) and Random Forest algorithms. Comparison with other fragmental descriptors generation software showed better predictive ability of our models.

References

P12 Simulation of the influence of oxygen on the chemical stage of radiobiological mechanism using Petri nets
J Barilla1, M Lokajíček”, H Piskaˇrová”, P Simˇe1
1 J. E. Purkinje University in Usti nad Labem, Faculty of Science, Czech Republic, Institute of Physics, Academy of Sciences of the Czech Republic Journal of Cheminformatics 2014, 6(Suppl 1)P12

The radiobiological effect of densely ionizing ends of primary or secondary charged particles may be influenced significantly by processes running in the chemical stage of radiobiological mechanism; especially the influence of present oxygen may be very important. The effect of its or of other species (radiomodifiers) present in water medium during irradiation may be studied with the help of corresponding mathematical models. The model based on the use of Petri nets will be proposed and described. Two parallel processes, i.e., diffusion of radicals and their chemical reactions, running in corresponding radial clusters formed during energy transfer may be represented with the help of the given model. A great number of chemical species may be easily taken into account. The model enables to study the concentrations of individual radicals changing during cluster diffusion and to estimate their damaging effects on corresponding DNA molecules in given cells. The results demonstrating the influence of oxygen under different concentrations will be presented.

P13 Evaluation of molecular model-based discovery of e5'-nucleotidase inhibitors on the basis of X-ray structures
Norbert Furtmann1, Jürgen Bajorath2
1 Department of Life Science Informatics, BHT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, Dattelnstrasse 2, D-53113 Bonn, Germany
2 Journal of Cheminformatics 2014, 6(Suppl 1)P13

Ecto-5'-nucleotidase (e5NT) belongs to the family of metallophosphoesterases, hydrolyses AMP to adenosine, and is a regulator of the adenosine signaling pathway [1]. It has been shown, that free adenosine is involved in various diseases and cancer progression [2,3]. In a previous study, a molecular model of e5NT has been created and used for the identification of new sulfonamide inhibitors [4]. Recently, X-ray structures of human e5NT in complex with different inhibitors were published [5]. This made it possible to reevaluate the model building and virtual screening efforts in detail. An extensive analysis of the comparative e5NT model, built using a bacterial enzyme in the presence of 35% sequence identity as a template, showed that the model was topologically correct and had high accuracy within the active site region. Comparative docking studies were carried out to explore inhibitor binding characteristics within the X-ray structure and the model. The results provided plausible explanations for the successful identification of new e5NT inhibitors by model-based virtual screening and highlighted important parameters [6].

References

P14 Fishing out the signal in polypharmacological high-throughput screening data using novel navigator cheminformatics software
Denis Fourches1, Alexander Trophsa
1 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 27598 USA
Journal of Cheminformatics 2014, 6(Suppl 1)P14

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 untreated output files generated by plate readers. We have developed the freely-accessible HTS Navigator software to enable and facilitate the processing and analysis of polypharmacological HTS data. We report on the capabilities of Navigator and present several case studies where we employed cheminformatics approaches embedded within the Navigator to curate and analyze large datasets of compounds tested toward different panels of targets. Examples include libraries of compounds tested for their inhibition potencies across several CYP450; or for their inhibition of multiple protein kinases; or their binding profiles against multiple GPCRs. We show how to quickly identify and highlight 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 and 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-Accross (CBRA) approach [1] also implemented in the Navigator to infer the activity of external compounds from both chemical (defined by chemical similarity) and biological (defined by the similarity of HTS profiles) analogues.

Reference

P15
Chemistry-wide association studies (CWAS) to determine joint toxicity effects of co-occurring chemical features
Yen Low, Alexander Sedykh, Denis Fourches, Alexander Trophsa*
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
E-mail: alex_trophsa@unc.edu
Journal of Cheminformatics 2014, 6(Suppl 1):P15

Individual structural alerts often fail to accurately predict chemical toxicity as they tend to overlook the moderating effects of other co-occurring alerts. Features are said to have statistical interaction effects when one changes or modulates the effect of another on the target property. Here we introduce Chemistry-Wide Association Study (CWAS; by analogy with GWAS in genomics) to systematically eliciting the individual and interaction effects of chemical features on the target property. A mutagenicity dataset of 5,439 compounds was used in this proof-of-concept study. We utilized QSAR models built with random forest and ISIDA fragment descriptors to select the most important chemical features and identify pairs of features with significant interaction effects. These interacting feature pairs revealed how subtle structural changes affect mutagenicity (e.g., ortho-substitution reduces mutagenicity caused by nitroaromatic moieties). We also found that feature pairs can be integrated into more specific structural alerts with fewer false positives. We believe the interaction effects uncovered by CWAS are useful for refining structural alerts and enhancing model interpretation, enabling more effective design of safe chemical substances and leading to the improved regulatory chemical risk assessment.

P16
Scaffold dependencies for halogen bonding: quantum mechanical investigation of nitrogen-bearing heterocycles
A Lange*, WO Zimmermann, MF Boeckler
Laboratory for Mol. Design & Pharm.Biophysics, Eberhard Karls University Tuebingen, Auf der Morgenstelle 8, 72076 Tuebingen, Germany
E-mail: andreas.lange@uni-tuebingen.de
Journal of Cheminformatics 2014, 6(Suppl 1):P16

Halogen bonding is a rather new but promising type of interaction for the drug discovery process. It is rather directional and involves an electron donor as binding partner. Employing quantum chemical calculations [1-3], we explore the applicability of halogen bonds in molecular design with respect to halogen-enriched fragment libraries (HEFlibs) [4]. Computational studies on protein-ligand complexes have been reported indicating varying halogen-bond strengths that depend on the chemical nature of the ligand and the surrounding binding pocket. The strength and behavior of a halogen bond certainly depends on the core scaffold to which the halogen is attached. Due to the relative novelty of halogen bonding in medicinal chemistry, only a few experimental studies exist [4-6]. Therefore, we computed 30 different nitrogen-bearing heterocycles substituted by chlorine, bromine, or iodine (in total 459 structures) to increase our understanding of scaffold-based tuning of halogen bonds (see Figure 1 below for a few examples). For each structure the electrostatic-potential isosurfaces will be plotted (3D and 2D) to illustrate changes in halogen bond strength and the appearance of binding motifs in certain substituted heterocycles. Furthermore, we aim at providing simplified rules and models to estimate halogen bonding strength based on the molecular electrostatic potentials. Our data highlights that the use of an appropriate scaffold exerts a most important influence on the applicability of a halogen bonding. References
3. Lange, et al., unpublished data.

P17
Introducing fuzziness into maximum common substructures for meaningful cluster characterisation
Christian Herhaus
Global Computational Chemistry, Merck Serono, Darmstadt, 64293, Germany
E-mail: christian.herhaus@merckgroup.com
Journal of Cheminformatics 2014, 6(Suppl 1):P17

Arranging similar structures in clusters is one of the typical tasks of modern Chemoinformatics with high impact in HTS follow-up, generation of structure activity relationships (SAR) and selection of starting points for compound optimisation. Methods for cluster generation are as diverse as the structures which they are applied to [1], may they be e.g. similarity- or substructure-based. Typically, medicinal chemists tend to orientate themselves in structure subsets like clusters with the help of substructures, so-called “scaffolds”, which intuitively characterise the structural relationships between the molecules of the subset. In the case of substructure-based clustering, well established methods are existing for the generation of Maximum Common Substructures (MCS) which are present in all members of the structure population or a defined proportion thereof [2]. But in the case of similarity-based clusters, such MCS may either not be existing for the required dataset proportion or the common substructure may be so small that it is no longer representative and therefore meaningless.

The approach presented here allows the generation of MCS also for similarity-based clusters with a given inherent structural diversity. It does so by generating an MCS of reduced graphs in a first step, followed by mapping atom and bond indexes of this reduced MCS onto the full structures and aggregation of atom and bond information for each indexed atom/bond. In a final step, query features of the MDL SDF format (atom lists, query bonds) are utilized to map aggregated element and bond information onto the reduced MCS. As a result, “fuzziness” in atom and bond information is added to the MCS which, although still being fully database-searchable, is more meaningful for the characterisation of clusters as it can cover larger parts of the full structures than a conventional MCS could do. The approach was implemented in Pipeline Pilot™ for proof of concept but is general enough to be transferred to other technical platforms as well.

References

P18
Fragment docking supported by NMR shift perturbations
Tim ten Brink*, Clementine Aquare, Isabelle Krimm
Institut des Sciences Analytiques – CNRS UMR5280, Universite Claude Bernard - Lyon 1, Villeurbanne, F-69622, France
E-mail: timten-brink@univ-lyon1.fr
Journal of Cheminformatics 2014, 6(Suppl 1):P18

Fragment-based approaches have become popular tool in drug design due to their ability to screen large portions of chemical space with comparatively small libraries. However fragments can exhibit unspecific binding and even if they bind to a specific binding site in some cases more than one binding mode is observed [1]. For computational approaches like molecular docking fragments pose also new challenges. Score differences between different
binding modes generated by docking are often small, making the identification of the correct, natural binding mode difficult. The sensitivity of a nuclei’s NMR chemical shift to changes in its chemical environment can be used to measure chemical shift perturbations (CSP) of protein atoms upon ligand binding. Especially 1H and 15N CSP are easily obtainable from 1H HSQC spectra and can be used as probe for ligand orientation but also include information about conformational changes on the protein side. CSP data has been used to orientated drug like molecules into protein binding sites [2,3] and can be included into the scoring function to improve docking [4]. Here we show how CSP can be used to quickly validate docking poses of smaller fragments by filtering them for their agreement between experimental CSP and simulated CSP for the docked poses. Additionally a more detailed analysis of the differences between the experimental and the simulated CSP profiles can be used to highlight protein regions and even single residues which undergo structural changes upon docking binding.

References

P19
Probing the impact of protein and ligand preparation procedures on chemotype enrichment in structure-based virtual screening using DEKOIS 2.0 benchmark sets
Tamer M Ibrahim1*, Matthias R Bauer1,2, Frank M Boedeker1
1Lab. of Molecular Design & Pharmaceutical Biophysics, University of Tuebingen, Tuebingen, 72076, Germany; 2MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK
E-mail: tamer.abdelrehim@student.uni-tuebingen.de
Journal of Cheminformatics 2014, 6(Suppl 1)P19

Structure-based virtual screening techniques can help to identify new lead structures and complement other screening approaches in drug discovery. Prior to docking, the data (protein crystal structures and ligands) should be prepared with great attention to chemistry-related molecular details. In all cases, a wide choice of commercially and non-commercially packages are available to perform such preparation schemes.

Using the DEKOIS 2.0 benchmark sets [1,2], we found differences in the respective virtual screening performance when employing different preparation schemes. We demonstrate how docking performance, particularly early enrichment, can be affected by these differences. To investigate these interesting results, we have developed an automated protocol to match and visualize ligand chemotype information in combination with the pROC profile obtained by docking. We can utilize this new tool to identify and highlight chemotype-specific behaviour, e.g. in dataset preparation. This can help to overcome chemistry-related issues in virtual screening.

References

P20
HTS explorer
Christoph Müllera, Isabella Feierberg, Ola Engkvist, Christian Tychan AstraZeneca R&D, Discovery Sciences Chemistry Innovation Centre & RIA iMed Medicinal Chemistry, Mölndal, Sweden
Journal of Cheminformatics 2014, 6(Suppl 1)P20

In the early stages of a drug discovery project it is often necessary to narrow down the search space for potential new leads substantially [1,2]. This crucial step identifies a set of molecules (a hit series) that have a high likelihood of being relevant to the drug discovery project. In many cases high throughput screening (HTS) is used to test (in-vitro) large amounts of molecules against a biological target in order to validate a molecule’s potential to interact with the target and therewith its relevance to the drug discovery process. Since there are time and money constraints associated with such a process, it is not feasible to pipe the full HTS compound set through very detailed testing. Rather, an HTS process consists of several stages: a primary (spot test, SP) stage, a confirmation (retest) stage and a concentration-response (CR) stage. The later is the most resource-intensive
where compounds are tested at a range of different concentrations, which allows for curve-fitting and determination of the potencies.

The HTS Explorer enables HTS evaluators across AstraZeneca R&D sites to do a comprehensive and effective screening analysis for compound prioritization within a single tool. This tool is provided as an extension to the visualization platform TIBCO Spotfire [3]. The use of Spotfire as a platform streamlines the process for the HTS evaluator and facilitates distribution of the tool as sharing of the HTS data evaluations. Further new approaches and methods can be made easily and readily available across AZ. Data is retrieved by calling PipelinePilot protocols, various web-services and queries to in-house databases [4,5]. The main features include many different options for clustering of compounds as well as commenting on clusters, cluster visualization, prioritization and interactive reclustering. Further it enables compound annotation with e.g. the known Structure-Activity Relationship (SAR) space and offers different types of structural searches in internal and external databases.

References
3. [http://spotfire.tibco.com/].
5. [http://accelrys.com/products/pipeline-pilot/].

Meat quality prediction using Raman spectroscopy and chemometrics
Marius Nache1*, Rico Scheier1, Heiner Schmidt2, Bernd Hitzmann1
1Department of Process Analytics and Cereal Technology, Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, 70599, Germany; 2Research Centre of Food Quality, University Bayreuth, Kulmbach, 95326, Germany
E-mail: Nache.M@uni-hohenheim.de
Journal of Cheminformatics 2014, 6(Suppl 1):P21

The feasibility of using Raman spectroscopy as a fast and non-invasive method to monitor the quality parameters in pork meat has been investigated. For this application an online prediction methodology has not been established yet. Based on raw Raman spectra of 10 pork semimembranosus muscles a range of data pre-processing and multivariate calibration methodology have been used to develop online predictive models for the meat quality parameters: the lactate and pH. The linear and nonlinear algorithms studied were comparatively analysed for speed, robustness and accuracy, identifying the best “efficiency” evaluation procedure represented the final milestone of the present study. Thus with a cross-validated r2 value for both pH and lactate of 0.97, a RMSECV of 4.5 mmol/l for the lactate prediction and 0.06 units for the pH prediction, locally weighted regression provided the most accurate and robust model. This prove the feasibility of using Raman spectroscopy for online meat quality control applications.

P22
Exploring and cataloguing the substrate space of prenyltransferases: automatic generation of SMARTS
Jakub Gunera*, Peter Kolb
Department of Pharmaceutical Chemistry, Philipps-Universität Marburg, Marbacher Weg 6, 35032 Marburg, Germany
E-mail: jakub.gunera@uni-marburg.de
Journal of Cheminformatics 2014, 6(Suppl 1):P22

Structure-activity relationship (SAR) is one of the foundational principles of biomolecular activity and states that similar molecules have similar biological activity. This work is focused on the utilization of enzymatic reactions in...
P23 Interaction studies of Alzheimer's Cathepsin B protein with inhibitors in presence and absence of water
Nitin Chitranshi1,2*, Pushpendra K Tripathi1, Prahlad K Seth3
1Gautam Buddha Technical University, Lucknow 227202, Uttar Pradesh, India; 2Bioinformatics Centre, Biotech Park, Sector-G, Jankipuram, Lucknow 226021, Uttar Pradesh, India
E-mail: nitinchitranshi@gmail.com
Journal of Cheminformatics 2014, 6(Suppl 1) P23

Background: The accuracy of ligand-protein docking may be affected by the presence of water molecules on the surface of proteins. Water can form complex bridging networks and can play a critical role in dictating the binding mode of ligands. A recent analysis of high-resolution crystal structures of ligand-protein complexes revealed that 85% of the complexes had one or more water molecules bridging the interaction between ligand and protein. For predicting the binding modes and energies of protein-ligand interactions, molecular docking methods are commonly used. In order to obtain an accurate complex geometry and binding energy estimation, an appropriate method for calculating partial charges is essential. AutoDockTools software, widely used as interface for preparing input files, utilizes the either Gasteiger or Kollman partial charge calculation method for both protein and ligand charge calculations. However, it has already been reported that more accurate partial charge calculation and as a consequence, more accurate docking can be achieved by using quantum chemical methods. In common practice so far, the quantum chemical partial charges were applied to the ligands for docking calculations. The newly developed Mozyme function of MOPAC2009 allows fast partial charge calculation of proteins by quantum mechanical semi-empirical methods. Thus, in the current study, we use the semi-empirical quantum-mechanical partial charge calculations to investigate the interaction energies and polarization effects of the various components of the binding pocket on a set of Cathepsin B protein.

Results: The docking accuracy was computed by using the original AutoDock scoring function with the set of 19 protein ligand complexes using Gasteiger, AM1 and PM3 partial charge calculation methods. This helped us to compare the effect of the partial charge calculation method on docking accuracy. It was seen that the docking accuracy in regard to complex geometry significantly increased when partial charges of the ligands and proteins were calculated with the semi-empirical PM3 method. Our results demonstrate that (i) the energetic of the key water molecule are more favorable for the binding site in the Cathepsin B protein (ii) Water bridging and triangle formation were seen between the key amino acid residue and the ligand (iii) The internal energy is significant factor for the binding modes of various ligands. It was also observed a statistically significant overall increase in accuracy when water molecules are included during docking simulations. Out of the 19 complexes analyzed in the course of our study, the geometry of 17 complexes were accurately calculated

order to augment ligands with small apolar substituents. These modified molecules will then be employed for establishing the SAR of a ligand series. For this purpose, prenyltransferases (PT) seem like an attractive avenue. Their enzymatic attachment of the apolar prenyl moiety [1] (Figure 1) is desirable, as it does not require the precise geometric orientation of a polar interaction. In the initial phase of this project, PTs and their reactions were catalogued and digitalized: A MySQL-database was established where reactions – extracted from primary literature – are stored as SMARTS strings. These reaction patterns will be used for generating a library of molecules accessible to prenylation. In order to automatically generate SMARTS, the atoms involved in bond cleavage and formation have to be enumerated. Fragmentation of the substrate followed by atom mapping onto the product molecule makes it possible to identify the nucleophilic atom in an automated manner (Figure 2). From the reactive atom as starting point, an atomic pattern is reconstituted within a distinct distance. This reconstitution approach ascertained the connectivity of the atoms and allowed for ring closures. The resulting SMARTS string bears atom types that can be modulated to any generality level supported by the SMARTS definitions. Studies of over 60 publications and 400 catalogued reactions revealed a huge predominance of PTs of the DMATS superfamily (dimethylallyltryptophan synthases) and correspondingly of prenylated indole derivatives. The reactions can further be subdivided into classes of i) reverse and regular prenylation, ii) endo- and exocyclicity of the prenylation site, iii) aromaticity of the prenylation site and iv) element name of the nucleophilic atom.

Reference
1. Li SM. Natural Product Reports 2010, 27:57-78.
using PM3 partial charges, while the use of Gasteiger charges resulted in only 8 accurate geometries.

**Conclusion:** Our findings indicate that the inclusion of water molecules in ligand-protein docking results in significant increases in docking accuracy when the use of quantum chemical partial charge assignment on both ligand and proteins for predicting the docking simulations.

DACS (Database of available chemical compounds) was developed to design a HTS collection or focused libraries for the FMP Screening Unit. But the dramatically increase of unique compounds of one order of magnitude from less than 10 to around 30 to 40 million compounds currently and approximately hundred million compounds in the next years requires a reorganization and redevelopment of the storage and searching strategy. To manage such a massive amount of data a relational database for registration and normalization of vendor catalogs was developed. This database contains the highly redundant data of the vendor catalogs and is converted in a second step into the non-redundant Unique Structure Database which represents a data warehouse combining vendor data, structural descriptors and in-house classification tools including our earlier developed ADMET- and reactivity filters as well as our in-house fragment-based fingerprints used for library design tasks. The management of both database systems is part of a new developed Java application, which handles the user management for the data upload in the Registration Database and the conversion into the Unique Structure Database. Moreover a first version of a web service is in preparation. This service allows the scientist not only to search for compounds and fragments in the Unique Structure Database but also to combine such a search with the FMP tools to classify compounds for their usability in biological assays.

**Reference**

De novo design of selective compounds: a fragment-based pipeline applied to the \( \beta_2 \) adrenergic receptor

Florent Chevillard*, Peter Kolb
Kolblab, Institute of Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6-10, 35037 Marburg, Germany
E-mail: florent.chevillard@uni-marburg.de

Journal of Cheminformatics 2014, 6(Suppl 1):P25

GPCRs play a key role in transmembrane signaling and are involved in many physiological processes, such as regulation of behavior, heart rate and the immune system. Therefore they are very important targets for pharmaceutical agents. Our project focuses on the \( \beta_2 \) Adrenergic Receptor [1,2] (\( \beta_2 \)AR). The \( \beta_2 \)AR is mainly involved in vasodilation and bronchodilation in the human body. The recently solved structures of the \( \beta_2 \)AR open up new possibilities in the design of novel specific ligands using structure-based approaches. Here, we describe a pipeline to grow an unspecific fragment-sized scaffold for the \( \beta_2 \)AR. The protocol uses focused docking of fragments in two different zones identified within the binding site. The top ranked fragments are then computationally added to the core scaffold, filtered, minimized, evaluated by flexible ligand docking and inspected for later synthesis. Our initial results show that promising ligands can be identified by adding discriminating fragments to a core scaffold and that the generated compounds provide a reasonable synthetic accessibility.

References

Estimation of the biogas production rate, a chemometrical approach

Tetyana Beltramo1*, Susanne Theuerl1, Michael Klacke1, Bernd Hitzmann1
1Department of Process Analytics and Cereal Technology, Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, 70599,

Figure 1(abstract P23) Charge and force field applied to inhibitors for better docking score.
Biogas production rate is an important criterion for the entire biogas production process. In the present study the biogas production rate was evaluated using more than 30 process variables measured at an agricultural biogas plant in Germany during two months. The measured variables include the chemical measurements (such as pH, dried matter, amount of organic acids), energy supply specifications, temperature level and substrate ingredients. The prediction of the biogas production rate was done using chemometric methods. The results of the different methods were compared and the most accurate method was identified. Here the cross-validated prediction error (RMSECV) computed using leave-one-out method was less than 5 percent for both PCR and PLSR models (less than 190 $\text{m}^3/\text{d}$), while the calculated correlation coefficient ($r^2$) for PLSR reached 0.85 and 0.75 for PCR. For better prediction accuracy a metaheuristic search of the process relevant variables was performed. Here the Ant Colony Optimisation (ACO) improved the prediction performance of PLSR, decreasing the RMSECV to less than 2 percent ($95 \text{m}^3/\text{d}$) while increasing the $r^2$ to 0.98. These are promising results, which prove the feasibility of using this evaluation methodology for monitoring in biogas production processes.

P27
Hit series selection in noisy HTS data: clustering techniques, statistical tests and data visualisations

Christoph Müller1,*, Daniel Ormsby1, Isabella Feierberg2, Olga Engkvist2, Christian Tyran3, Michael J Hartshorn1
1Dotmatics Ltd, Windhill, Bishop's Stortford, CM23 2ND, UK; 2AstraZeneca AB, Peppardsleden 1, Mölndal, 43183, Sweden
E-mail: christoph.muller@dotmatics.com
Journal of Cheminformatics 2014, 6(Suppl 1)P27

High throughput screening (HTS) is one of the most prominent techniques used in the beginning stages of a drug discovery programme to identify those few hit compounds that can be used as starting points in subsequent studies [1,2]. However, an HTS experiment often entails a very data-intensive and challenging hit prioritization process that yields the mentioned hit compounds. The workflow described in this study aims to make this decision-making process easier by combining the structural and biological information of compounds used in an HTS. In particular, the workflow combines various clustering and nearest neighbourhood schemes with a non-parametric statistical test in order to prioritize those groupings of compounds that are likely of being relevant to the biological target of interest [3].

The novel workflow was evaluated under various aspects in a retrospective study using publicly available quantitative HTS (qHTS) datasets [4]. One of the main benchmarking aspects in this study was the ability to correctly identify as many true active compounds as possible. Therefore different chemical descriptors and clustering schemes were tested in combination with the statistic to measure their classification performance.

The workflow was integrated into Dotmatics’ Vortex, a platform for analysing chemical information using chemoinformatics methods and data visualisations tools [5]. This integration enables researchers to easily extend their current HTS workflow in order to discover new hit series and reveal hidden relationships between compounds, scaffolds and clusters.

References

P28
Dynamic information system for small molecules

Kiran K Telukunta, Xavier Lucas, Kersten Döring, Björn A Grüning, Stefan Günther
Institute of Pharmaceutical Sciences, Research Group Pharmaceutical Bioinformatics, Freiburg, Baden-Württemberg, 79104, Germany
E-mail: kiran.telukunta@pharmazie.uni-freiburg.de
Journal of Cheminformatics 2014, 6(Suppl 1)P28

The analysis of the biological effects of small molecules by mining lot of growing databases is an important task in the field of pharmaceutical sciences. To identify potential new similar drugs or to assess health risks from chemicals requires detailed knowledge about compounds. The pursued project is aiming at the integration of existing data resources of small molecules that are combined with tools for the prediction of biological effects based on molecular interactions. Users can query the system for a given compound and retrieve information on questions such as:

- Which other similar compounds exist?  
- Which proteins are mentioned in the literature in the same context of the given or similar compounds(s)?  
- Has the compound been tested in a bioassay?  
- Is the compound toxic?  
- Is the compound purchasable?  
- Is the compound patented?

Answers to the above questions will be given by the web interface dynamically from the web server. The system further records the users’ choices and learns from them such that more important information is shown more prominently.

The system is based on a large compound library consisting of above 70 million molecules collected from different publicly available resources (e.g., PubChem [1], ChEMBL [2], ZINC [3], StreptomeDB [4]) and will have the ability to accommodate new databases easily. An integrated text mining tool (CIL [5]) provides information on functional relations of queried compounds to proteins and refers to articles describing them. Tools for the prediction of biological effects such as SuperPred [6] will also be included. Subsequently, the results will be adapted to the requirements of the user and gives an extensive digest of relevant information for pharmaceutical researchers.

References

P29
Making the most of approximate maximum common substructure search

Péter Englert1,2, Péter Kovács1,2
1Department of Algorithms and Applications, Eötvös Loránd University, Budapest, H-1117, Hungary; 2ChemAxon Ltd., Budapest, H-1031, Hungary
E-mail: pkovacs84@chemaxon.com
Journal of Cheminformatics 2014, 6(Suppl 1)P29

The maximum common substructure (MCS) problem is of great importance in multiple aspects of chemoinformatics. It has diverse applications ranging from lead prediction to automated reaction mapping and visual alignment of similar compounds. Many different algorithms have been developed [1], both exact and approximate. Since the MCS problem is NP-complete, the
strict time constraints of most applications can only be realistically satisfied by fast and robust approximation methods.

We developed two efficient heuristic algorithms. One is based on the popular approach of reducing the MCS problem to finding the maximum clique in the modular product of the two molecule graphs. The other is based on a new algorithm by Kawabata, called the build-up method [2]. We also incorporated other techniques, for example, the topological fingerprinting primarily used in substructure and similarity searching. We optimized our implementations for use in multiple applications developed at ChemAxon. In some applications, for example, hierarchical MCS-based clustering or similarity search in large databases, the algorithms are required to give close to optimal results in limited time. To meet these conflicting demands, our implementations were enhanced with strong heuristics. Upper bound calculation methods were also applied for screening and early termination purposes.

In other applications, for example, reaction mapping or visual alignment, the challenge is that topological features must also be taken into account. Apart from the size of the found common substructure, the determined one-to-one correspondence between the atoms of the molecules is also very important. Effective heuristics were developed to guide the algorithms to prefer those solutions in which the relative positions of the common fragments of the input molecules are as similar as possible.

Our implementations have been thoroughly tested and benchmarked. They also have been compared to publicly available solution methods, and integrated into different products at ChemAxon. This has shown that the presented MCS algorithms can adequately cover the conflicting requirements of typical applications. We present our methods and heuristics along with their effects on running time, memory usage, as well as the size and features of the result.

References

P30

Computational modeling, docking and molecular dynamics of the transcriptional activator ComA bound to a newly-identified functional DNA binding site
Juan C Mobarec1*, Diana Wolf2, Ilka B Bischof2, Peter Kolb1
1Department of Pharmaceutical Chemistry, Philipps University Marburg, Marburg, Hesse, 35032, Germany; 2Ruprecht-Karls-University Heidelberg, Heidelberg, Germany
E-mail: Juan.Mobarec@uni-marburg.de
Journal of Cheminformatics 2014, 6(Suppl 1)P30

Quorum sensing is the mechanism by which bacterial cells communicate to each other in response to changes in cell density. Secreted signaling molecules reach other bacteria and trigger an internal physiological response, like production of degradative enzymes and antibiotics, competence development, sporulation and pathogenesis. In Bacillus subtilis, the transcriptional activator ComA regulates several genes of the quorum sensing response. ComA binds to recognition elements (RE) in bacterial promoters, activating transcription. The DNA binding domain of ComA has four α-helices, which contain a helix-turn-helix (HTH) motif. Several promoters in B. subtilis have known inverted repeat motifs (RE1 and RE2), where a ComA dimer can bind. However, a third and fourth recognition-like element (RE3 and RE4) in a direct repeat (DR) arrangement have recently been identified downstream of the known ComA box. Currently there is no structural information on how a DR would bind to ComA. Here, we present computational results supported by experiments that the DR (RE3 and RE4) form a functional domain for recognition of a ComA dimer. Flexible protein-DNA docking of the data used to get insight into the putative binding mode of ComA bound to the new direct repeat. The lowest-energy docked conformations of the ComA-DNA complex were tested for dynamic stability with explicit molecular dynamics simulations. Clustering of the sampled conformations was used to select a representative structure of the ComA-DR-DNA complex. In our results, the second α-helix of the HTH contributes most of the DNA recognition, binding to the major grooves of the DR-DNA, and interacting mostly with the DNA bases. We pinpoint specific ComA-DNA interactions that may have a key role for recognition and affinity. Furthermore, the physical interaction between ComA and the new DR was demonstrated in vitro, and its functionality was confirmed in vivo.

Our results strongly support the hypothesis that ComA dimers can bind to direct repeats, and provide an atomistic model for its recognition. Additionally, we suggest specific interactions for fine-tuning of transcription which will be tested experimentally.

P31

Large-scale docking approaches to the kinome
Denis Schmidt1*, Peter Kolb1
Pharmaceutical Chemistry, Philipps-University, Marburg, Germany
E-mail: denis.schmidt@uni-marburg.de
Journal of Cheminformatics 2014, 6(Suppl 1)P31

Docking nowadays is a widely used tool and has successfully and repeatedly been applied to identify hits with new chemical scaffolds [1]. Despite its successful applications, docking still suffers mainly from its scoring functions, which tend to be optimized for speed at the disadvantage of accuracy. Because of that, docking successfully enriches active molecules compared to inactive ones, however the false positive rate in docking runs is still very high. Several attempts have been made to improve docking by applying consensus scoring approaches using multiple structures or softwares or by rescoring compounds using normalization procedures [2-4]. At the same time, the number of large datasets of activity data derived from high-throughput screening methods has strongly increased during the last years, especially for kinases, allowing re-evaluation of these techniques [5-7]. We herein carried out a large-scale docking approach using 650 different kinase crystal structures. To the best of our knowledge, this is the largest number of targets docked to at the same time. Subsequently, different normalization techniques were re-evaluated using the experimental activity data as target function. Furthermore, we investigated the similarity of the kinase structures within the “docking universe” and correlated these similarities with other similarity metrics.

References

P32

Using structure- and Ligand-based pharmacophores as filters to discriminate Human Aryl Sulfotransferase 1A1 (SULT1A1) binders into substrates and inhibitors
Salwa M Soliman1*, Gerhard Wolf2
Freie Universitaet Berlin, Institute of Pharmacy, Department pharmaceutical chemistry, Koepenick-Luisestrasse 2-4, 14195 Berlin, Germany
E-mail: salwamosad@zedat.fu-berlin.de
Journal of Cheminformatics 2014, 6(Suppl 1)P32

Predicting metabolism transformation is one of major challenges in drug discovery [1]. Sulfotransferase 1A1 (SULT1A1), one of phase II metabolism enzymes, is the major SULT in adult liver catalysis [2]. It metabolizes
many endogenous compounds and is relevant in carcinogenesis due to its ability to modify diverse promutagens and procarcinogen xenobiotics [3]. In order to make a discriminative model that classifies group of SULT1A1 binders into substrates and inhibitors, a combination of structure, ligand-based, and docking based pharmacophores have been generated and validated by LigandScout [4].

On one hand, structure-based pharmacophores have been derived from PDB files of good binders (substrates and inhibitors). On the other hand, ligand-based interaction maps have been conducted from some drug classes that show different substrate/inhibitor activity towards SULT1A1. Finally, highly active substrates and inhibitors have been docked into the enzyme using GOLD [5], and subsequently molecular interaction fields have been developed for the most plausible poses. As a retrospective validation, all pharmacophores simultaneously have been used to screen more than 100 SULT1A1 binders covering several activity classes and different chemical scaffolds. The model showed good discriminative power to differentiate between inhibitors, substrates and mixed substrates/inhibitors.

References

Figure 1 (abstract P33) Iterative Workflow.
providing true “pythonesque” constructs. Since significant functionality of the toolkit is implemented as external Tcl script function snippets, and future enhancements will probably preferably be coded in Python without providing also a Tcl port, providing automatic and fully transparent access to language-mismatched components has been an important and rather peculiar design goal.

Examples of the new toolkit scripting capabilities shall be presented, as well as a documentation of the challenges involved in the design of a parallel multi-language interface to a large software system.

Acknowledgements: We gratefully acknowledge support by Vertex Pharmaceuticals, Inc. for this project.

P35
Entropy gain due to water release upon ligand binding
Mazen Ahmad1, Olga Kalinina, Thomas Lengauer
Max-Planck-Institut für Informatik , Saarbrücken, Campus E1-4, 66123, Germany
E-mail: mahmad@mpi-inf.mpg.de
Journal of Cheminformatics 2014, 6(Suppl 1):P35

Experimental thermodynamic data of the ligand-receptor association showed that the entropy changes upon binding are positive and large enough to be important driving forces of the binding process for a considerable number of ligand-receptor complexes [1]. The expected source behind such an entropy increase is the release of the water molecules from the binding pocket and around the ligand to get more freedom in the bulk phase. However, this important source of entropy is usually ignored due to the lack of a method to compute it. Therefore, we developed a method to compute the entropic gain due to water release upon ligand binding.

Water molecules close to a charged residue are restricted from free rotation due to the interaction between their dipole moments and the electric field. This restriction result in a decrease in their entropy. The loss of entropy can be gained again when the water molecules displaced from area that under the electric field upon the binding process. To compute the loss of entropy due to the electric field, we developed a method based on the knowledge of the electric field which affects the rotation of the water molecules using the continuum electrostatic methods. This provides a possibility for the calculation of the entropic loss from the partition function of the water molecules under an electric field. For monovalent ions, the computed loss in entropy upon solvation correlates well with experimentally measured one.

The loss of entropy of water close to hydrophobic molecules (Hydrophobic entropy) cannot be explained by the effect of the electric field from the molecule because the magnitude of the electric field is not strong enough to restrict the mobility of the water molecules. However, the loss in entropy results from the asymmetric interactions of the water molecules with the hydrophobic molecules and the neighbour water molecules. To compute the hydrophobic entropy we developed an empirical method which is derived from a linear relation between the hydration entropy of small hydrophobic molecules and the hydrophobicity of the water molecules around the surface of the small molecule. A training data set of 142 small molecules shows a significant linear relation between the hydrophobic entropy and the computed hydrophobicity.

Reference

P36
Elucidating protein-protein interactions using the HYDE scoring function
Eva Vennmann1, Nadine Schneider1, Gudrun Lange2, Matthias Rarey1
1Center for Bioinformatics, University of Hamburg, Bundesstr.43, 20146 Hamburg, Germany; 2Bayer CropScience AG, Industriepark Hoechst, 65926 Frankfurt am Main, Germany
Journal of Cheminformatics 2014, 6(Suppl 1):P36

Protein-protein interactions take place in every aspect of life. The diversity of those interactions and their significant role in regulatory pathways account for the special interest in them. As shown in various diseases, protein-protein interactions can be deregulated and the cause of illness. Therefore, a better and more detailed understanding of protein-protein interactions is of great importance.

Protein-protein interactions can be classified according to their stability into permanent, long-lasting or transient, functionality-dependent complexes [1]. The latter ones are of special interest for using them to influence regulatory pathways. Deeper insights into protein-protein interactions can be achieved by comprehending the stability of protein-protein complexes and especially of the interface in all its details.

In our study we analyzed the stability of protein-protein interactions using the recently developed HYDE scoring function [2-4]. HYDE consistently describes hydrogen bonds, the hydrophobic effect and polar dehydration and has been proved successful in estimating protein-ligand binding affinities. In this way, HYDE enables to estimate the stability of protein-protein interactions as well as the energetical contribution of single amino acids e.g. to identify so called ‘hotspot’ residues.

References

P37
Identification of SUMO activating enzyme 1 inhibitors utilizing virtual screening approach
Ashutosh Kumar1,2, Akhiro Ito1, Mikako Hirohama1, Minoru Yoshida3, Kam YJ Zhang1
1Zhang Initiative Research Unit, Institute laboratories, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan; 2Chemical Genetics Laboratory/Chemical Genomics Research Group, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
Journal of Cheminformatics 2014, 6(Suppl 1):P37

Sumoylation is a post-translational modification affecting diverse cellular processes including DNA replication and repair, chromosome packing and dynamics, genome integrity, nuclear transport, signal transduction and cell proliferation [1]. Sumoylation involves the covalent attachment of a small ubiquitin like modifier (SUMO) protein to ε-amino group of lysine residues in specific target proteins via a sequential action of an activating enzyme E1 (SUMO E1), a conjugating enzyme E2 and a ligase E3. Among the sumoylation proteins, SUMO E1 is responsible for the activation of SUMO in the first step of the sumoylation cascade [2]. SUMO E1 is linked to many human diseases including cancer and thus making it a potential therapeutic target [3]. However, only a few inhibitors were reported up to now that includes three natural products, semi-synthetic protein inhibitors and one AMP mimic [4-6]. Here in this research, the combination of structure based virtual screening and in vitro sumoylation assay was used to identify potential small molecule inhibitors of SUMO E1 that could be used in chemical biology and therapeutic studies. Our virtual screening protocol involves the fast docking of a small molecule library to rigid protein followed by redocking of top hits using a method that incorporates both ligand and protein flexibility. Subsequently, the top ranking compounds were prioritized using molecular dynamics simulation based binding free energy calculation. The result of biological assay and subsequent similarity search resulted in the identification of two classes of small molecules that shared biaryl urea scaffold. Both of these chemical classes displayed moderate inhibitory potency against SUMO E1. The most potent compound of each class inhibited the in vitro sumoylation with an IC50 of 11.1 and 13.4 μM. These compounds inhibit sumoylation by blocking the formation of SUMO-E1 thioester intermediate. Our study presents starting points for the development of novel therapeutic agents against various diseases targeting SUMO E1.

References


P38

Accelerating turbo similarity searching in chemoinformatics on multicore and GPU platforms

Marwah Haritham Al Laila*, Nurul Malim, Nur‘Aini Abdul Rashid School of Computer Sciences, Universiti Sains Malaysia, 11800, Penang, Malaysia
E-mail: marwah_allaya2000@yahoo.com

Journal of Cheminformatics 2014, 6(Suppl 1):P38

The increase in the database size of chemical compound requires a longer time in processing for any searching algorithms [1]. With the focus on an algorithm called Turbo Similarity Searching which have been proven to have good accuracy in retrieving actives [2], we propose that this algorithm be implemented on the widely-used many-cores and multi-cores processors which are easily available at lower cost. This would help medicinal chemists run virtual screening faster while maintaining the accuracy. The many-cores processors are on-chip processors that could run simultaneously at a clock-cycle. Whilst the multi-cores processors are originally being developed to support graphics processing hence being called Graphics Processing Unit (GPU). However, the usage of GPU is now being generalized to include other general purpose operation [3]. Many works on the parallel field have tried implementing computational algorithms on this unit [4].

Taken into consideration on the compute intensive characteristics of TSS, we investigate the best method to parallelize it for better execution time. TSS is a two-phase algorithm in which its second phase is a compute intensive portion of the algorithm. Hence, if this phase could be accelerated, TSS would run faster and searching through a larger database would not be less a problem. This poster describes our experimental details and results on the matter.

References


P39

Computational evaluation of the η6-arene during the ATH of imines on Noyori’s Bull catalyst

Petr Kacer*, Sot Petr, Jan Pechacek, Jakub Janusčak, Marek Kuzma Institute of Chemical Technology, Prague, 166 28, Czech Republic
E-mail: Petr.Kacer@vscht.cz

Journal of Cheminformatics 2014, 6(Suppl 1):P39

Asymmetric hydrogenation ranks to the most intensively researched way of preparation of enantiothermic pure compounds which are demanded e.g. in pharmaceutical industry, cosmetics or agriculture. In the field of asymmetric transfer hydrogenations (ATH) of C=N and C=O double bonds Noyori’s ruthenium (II) complexes represent significant breakthrough. This catalytic system consists of three integral parts - chiral monotosylated diamine ligand – i.e. N-(p-toluenesulfonyl)-1,2-diphenylethylenediamine (TsDPEN), η6-coordinated aromatic molecule (e.g. benzene or p-cymene) and halogen counteranion (usually chloride). General formula of the catalyst can be written as [RuCl(n^6-arene)(N-aryl sulfonyl)DPEN]). Aforementioned fragments/ligands offer countless number of possibilities for structural modifications – e.g. elongation of carbonaceous spacer between phenyl rings within 1,2-diphenylethylendiamino fragment, alkylation of amino group, usage of diversely substituted η6-aromatic molecule, employment of different aryls within arylsulfonyl fragment etc. Systematic evaluation of these modifications has multilateral benefits because it not only helps to clarify mechanistic phenomena but also contributes to the deeper understanding of relationship between structure and catalytic activity. With sufficiently big and rich data base it should be possible to tailor catalyst’s properties specifically for given substrate (or class of substrates) and reaction conditions (solubility, stability, etc.). Our research is focused primarily on comprehension of role of the η6-aromatic molecule during asymmetric transfer hydrogenation of imines. This ligand plays very important mechanistic role because its structure (respectively interaction with the substrate) allows asymmetric course of the reaction. Arene ligand can in certain cases form stabilizing CH/π interaction between aromatic part of substrate and therefore lower energy of transition state. This led us to the hypothesis that alteration of its structure could strongly affect enantioselectivity and reaction rate. This hypothesis has been brought up an case of ATH of C=O bonds, which dramatically differs from hydrogenation of C=N bonds. Usually only alkyl-substituted arene molecules are used as aromatic ligands. In our study we have prepared and compared four catalysts with different aromatic ligands (benzene, p-cymene, mesitylene, 1,2,3,4,5,6-hexamethylbenzene) according to their performance (reaction rate, enantioselectivity) during hydrogenation of variously substituted 3,4-dihydroisoquinolines and tried to interpret obtained results via means of computational chemistry.

Acknowledgements: This work has been financially supported by the Grant Agency of the Czech Republic (Grant GACR 106/12/1276).

References


**P41**

Consistency of sugar structures and their annotation in the PDB

Deepti Jaswal1, David Sehnal, Radka Svobodová Valčková, Círa-Maria Ionescu, Jaroslav Koča
National Centre for Biomolecular Research, Faculty of Science and CEITEC -Central European Institute of Technology, Masaryk University, Brno, 625 00, CZ, Czech Republic
E-mail: deepti@chemi.muni.cz
Journal of Cheminformatics 2014, 6(Suppl 1):P41

Cell-cell recognition is the first stage in many important phenomena such as infection by bacteria and viruses, communication among cells of lower eukaryotes, binding of sperm to egg, etc. [1]. Cell-cell recognition relies on sugar (carbohydrate) specific interactions at the cell surface. Theoretical studies typically involve molecular modeling of sugars and sugar-specific protein receptors. These studies rely on structural information obtained mainly by crystallography and nuclear magnetic resonance, and deposited in the Protein Data Bank (PDB). Since the main purpose of PDB is to store the structure of proteins and nucleic acids, thus, it is expected that PDB structure files are complete and correctly annotated. Nonetheless, sugars exhibit a structural diversity larger than amino acids or nucleotides, a property which makes them ideal for recognition. At the same time, sugars are characterized by specific and very sensitive structural features such as multiple chiral centers on each ring. Because of these peculiarities, the validation and annotation of sugar structures is not straightforward.

Our first goal was to develop a methodology that can identify whether a sugar structure is complete and correctly annotated. Our second goal was then to check all PDB entries containing sugars, and record whatever problems we encounter in the sugar structures. For this purpose we collected all sugar structures which appear as ligands in PDB entries, and compared them to model structures available in Ligand Expo [2], a curated repository of ligand chemical and structural information. In order to perform the comparison we used several tools for structural comparison currently available (SiteBinder [3], Open Babel [4]), as well as two in-house programs. We report here on our findings regarding the complete and correctly annotated sugar structures in PDB, together with the problematic cases.

**References**


**P42**

Towards understanding the chemical environment effect on gold-containing clusters

Doreen Mollenhauer1*, Nicola Gaston2
1Callaghan Innovation Research Limited, 69 Gracefield Road, 5010 Wellington, New Zealand; 2MacDiarmid Institute for Advanced Materials and Nanotechnology, Victoria University of Wellington, P. O. 600, 6140 Wellington, New Zealand.
E-mail: Doreen.Mollenhauer@callaghaninnovation.govt.nz
Journal of Cheminformatics 2014, 6(Suppl 1):P42

Gold clusters and nanoparticles have attracted continuous attention due to interesting and important electronic, catalytic and optical properties [1,2]. As the chemical environment strongly affects the catalytic properties, an understanding of this is essential to be able to control these properties. In order to study the influence of the ligand shell on the catalytic properties we have studied various gold clusters in interaction with different ligands by performing DFT-D3, SCS-MP2 and CCSD(T) calculations [3,4]. The effect of the ligands to the geometric and electronic structure of the gold clusters is analysed in a systematic way [5,6]. Furthermore as bimetallic gold-palladium catalysts have been found to have improved catalytic properties in various reactions in comparison to the monometallic clusters, the influence of the ligand shell is investigated for small systems.

**References**


**P43**

Exploiting solvent effects in drug design and optimization

Jean-Francois Truchon1, Kristina Grabowski2, Barbara Sander3, Alain Ajamian2
1Vertex, Laval, H7V 4A7, Canada; 2Chemical Computing Group, Köln, 50672, Germany; 3Chemical Computing Group, Montreal, H3A 2R7, Canada
E-mail: ajamian@chemcomp.com
Journal of Cheminformatics 2014, 6(Suppl 1):P43

Upon ligand binding, solvent molecules around the binding pocket and the ligand become displaced or rearranged. These desolvation energies can be a significant portion of the total binding energy, and thus represent opportunities for ligand design. Computing desolvation energetics typically requires lengthy simulations, but this talk presents a fast and easy-to-use method (3D-RISM) which computes desolvation energies in minutes, without using explicit simulations. Application to ligand optimization is demonstrated using case studies.

**References**


**P44**

Fuzzy context specific matched molecular pairs

Peter Schmidtko1, Vincent le Guilloux1
1Discrine SAS, Paris, 75011, France
E-mail: peter.schmidtko@discrine.com
Journal of Cheminformatics 2014, 6(Suppl 1):P44

Matched molecular pairs (MMPs) are commonly used to assess the importance of chemical modifications on small molecules versus a particular property [1]. One major advantage of MMPs is the direct interpretation by medicinal chemists. Despite the popularity of MMPs several drawbacks hinder their systematic applications in drug discovery projects. First, in order to derive sufficiently representative statistics for extracting design rules rather large datasets of molecules and activities are needed. Second, MMPs are often used without context specificity, likely because of the lack of sufficient data. This results in weak statistics. One transformation could result
to be beneficial, but is detrimental in reality once the context taken into account [2]. Context specific MMP analysis is clearly advantageous for in-depth understanding of SAR, but hindered due to lack of data. Here we present a novel methodology for providing robust statistics for fuzzy context specific MMPs even on medium sized data-sets. Molecules are transformed to a reduced graph using the Discngine Pharmacophore Graph methodology [3], a molecular graph reduction method closely related to classical reduced graphs [4]. Next, molecular fragmentation and MMP detection is directly performed on the pharmacophore graph. The herein used pharmacophore graph representation allows to group together very similar contexts and/or fragments and thus increase population size compared to classical MMP analysis. Validation of the fuzzy context specific MMP (fcsMMP) is presented and outcome is compared to classical MMP analysis. Last the Discngine Network framework is used to organize the derived design rules for efficient large scale mining and results extraction in a real world Med Chem context and applications are shown.

References

P45
A new method for rapid comparison of protein binding pockets by capturing spatial distributions
Timo Krotzky, Gerhard Klebe
Department of Pharmaceutical Chemistry, Philipps-Universität, Marburg, 35032, Germany
E-mail: krotzky@uni-marburg.de
Journal of Cheminformatics 2014, 6(Suppl 1):P45
Efficient determination of structural similarities between protein binding pockets is an important challenge in computational chemistry. A degree of similarity in the mutual comparison is often estimated in terms of graphs and by calculating a metric such as the maximum shared common subgraph. Cavbase [1] was developed as a tool for the automatic detection, storage and classification of putative protein binding sites. Cavbase assigns so-called pseudocenters to the cavity-flanking amino acids, which characterize their physicochemical properties with respect to molecular recognition. Subsequently, the pseudocenters are used as graph nodes to accomplish mutual binding site comparisons. This way of modeling protein binding sites, however, tends to be computationally very demanding, which often leads to very lengthy evaluations of the similarity measures.

In this study we propose Rapid Pocket Matching using Distances (RAPMAD), a new modeling formalism for Cavbase entries which allows for highly efficient similarity calculations. Here, protein binding sites are represented by sets of distance histograms based on specific spatial reference points [2] in order to characterize the distribution of pseudocenters within the cavity. The histograms can be both generated and compared with linear complexity. Attaining a speed of approximately 20,000 comparisons per second, pocket comparisons across large datasets and even screenings of entire databases become easily feasible.

We demonstrate the discriminative power and the orders of magnitude faster runtime of this novel method by carrying out several classification and retrieval experiments. Among others, datasets of protein cavities hosting specific cofactors are used for classification experiments, where RAPMAD results in a considerably higher rate of correct classifications compared to other alternative approaches while it requires only a fraction of their runtime. Moreover, a set of proteases [3] was investigated, where it turned out that RAPMAD is able to distinguish between different Merops clans such as serine or metallo proteases.

References

P46
Peptide lineage against Gram-negative bacterial infection – first-in-class peptide inhibitor of H. pylori HtrA
Anna M. Peina1,*, Thomas Schmidt1, Tim Fugmann1, Nicole Tegtmeier1, Steffen Backert4, Silja Wessler3, G. Schneider2
1Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Zurich, 8093, Switzerland; 2Department of Molecular Biology, University of Salzburg, Salzburg, 5020, Austria; 3Philochem AG, Oetelfingen, 8112, Switzerland; 4Department of Biology, University of Erlangen-Nuremberg, Erlangen, 91058, Germany
Journal of Cheminformatics 2014, 6(Suppl 1):P46
More than 50% of the world population is infected with Helicobacter pylori (H. pylori) and the actual H. pylori treatments fail with increased regularity because of continuously rising antibiotic resistances. To meet this challenge, we focus on the development of a new anti-infective therapy against H. pylori by targeting a secreted enzyme, high temperature resistance A (HtrA). Release of the serine protease HtrA near the host's gastric epithelial cells leads to loss of cellular adhesion due to E- cadherin cleavage [1]. We investigated the substrate cleavage sites of HtrA in its natural substrate E-cadherin performing a label-free mass spectrometry-based proteomic analysis and identified preferred cleavage positions by Edman sequencing. Further, we developed the first HtrA peptide inhibitor by synthesizing cleavage site fragments and analogues. Surface plasmon resonance (SPR) was used to perform binding studies. In vitro substrate cleavage assays as well as cellular infection assays fully support the biophysical data.

References

P47
Target prediction by cascaded self-organizing maps for ligand de-orphaning and side-effect investigation
Daniel Reker, Tiago Rodrigues, Petra Schneider, Gisbert Schneider
Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Zurich, 8093, Switzerland
Journal of Cheminformatics 2014, 6(Suppl 1):P47
Computational chemogenomics approaches have emerged as a means to predict modulations of biomolecules by ligands. We implemented a method for the prediction of the macromolecular targets of small molecules combining state-of-the-art approaches that compare physicochemical properties and pharmacophoric features of query molecules with known drugs. Investigating similarity from multiple vantage points has been shown to increase the prediction accuracy in a retrospective evaluation. The method has been applied in multiple projects to “de-orphan” molecules with unknown main target and investigate potential side-effects of drug candidates. In a first application, the method identified a molecular scaffold as a potentially privileged structure of druglike compounds for chemotherapy resistant tumor therapy [1]. In a second project, the tool revealed the potential of up to 5% of known bioactive substances to have unrecognized epigenetic effects by modulating histone deacetylase (HDAC) activity – thereby stressing the importance of probing for epigenetic effects in long-term drug toxicity studies [2].
P48
Targeting flexibility: a structure-based computational study revealing allosteric HIV-1 protease inhibitors
Jens Kurzat1, Nickolay Todoroff, Tiago Rodrigues, Petra Schneider, Gisbert Schneider
Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland
Journal of Cheminformatics 2014, 6(Suppl 1)248

We present the discovery of innovative low molecular weight inhibitors against human immunodeficiency virus 1 (HIV-1) protease. Structure-based virtual screening focused on potential allosteric surface cavities revealed these compounds [1]. To identify and prioritize such cavities we performed a molecular dynamics simulation were we concentrated on flexible and transient potential binding sites. For several time-points of the simulation we computed receptor-derived pharmacophore models in the so-called hinge region (‘Exo site’) and screened a large screening compound library [2]. The most potent hit shows inhibition in a non-competitive mode of action.

References
2. Beck M, Srivastava S, Khoury K, Herdtweck E, Dömling A:

P49
Go with the flow: de-orphaning focused combinatorial libraries
Michael Reutlinger1,2, Tiago Rodrigues, Petra Schneider, Gisbert Schneider
Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology (ETH), Zurich, 8093, Switzerland
Journal of Cheminformatics 2014, 6(Suppl 1)249

The fast pace of drug discovery programs, aided by high-throughput screening campaigns, often relies on the generation of combinatorial libraries to identify new chemical entities. The Ugi 4- and 3-component reactions in particular [1], have proven to be robust in producing both tool compounds and drugs [2,3]. Here we report a high-throughput entry into the imidazopyridine scaffold, using a microfluidic-assisted synthesis setup, coupled to a target prediction tool to de-orphan a focused compound library with high success rate, and identify an innovative GPCR-inhibiting chemistry. Combinatorial compounds were correctly identified as ligand-efficient adenosine A2aR and adrenergic α1bβ2 inhibitors with Kᵢ values in the low micromolar range.

References

P50
Discovery of novel α-amylase inhibitors using structure-based drug design
Jamil Al-Asri1, Gerhard Wolber
Computer-Aided Molecular Design, Pharmaceutical Chemistry Department, Free Universität Berlin, Berlin, 14195, Germany
E-mail: jmlasa@zedat.fu-berlin.de
Journal of Cheminformatics 2014, 6(Suppl 1)250

α-Amylase is an endoamylase and belongs to glycoside hydrolase family 13 (GH 13) according to the classification of carbohydrate-active enzymes [1]. It initiates starch hydrolysis into smaller oligomers. Inhibitors of this enzyme are of pharmacological importance as α-amylase is considered as attractive target for treating elevated post-prandial blood glucose levels resulting in obesity and type II diabetes. Besides the application as a drug, it is highly interesting to classify nutritional components, such as food additives or secondary plant metabolites with respect to their modulation of α-amylase.

We present a model that predicts the affinity of small organic molecules to α-amylase. On the basis of available crystal structures (Figure 1) [2], we developed a virtual screening workflow for the identification of novel non-peptidic, non-carbohydrate α-amylase inhibitors. In addition to virtual screening using structure-based 3D pharmacophore models [3], molecular docking and clustering for diversity selection have been applied as post-screening filters. Fourteen virtual hits were purchased and tested in vitro using a kinetic assay with p-Nitrophenyl-t-d-maltopentaoside (PNPG) as a chromogenic substrate. Three of those fourteen compounds showed concentration-dependent inhibition with promising IC₅₀ values (hit rate: 21%).

References

P51
ChemicalToolBox and its application on the study of the drug like and purchasable space
Xavier Lucas1, Björn A Grüning, Stefan Günther
Pharmaceutical Bioinformatics, Institute of Pharmaceutical Sciences, Albert-Ludwigs-Universität, Freiburg, D-79104, Germany
E-mail: xavier.lucas@pharmazie.uni-freiburg.de
Journal of Cheminformatics 2014, 6(Suppl 1)251

The ever increasing amount of data and computational capabilities in the cheminformatics field has led to a scenario where efficient techniques for storage and processing in an integrated, modular, and easily accessible platform are in vital demand. Here, we present ChemicalToolBox, a compilation of more than 30 tools integrated into a single computational chemistry and cheminformatics platform based on the Galaxy workflow management system [1,2]. We have recently designed a workflow within the ChemicalToolBox to generate a library of compounds containing around 70 million unique commercially available small molecules, i.e. the purchasable space [3]. Subsequently, we have used filtering rules based on structural patterns and chemical alarms to discard problematic molecules, representing a very large portion of the drug-like and purchasable space, along with other drug discovery data sets including more than 2 million fragments (Figure 1). Furthermore, we have computed several physicochemical descriptors to discover general trends applying to each subset.

References
1. [https://github.com/bgruening/galaxytools/tree/master/chemicaltoolbox].

P52
Accessing Open PHACTS: interactive exploration of compounds and targets from the semantic web
Katrin Stierand1, Tim Harder1, Lothar Wissler2, Christian Lemmern2, Matthias Rarey3
1Center for Bioinformatics (ZBH), University of Hamburg, Bundesstr. 43, 20146 Hamburg, Germany; 2BiosolveIT GmbH, An der Ziegelei 79, 53757 Sankt Augustin, Germany
E-mail: stierand@zbh.uni-hamburg.de
Journal of Cheminformatics 2014, 6(Suppl 1)252

Pharmacological research is hampered by scattered data which have to be retrieved by varying methods and in different data formats. This heterogeneity increases research costs and limits throughput. Over the last two years, the Open PHACTS Discovery Platform [1] has been
developed as a centralized repository, integrating pharmacological data from a variety of information resources and providing tools and services to query these integrated data in pharmacological research. Following an application-oriented approach, the Open PHACTS project started with the definition of potential use cases in the form of prioritized research questions [2], most of which can only be answered by accessing multiple data sources in the web. The development of the platform as well as the services has been guided by these questions. Here, we present the ChemBioNavigator (CBN) [3], a web application allowing to navigate the Open PHACTS chem-bio space with a focus on small molecules and their targets. CBN comprises of a large visualization area with different view modes and two information panels, allowing a deeper insight in information for compounds and targets. It allows interactive exploration of compound sets through sorting and subset selection as well as extending sets by substructure or similarity search. The relation between compounds and targets is defined by assay data from the Discovery Platform. Each compound and each target is annotated with information from multiple data sources which is provided together with the provenance for each data point. In this contribution we roughly outline the OpenPHACTS/CBN technology and present a number of high-priority research questions, highlight the advantages of exploiting the integrated data through the CBN’s smart and intuitive interface.

References

PS3
Parameterization to NDDO-based polarizable force field
Heike Thomas¹, Matthias Hennemann³, Stefan Guissregen⁴, Timothy Clark¹,²*
¹Computer-Chemie-Centrum, Friedrich-Alexander-Universitat Erlangen-Nurnberg, Naegelsbachstrasse 25, 91052 Erlangen, Germany; ²Sanofi Deutschland GmbH, R&D, LGCR, Structure, Design and Informatics, Building G878, 65926 Frankfurt am Main, Germany; ³Centre for Molecular Design, University of Portsmouth, King Henry Building, Portsmouth PO1 2DY, UK
E-mail: Tim.Clark@fau.de
Journal of Cheminformatics 2014, 6(Suppl 1):PS3
In Computer-Aided-Drug-Design (CADD), the electrostatic interactions contribute strongly to the interaction between the drug-molecule and the target. Further, the Coulomb term is crucial for calculating the electrostatic contribution to the solvation energy. In spite of this, conventional Force Fields use the obsolete physical concept of point-monopoles (net atomic charges) and thus, are not able to represent the molecular electrostatic potential (MEP) accurately or are even wrong for atoms that have positively and negatively charged areas on their surface [1]. A far better way to...
describe the MEP is the is multipole-based semiempirical MO-theory [2,3].
For the parameterization of the polarizable hpCADD Force Field, the two methods are combined in order to obtain both the MEP and structures and energies. Additionally, the differentiation of atom-types leads to more detailed information about the MEP.

**References**

**P54**
Pharmacophore annotation using extended Hückel theory
Paul Labute\(^1\), Markus Kossner\(^2\), Alain Ajamian\(^1\), Martin Santavy\(^1\), Anna Lin\(^1\)
\(^1\)Chemical Computing Group, Montreal, H3A 2R7, Canada; \(^2\)Chemical Computing Group, Köln, 50672, Germany
E-mail: mkossner@chemcomp.com
Journal of Cheminformatics 2014, 6(Suppl 1):P54

Pharmacophore models play an essential role in drug discovery. Generating pharmacophore models which encode accurate molecular recognition features are highly dependent on properly defined annotations. Simplistic or ill-defined pharmacophore annotations which do not capture subtle electronic or geometric effects lead to many inaccuracies. SMARTS patterns which are often used to specify annotation “rules” are subject to such inaccuracies.

The application of Extended Hückel Theory (EHT) to pharmacophore annotations compensates for deficiencies observed in “rule” based methods by taking into account electron withdrawal and resonance effects and treating these effects in a consistent manner independent of structural depiction. The application of the EHT approach will be further described and discussed through a number of case studies.

**P55**
Surflex-QMODO: physically meaningful QSAR
Alexander Steudle\(^1\), Rocco Varel\(^1\), Ajay N Jain\(^2\)
\(^1\)Certa International, Martin-Kollár-Straße 17, 81829 München, Germany;
\(^2\)Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94143-0912, USA
E-mail: Alexander.Steudle@certara.com
Journal of Cheminformatics 2014, 6(Suppl 1):P55

Computational methods for predicting ligand affinity where no protein structure is known generally take the form of regression analysis based on molecular features that have only a tangential relationship to a protein/ligand binding event. Such methods have utility in retrospective rationalization of activity patterns of substituents on a common scaffold, but are limited when either multiple scaffolds are present or when ligand alignment varies significantly based on structural changes. In addition, such methods generally assume independence and additivity of effect from scaffold substituents. Collectively, these non-physical modeling assumptions sharply limit the utility of widely used QSAR approaches for prospective prediction of ligand activity.

The approach we report builds upon the Compass approach by constructing physical models of a protein binding site based upon ligand binding data. The result is a binding site composed of molecular fragments that can be treated as a target for molecular docking. The binding site model consists of molecular fragments that can account for multiple positions of protein residues. It is not a literal reconstruction of a single configuration of protein residues. New molecules are docked directly into the binding site, with their binding scores assessed as the predicted binding affinity. By deriving a virtual binding pocket at the same time as the relative poses of ligands are identified, the key analogy is that one can treat a computational model of a binding site as one treats a protein binding pocket. We seek the optimal fit of ligands into the binding site. One begins with a guess as to the initial alignment of ligands, then constructs a model of activity that depends on the ligand’s poses. The model can be thought of as a virtual receptor. Next, one seeks poses for each ligand that optimize their interaction with the virtual receptor. Then, the virtual receptor is refined, making use of the new ligand poses, and the process iterates between pose refinement and virtual receptor refinement. As the virtual receptor evolves, the changes in ligand scores due to pose optimization decrease. When the iterative process converges, the final poses of the ligands are optimal with respect to the final virtual receptor. The software implementing the algorithms for pocket construction and ligand activity prediction constitute a new module within the Surflex platform, called Surflex-QMODO (Quantitative Modeling). The Surflex-QMODO approach addresses the physical linkage between activity model and molecular binding mode with pockets having detailed structure comparable to true protein binding sites. Because the model building process results in a model that selects ligand alignments based on mutual interaction, there is a direct correspondence between the physical process of protein/ligand binding and the act of prediction.

**P56**
Is there a sodium effect in fibrillar amyloid-\( \beta \) oligomers?
Anselm HC Horn\(^1\), Danyi Huraskin, Heinrich Sticht
Bioinformaik, Institut für Biochemie, Friedrich-Alexander-Universität Erlangen-Nürnberg, Fahrstr. 17, 91054 Erlangen, Germany
E-mail: Anselm.Horn@fau.de
Journal of Cheminformatics 2014, 6(Suppl 1):P56

Hall mark in Alzheimer’s Disease (AD) is the aggregation of amyloid-\( \beta \) (A\( \beta \)) peptide into oligomers and fibrils. Nowadays, the soluble oligomers are believed to be the most neurotoxic species, probably the causative agents in AD. Amyloid-beta fibrils on the other hand may serve as reservoirs for small toxic oligomers. While it is well known from experiment, that the aggregation process is modulated by salt concentration in solution, the molecular details of the underlying interactions are not. Salts occur ubiquitously in physiological environments and are known to have profound effects on the solubility of proteins (Hofmeister series). Monovalent alkali metal ions exhibit a more subtle effect on A\( \beta \) aggregation in experiment than doubly charged species [1,2].

In this contribution we investigate the presence of the so-called ‘sodium-effect’ in fibrillar A\( \beta \) oligomers. This effect was originally described for dendrimer micelles formation modulating the self-organization of amphiphilic carboxylates: Na\(^+\) forms bridging complexes with carboxylate groups, in contrast to K\(^+\) [3].

Molecular dynamics simulations for a systematic series of single and double layer fibrillar A\( \beta \) oligomers in aqueous 150mM salt solution provide insights about the stabilizing interactions between the cations and charged A\( \beta \) key residues (e.g. Glu22). Comparison with a previous computational study at low ion concentration [4] show similarities and differences on a structural level. Interestingly, ions access the A\( \beta \) water channel region via entry paths suggested previously [4].

**References**

**P57**
SiteHopper - a unique tool for binding site comparison
Jose Batista\(^1\), Paul CD Hawkins\(^2\), Robert Tolbert\(^2\), Matthew T Geballe\(^2\)
\(^1\)OpenEye Scientific Software, Cologne, 50672, Germany; \(^2\)OpenEye Scientific Software, 9 Bisbee Court, Suite D, Santa Fe, NM, 87508, USA
E-mail: batista@eyesopen.com
Journal of Cheminformatics 2014, 6(Suppl 1):P57
Knowledge of and information about protein binding sites has become increasingly important in the drug discovery process and not just for molecular biologists [1]. By comparing binding sites within and across protein families, relevant details about the functionality and selectivity of a target protein can be extracted leading to useful insights for the development of new ligands [2]. SiteHopper provides a powerful alternative method to the traditional use of sequence alignment for this purpose. Using OpenEye’s Shape [3] and Spicoli toolkits [4], SiteHopper quickly calculates a 3D shape representation of the active site colored by the chemical properties of the protein residues defining the active site. These active site representations can be rapidly aligned and assessed for shape and chemistry similarity. As SiteHopper is built on the OpenEye toolkits, it is highly flexible and customizable for a variety of end-uses. In this presentation, the methodology behind SiteHopper will be introduced and multiple relevant applications will be shown.

References

P58
Theoretical illustration of the effect of 1-ethyl-pyridinium trifluoroacetate ionic liquid in the enhancement of the Diels–Alder reaction of isoprene with acrylic acid
Hafida Chemouri1,2, Sidi Mohamed Mekelleche1*
1Laboratory of Applied Thermodynamics and Molecular Modeling, Department of Chemistry, Faculty of Science, University A. Belkaïd, B. P. 119, Tlemcen, 13000, Algeria; 2Preparing School in Sciences and Techniques of Tlemcen, BP 165 RP Bel Horizon, 13000, Tlemcen, Algeria
E-mail: sm_mekelleche@mail.univ-tlemcen.dz
Journal of Cheminformatics 2014, 6(Suppl 1):P58

The Diels–Alder reaction is a powerful tool in organic synthesis and in the chemical industry. Recently an increased attention has been focused on the development of green methods for the purpose of improving rate and selectivity of this reaction. In recent years, ionic liquids (ILs) have gained a lot of attention as green solvents in organic synthesis and other chemical processes. This is mainly due to their favorable inherent properties such as chemical and thermal stability, no measurable vapor pressure, nontoxicity, nonflammability, catalytic activity, high polarity, ease to recycle, etc. Diels–Alder reactions of isoprene (1) and acrylic acid (2) (Figure 1) have been investigated in pyridinium based ILs. The ionic liquid 1-ethyl-pyridinium trifluoroacetate [EtPy] [CF3COO] is found to be an excellent reaction solvent with significantly increased rate for this reaction compared to organic solvent.

In order to confirm this experimental finding, theoretical investigation of the regioselectivity meta/para of this Diels Alder reaction have been carried out. The calculations have been performed in gas phase, in CH2Cl2 organic solvent and in 1-ethyl-pyridinium trifluoroacetate protic IL. Asynchronous concerted mechanisms yielding to the formation of the para regioisomers as the major cycloadducts is shown by the analysis of the relevant stationary points of the potential energy surface and intrinsic reaction coordinate calculations carried out in gas phase. The calculation of activation and reaction energies indicates that the para cycloadducts are favored both kinetically and thermodynamically. The calculations, performed using explicit solvent model, show a considerable decrease of the activation barrier in the IL in comparison with gas phase and CH2Cl2. The obtained results put in evidence the importance of hydrogen bonding formed between the IL and the dienophile fragment in the promotion of this Diels-Alder reaction. The calculations are carried out with the Gaussian 09 suite of programmes using the B3PW91 exchange-correlation functionals together with the 6–31G(dp) basis set and the obtained results are in good agreement with experimental outcomes.

References

Cognate docking has been used as a test for pose prediction quality in docking engines for decades. While cognate docking is not the problem that docking engines are put to in their normal use (that being cross docking), it is expected that good performance in cognate docking is a necessary but not sufficient condition for good performance in cross docking. In this talk we report a statistically rigorous analysis of cognate docking using tools in the OEDocking suite [1,2]. We address a number of critically important aspects of the cognate docking problem that are frequently poorly handled in publications in this area; dataset quality, methods of comparison of the docked pose to the ligand model pose and analysis of the results to determine if and by how much a given method is actually better than another.

The first problem is handled through the use of our recently published Iridium-HT dataset [3]. To overcome the second problem we use a variety of measures to compare a docked pose to the ligand model pose. In addressing the third problem we utilize a variety of statistical methods to determine whether, and by how much, changes in the scoring functions actually improve cognate docking performance; a major challenge in this area is the paired nature of the deviation data. We caution against the

Figure 1 (abstract P58) Diels–Alder reactions of isoprene (1) and acrylic acid (2).
mechanical application of statistical tests, however, and advocate for searching for substantive and meaningful significance, as well as statistical significance.

References

P60
Empirical charges for chemoinformatics applications
Tomáš Bouchal 1, Radka Svobodová Vařeková 2, Tomáš Raček 3, Crina-Maria Ionescu 4, Stanislav Geidl 5, Aleš Řenek 2, Jaroslav Koča 6
1 National Centre for Biomolecular Research, Faculty of Science and CEITEC - Central European Institute of Technology, Masaryk University, Brno, 625 00, CZ, Czech Republic; 2 Institute of Computer Science and Faculty of Informatics, Masaryk University, Brno, 620 00, CZ, Czech Republic
E-mail: tbouchal@mail.muni.cz
Journal of Cheminformatics 2014, 6(Suppl 1):660

Partial atomic charges describe the distribution of electron density in a molecule, and therefore they provide clues regarding the chemical behaviour of molecules. Atomic charges are frequently used in molecular modelling applications such as molecular dynamics, docking, conformational searches, binding site prediction, etc. Recently, partial atomic charges have also become popular chemoinformatics descriptors [1].

Partial atomic charges cannot be determined experimentally, and they are also not quantum mechanical observables. For this reason, many different methods have been developed for their calculation. These charge calculation methods can be divided into two main groups, namely quantum mechanical (QM) approaches and empirical approaches. QM approaches provide accurate charges, but they are very slow and therefore not feasible for large sets of molecules. Empirical charges can be calculated quickly and their accuracy is similar to QM, making empirical charges more appropriate for chemoinformatics applications. A very useful empirical charge calculation method is EEM (Electronegativity Equalization Method) [2,3]. This method provides charges comparable to the QM approach for which the given EEM model was parameterized.

The weak point of this empirical method, as well as of other empirical methods, is the necessity for parameterization, and also the insufficient coverage of currently available EEM model parameters.

In our work, we first analysed, how applicable are currently published EEM methods, is the necessity for parameterization, and also the insufficient coverage of currently available EEM model parameters.

In our work, we first analysed how applicable are currently published EEM models, how many molecules from datasets can be parameterized, and also its results. We found, the coverage is about 50-75%. We would like to show a methodology for preparation of parameters with higher coverage (>95% of molecules) and also its results.

References

P61
QM quality atomic charges for proteins
Stanislav Geidl 1, Crina-Maria Ionescu, Radka Svobodová Vařeková, Jaroslav Koča
1 National Centre for Biomolecular Research, Faculty of Science and CEITEC - Central European Institute of Technology, Masaryk University, Brno, 625 00, CZ, Czech Republic
E-mail: standag@chemi.muni.cz
Journal of Cheminformatics 2014, 6(Suppl 1):P61

The concept of atomic point charges is well established in theoretical chemistry. Atomic point charges have played an important role in understanding and modeling chemical behavior by allowing to extract and quantify information stored in the molecular electron distribution of chemical compounds. Thus, atomic point charges have been used to estimate reactivity indices, dissociation constants, partition coefficients, the electrostatic contribution in molecular dynamics or docking studies. It is therefore desirable to have knowledge of the values of atomic charges in proteins (see, e.g., [1]). Unfortunately, accurate and universally applicable approaches for atomic charge calculation based on quantum mechanics (QM) are very time consuming and thus cannot be employed for large biomolecules like proteins. An alternative is to use empirical charge calculation methods, such as the electronegativity equalization method (EEM) [2], which is very fast and has accuracy comparable to QM. The challenge is to calibrate (i.e., parametrize) this method for proteins. This parameterization can be done using atomic charges calculated by different types of QM approaches. EEM can be as accurate as the QM approach for which EEM was calibrated.

In our work, we present the workflow of the EEM calibration process. Afterwards, we validate EEM models for 12 types of QM charges, including the newest approaches like iterative Hirshfeld [3]. The accuracy of the obtained EEM models is evaluated on insulin and ubiquitin. We also show two case studies demonstrating the applicability of atomic charges computed via EEM: a small docking study, and the calculation of electrostatic potential based on the EEM charges [4].

References

P62
IterTunnel: a method for predicting and evaluating ligand EgressTunnels in proteins with buried active sites
Laura J Kingsley 1, Markus A Lill 2
1 Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN, 47906, USA
Journal of Cheminformatics 2014, 6(Suppl 1):P62

The computational prediction of ligand entry and egress paths in proteins has become an emerging topic in computational biology due to the potential for estimating kinetic properties of drug binding. These properties are related to important pharmacological quantities such as the k off and k on rate of drugs [1,2]. We have investigated the influence of protein flexibility on tunnel prediction using geometric methods by comparing tunnels identified in static structures with those found in structural ensembles of three CYP isozymes. We found drastic differences between tunnels predicted in the crystal structures as opposed to those predicted in the ensembles [3]. Furthermore, we found significant differences between tunnels identified in the apo versus the holo protein ensembles [3].

While geometric prediction provides a good starting point for tunnel prediction, in order to estimate kinetic properties, more detailed investigations of the ligand binding process are required. We have developed a tunnel prediction methodology, IterTunnel, which predicts tunnels in proteins and estimates the free energy of ligand unbinding using a combination of geometric tunnel prediction with steered molecular dynamics and umbrella sampling [4]. Applying this new method to cytochrome P450 2B6 (CYP2B6), we demonstrate that the ligand itself plays an important role in reshaping tunnels as it traverses through a protein. This process results in the exposure of new tunnels and the closure of pre-existing tunnels as the ligand migrates from the active site. We found that many of the tunnels that are exposed due to ligand-induced conformational
changes are amongst the most energetically favorable tunnels for ligand egress in CYP2B6 [4].

References