Friday, 24th September, 2021
9AM ET / 2PM BST / 3PM CET. Online Event

Sponsors and Partners:

astex

evotec

NovaData Solutions

CCPBioSim

SCHRÖDINGER

Exscientia
DRIVEN BY KNOWLEDGE

MGMS at 40 agenda

Note: Talk times might change. All times given in UK Time, a.k.a. British Summer Time (BST).

2.00 – 2.10 pm	Welcome
2.10 – 2.40 pm	**MGMS Frank Blaney Award Winner 2018** Christopher Rowley, Department of Chemistry, Carleton University *"Simulating Protein-Ligand Binding with Neural Network Potentials"*
2.40 – 3.10 pm	**MGMS Frank Blaney Award Winner 2019** Isidro Cortes, European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL-EBI) *"What is realistic and what are illusions in Artificial Intelligence in drug discovery? A discussion on ways to impact, and why we are not there yet"*
3.10 – 3.40 pm	**MGMS Frank Blaney Award Winner 2020** Fernanda Duarte, Department of Chemistry, University of Oxford *"Efficient strategies to explore chemical reactivity"*
3.40 – 3.50 pm	Break
3.50 – 4.30 pm	Lightning talks, poster session
4.30 – 5.30 pm	**Keynote speaker** W. Graham Richards, Oxford Drug Design, a founding member of the Molecular Graphics Society (now the MGMS) *"The origins of the MGMS"*
5.30 – 6.00 pm	MGMS Annual General Meeting

List of poster presenters:

1. Alex Jamieson-Binnie, School of Chemistry, University of Bristol (UK)
   "Applying Rotational Forces to Atomistic and Coarse-Grained Models using Interactive Molecular Dynamics"
2. Charmaine Chu, Department of Chemistry, University of Liverpool (UK)
   "Potential amphiphilicity: novel descriptor for surfactants"
3. Stefania Monteleone, Research Informatics, Evotec (UK)
   "Comparative Analysis of Free Energy Perturbation Tools"
4. Blaine Mooers, University of Oklahoma Health Sciences Center (USA)
   "PyMOL in RStudio"
5. Atanas Patronov, Discovery Sciences Molecular AI Team, Astra Zeneca (Sweden)
   "Application of generative models in AI driven De Novo design of small molecules"
6. Rachael M. E. Pirie, Cole Research Group, Newcastle University (UK)
   "Evaluating 3D Shape-Similarity Using the Kähler Potential"
7. David J. Ponting, Lhasa Limited (UK)
   "Automated chromophore identification to support prediction of light-mediated reactivity"
8. Thomas Steinbrecher, Applications Science, Schrodinger (Germany)
   "PyMOL - An open-source visualisation environment for communicating results of chemical and biochemical research"

Zoom breakout rooms:

Room1: Alex Jamieson-Binnie and Atanas Patronov
Room2: Rachael M. E. Pirie and David J. Ponting
Room3: Charmaine Chu and Stefania Monteleone
Room4: Blaine Mooers and Thomas Steinbrecher

Simulating Protein-Ligand Binding with Neural Network Potentials

Christopher Rowley

Department of Chemistry, Carleton University

Quantitative modeling of protein-ligand binding is a key objective in computer-aided drug discovery. Calculating the structure and relative stabilities of drug conformations using molecular mechanical force fields remains challenging because a huge number of parameters need to be defined to describe the chemical space of drug-like molecules comprehensively. These models are also limited by the conventional form of force fields (e.g., harmonic bonds…). Neural network potentials (NNPs) provide a radically different approach to this problem, where neural networks are trained to reproduce the energies predicted by quantum chemical calculations in a transferable manner. These can be used to simulate the dynamics of arbitrary molecules with the accuracy and transferability of a QM model but with a computational cost closer to an MM model. We have developed a new technique termed NNP/MM that uses an NNP to represent the intramolecular terms of the ligand while the rest of the system is described using a conventional force field. This method is effective in predicting the bound pose of a drug and can be used to calculate the conformational component of the absolute binding energy. In some cases, NNP/MM provides a vastly different binding energy than a pure MM model because of deficiencies in the MM model. This method is also effective in refining cryo-EM protein-ligand complexes by providing a chemically accurate but computationally efficient and parameterization-free model for the ligand.

What is realistic and what are illusions in Artificial Intelligence in drug discovery? A discussion on ways to impact, and why we are not there yet

Isidro Cortes-Ciriano

European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL-EBI), Wellcome Genome Campus, Hinxton

While Artificial Intelligence (AI) has had a profound impact on diverse scientific disciplines, such as computer vision, its contribution to the discovery of new drugs has been somewhat limited. In this presentation, I will firstly discuss in which stages of the drug discovery process improving either the time needed, the success rate of decisions, or decreasing cost would have the most profound impact on bringing new drugs to market. After cost of capital, changes in clinical success rates would have the strongest impact on improving success in drug discovery, indicating that the quality of decisions regarding which compound to take forward are more important than their speed or cost. Secondly, I will argue that while most efforts applying AI in the context of drug discovery have been centred around the question of 'how to make new compounds', the question of 'what compound to make' taking into account efficacy and toxicity endpoints has received comparatively less attention. As a result, the potential of AI to make significant progress in drug discovery is limited by the proxy data currently available, which are often of insufficient quality with respect to the in vivo assessment of efficacy and safety. Harnessing the power of AI to inform decision making at those stages of the drug discovery process with the strongest impact on bringing new drugs to market is thus contingent on the generation of practically relevant data in sufficient quantities to enable the discovery of novel chemistry with novel modes of action, and showing desirable efficacy and safety in the clinic.

Efficient strategies to explore chemical reactivity

Fernanda Duarte

Department of Chemistry, University of Oxford

Automating the search for new chemical reactions is one of the 'grand challenges' in computational chemistry. While characterizing reaction energy pathways has been routine for practitioners, efficiently exploring complex reaction mechanisms remains time consuming and a highly non-systematic endeavour. Even for an experienced computational chemist, reaction prediction is resource-intensive. For computational chemistry to become more practical and impactful, new developments are necessary to reduce technical barriers and increase efficiency and accuracy when modelling reactions in complex environments.

In this contribution, I will present our team's efforts to tackle these challenges by introducing automation and machine-learned potentials to study reactions mechanisms in the condensed phase. I will describe our recently developed tool, *autodE,* which automates the characterization of reaction pathways with minimal user input and expertise. Previous developments in the field have focused on the free exploration of unknown pathways, which require no *a priori* knowledge but are computationally costly and therefore limited to small systems.[1] Our approach is applicable to both organic and organometallic reaction classes, accounts for conformational sampling, is compatible with several electronic structure theory packages, and is freely available. I will illustrate the functionality and general applicability of *autodE* in a range of reaction classes, including complex organic and metal-catalyzed reactions.

Concerning increasingly complex systems in solution, e.g., including explicit solvation of flexible systems, I will present an efficient and general strategy that employs machine-learned (ML) potentials within the Gaussian Approximation Potential framework.[2] Our work demonstrates that reactive ML potentials can achieve *ab initio* accuracy at a much lower cost than related strategies. A diverse range of examples will be presented, for which expensive AIMD simulations would otherwise be needed.

References

1. T. A. Young, Joseph J. Silcock, Alistair J. Sterling, F. Duarte.* *autodE: Automated Calculation of Reaction Energy Profiles— Application to Organic and Organometallic Reactions.* *Angew. Chem. Int. Ed.* **2020**, 59, 2.

2. T. A. Young, T. Johnston-Wood, V. Deringer,* F. Duarte.* A Transferable Active-Learning Strategy for Reactive Molecular Force Fields. *ChemRxiv* **2021** Preprint. *Chem. Sci.*, 2021, Advance Article

The origins of the MGMS

W. Graham Richards

Oxford Drug Design, a founding member of the Molecular Graphics Society (now the MGMS)

The origins of the MGMS came from two streams, crystallographers who wanted to display the positions of nuclei and theoretical chemists who wanted to show electron distribution. Both had the same problem that the then existing journals would only publish colour pictures at very expensive cost to the authors. There was a need for a new journal devoted to the area. This was achieved by creating the new society with its own journal.

Applying Rotational Forces to Atomistic and Coarse-Grained Models using Interactive Molecular Dynamics

Alexander D. Jamieson-Binnie

School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS

Interactive and steered molecular dynamics are becoming more ubiquitous as methods for rare event sampling. Steered molecular dynamics applies predefined forces to a system, whilst interactive molecular dynamics (iMD) allows users to view simulations and modify forces in real time.[1] It achieves this by using technologies such as virtual reality to give users three-dimensional control over their system.

Interactive molecular dynamics has up to now only applied forces that translate atoms or groups of atoms through space while maintaining their orientation. However, in complex applications, such as drug docking, being able to control the orientation of molecules would be very useful. Current methods for inducing rotations using iMD lead to distortion of the internal molecular structure, whilst other methods that treat the ligand as a rigid body fail to allow conformational changes within it.

My proposed method applies an interactive force based on three terms:

- A translational critically-damped spring that moves the molecule towards the target position, driven by the position of the user's controller.
- A rotational critically-damped torsional spring that rotates the molecule as a whole towards the target orientation, which is based on the orientation of the user's controller.
- A centripetal force to maintain the molecule's shape as it rotates, by applying a force towards the current axis of rotation.

Using this kind of forces allows us to rotate a molecule in an intuitive way, while avoiding distorting the internal structure of the molecule.

This rotational protocol can equally be applied to atomistic systems, where it can perform tasks such as rotating enzyme ligands or protein side chains, and to coarse-grained systems, such as oxDNA,[2] where each particle has an orientation which can now be changed by the user interactively.

References:

[1] M. B. O'Connor, S. J. Bennie, H. M. Deeks, A. Jamieson-Binnie, A. J. Jones, R. J. Shannon, R. Walters, T. J. Mitchell, A. J. Mulholland and D. R. Glowacki, *J. Chem. Phys.*, (2019), **150**, 220901.

[2] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, (2012), **137**, 135101.

Potential amphiphilicity: novel descriptor for surfactants

Charmaine S.M. Chu[a], *Jerry Winter[b], Neil G. Berry[a]*

[a]Department of Chemistry, University of Liverpool, Crown Street, Liverpool L69 7ZD
[b]Unilever R&D Port Sunlight, Wood St, Birkenhead, Wirral CH62 4UY

Amphiphilicity, the extent to which a molecule is hydrophilic and hydrophobic, is an important property of surfactants. The balance of hydrophilicity and hydrophobicity in a molecule affects how it sits at the oil/water interface, amphiphilicity is crucial to the oil/water interfacial tension, a key characteristic for surfactants. Quantifying amphiphilicity for a surfactant molecule holds many difficulties as amphiphilicity is dependent on the hydrophilicity and hydrophobicity of the molecule, and the relative position of the hydrophilic and hydrophobic sections and consideration of the conformations the molecule can adopt.[1]

Although the identification of the boundary between the hydrophilic and hydrophobic sections can be achieved manually using relevant chemistry knowledge, when the analysis in on a library of hundreds of molecules, there is a necessity to automate this process.

Within this poster, we present the novel automated protocol constructed for calculating potential amphiphilicity descriptor. This protocol allows a library of surfactant to be analysed for their hydrophobic and hydrophilic section boundaries and output the AlogP, number of rotatable bonds and effective surface area of the hydrophobic and hydrophilic sections, and the AlogP difference between the hydrophobic and hydrophilic sections. Using such output, comparison with critical micelle concentration, the broadly used property to assess surfactants, can be carries using QSAR methods.

References:

1. G. Georgiou, S.-C. Lin and M. M. Sharma, *Bio/Technology*, 1992, **10**, 60-65.

Comparative Analysis of Free Energy Perturbation Tools

Stefania Monteleone, Inaki Morao, *Tahsin Kellici*

Research Informatics, Evotec (UK) Ltd, Abingdon, UK.

Free energy perturbation (FEP) calculations have recently shown great improvements in performance and remarkable impact in the drug discovery process. Nowadays, many computational software suite provide specific algorithms and alternative strategies to optimize the performance of such calculations. Here we focus on a challenging but everyday task: For a given a list of virtual compounds, which one(s) should be synthesized next and which program performs best?

With this aim, we present a comparison of different state-of-the-art software, such as AMBER20[1] and FEP+[2], with MOE-TI, a streamlined interface to setup GPU-enabled molecular dynamics simulations based on the Thermodynamic Integration (TI) method in AMBER. TI simulations were carried out on a benchmark subset extracted from literature, using MOE version 2020 and AMBER20. Accuracies and rankings have been evaluated together with the mean unsigned errors for each molecular pair.

Our findings showed that the accuracy is comparable for MCL1 and Tyk2 test cases, whereas significant inaccuracy was found for thrombin and p38 examples. Although the three FEP methods showed to be valid, overall their performances were affected by the system (both the target and the ligands) under consideration. We therefore recommend to carry out preliminary tests with all available methods in order to identify the best approach prior to any heavy computation on new virtual compounds.

References:
[1] T. Lee et al., *J Chem Inf Model.*, (2020) **60**, 5595-5623.
[2] L. Wang et al., *J. Am. Chem Soc.*, (2015) **137**, 2695−2703.

PyMOL in RStudio

Blaine Mooers

Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA 73104

The reproducibility crisis in science undermines the public's support for science. This crisis includes molecular graphics where the scripts for making images are not shared. Some authors address this problem by adding the code to the supplemental information. This solution is awkward when there are many plots and their scripts. However, the scripts and output files can be included in a single electronic notebook for easier downloading. The notebook can also have explanatory prose interleaved between the blocks of code. The blocks of code are executable and allow the reader to reproduce the images in the manuscript. The ubiquitous Jupyter notebook with Python code is now common in supplemental materials.

However, many users of PyMOL users are not conversant with Python, which is the default programming language used in Jupyter notebooks. Yet, some of these users are comfortable with R through their bioinformatics work. The PyMOL Python API enables running of PyMOL inside of R via the R package reticulate[1]. Any resulting output can be returned to a R Markdown notebook inside RStudio. RStudio's support for code snippets includes tab triggers and tab stops. The first feature accelerates the insertion of code templates. The second feature supports the complete editing of the template code. This reduces time spent debugging the code. We formatted the pymolpysnips library[2] to ease the use of PyMOL with reticulate in RStudio. We also demonstrate the passing of data structures from PyMOL to R packages for further structural analysis. The library and sample notebooks are available on GitHub (https://github.com/MooersLab/rstudiopymolpysnip).

[1] K. Ushey, J.J. Allaire, and Y. Tang, https://CRAN.R-project.org/package=reticulate, (2021).

[2] B.H.M. Mooers, *Computers in Engineering & Sci.,* (2021) **23**, 47-53.

Application of generative models in AI driven De Novo design of small molecules

Atanas Patronov

MolecularAI, Discovery Sciences, R&D, AstraZeneca Gothenburg, Sweden

The goal of *de novo* design is to identify novel active compounds that can simultaneously satisfy a combination of essential optimization goals such as activity, selectivity, physical-chemical and ADMET properties. It is a process of generating novel compounds from scratch that are normally not present in databases and are not being previously considered in the context of the given target. Identifying a compound that meets all these criteria is still far from trivial.

Here we demonstrate how AI has been adopted for the needs of *de novo* design of small molecules. We show this in the context of REINVENT, which is a platform developed internally within AstraZeneca and made open source. It is based on using different flavors of generative models that know how to generate small molecules. The generative models are placed in various learning scenarios and can learn to generate solutions that satisfy a diverse set of criteria.

Evaluating 3D Shape-Similarity Using Riemannian Geometry

Rachael M. E. Pirie,[a] *Daniel J. Cole,*[a] *Matthew Forshaw,*[b] *Stuart J. Hall.*[c]

[a] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.
[b] School of Computing, Newcastle University, Newcastle upon Tyne, NE4 5TG.
[c] School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, UK.

The use of 3D shape similarity in computer-aided drug design has gained traction in recent years due to the importance of protein-ligand shape complementarity in binding.[1] Mathematical approximations are used to describe the 3D shape of molecules. These are less expensive to compute than quantum-mechanical models, allowing for faster screening of large databases. Consideration of 3D shape offers the advantage of being able to account for multiple conformers. These methods also enable scaffold hopping, which can aid the rescue of drug candidates found to have issues such as poor solubility or toxicity.

We outline a new approach for approximating shape using Riemannian geometry. The surface is treated as topologically equivalent to a sphere embedded in complex projective space ($\mathbb{CP}^n$). For each point in $\mathbb{CP}^n$ a metric exists that describes the local geometry about the point. These local geometry descriptors can be clustered to give the Laplace-Beltrami spectrum,[2] which represents the global geometry of the surface using a series of 9 eigenvalues. Two representations are then compared using the Bray-Curtis metric to determine shape-similarity between molecules. As our method considers each atom individually we believe it will be more accurate and efficient than the existing methods. In addition, these representations are translation invariant and while they are not themselves rotation invariant, the behaviour of Hermitian matrices is well defined, allowing for straightforward identification of instances of rotation of the same molecule. The results of our initial development will be presented, along with an outline of our planned work benchmarking our method and investigating its usefulness in machine learning for drug discovery.

References:
[1] A. Kumar and K. Y. J. Zhang, *Front. Chem.*, (2018), **6**, 1-21.
[2] M. P. Seddon, D. A. Cosgrove, M. J. Packer and V. J. Gillet, *J. Chem. Inf. Model.*, (2019), **59**, 98–116.

Automated chromophore identification to support prediction of light-mediated reactivity

David J. Ponting*, Rachel Hemingway and Martin A. Ott
*david.ponting@lhasalimited.org
Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, UK

Many reactions in organic chemistry require initiation by absorption of a photon, exciting an electron to form a transient reactive species. Due to the quantisation of the energy levels of electrons within small molecules, the photon must have a specific energy defined by the energies of the molecular orbitals in the chromophore. This is principally defined by the size of the chromophore and the atom types contained therein. It should be noted, however, that absorption even of a photon of the correct energy is not certain; rather, it depends on parameters such as whether the transition is spin-forbidden or allowed and whether the relevant orbitals physically overlap. The prediction of the full absorption spectrum is therefore complex, but determination of chromophore size provides a rapid estimate of the potential for light-initiated reactivity.

We describe a method for the identification of chromophores by 'walking' through a molecular structure, counting as part of the chromophore those atoms that can contribute to the delocalised system – either by being part of a multiple bond, or having a lone pair or empty orbital. It is not, however, as simple as counting these: for example, the molecular orbitals in a C≡C triple bond are perpendicular to each other and thus only one can contribute, and while oxygen has two lone pairs and the potential for multiple bonds not all of these can contribute. Methods for determining solely from the structure which of these are involved are discussed.

The excited species formed may, if it does not revert to ground state by emission of a photon, react through a variety of mechanisms. Examples where this method could be used to improve the selectivity of predicted forced degradation reactions, without the need to determine either the molecular orbital energies themselves or the absorption spectrum, are presented.

PyMOL - An open-source visualisation environment for communicating results of chemical and biochemical research

Thomas Steinbrecher

Applications Science, Europe P, Schrödinger GmbH, Mannheim, Germany

PyMOL (https://pymol.org/) is a popular visualization environment that allows chemists and biochemists to showcase their work with eye-catching graphics and movies. Users can create realistic ray-traced graphics, produce molecular animations, send step-by-step presentations to colleagues, and embed PyMOL presentations within PowerPoint slides. It furthermore allows a high degree of automation using custom scripting or Python. Here we will review some of the highlights of more than two decades of development and usage as well as look at some of the latest improvements.