E-CAM Scoping Workshop. Solubility prediction

Location: CECAM-FR-RA, Ecole Normale Supérieure de Lyon, France
Webpage: https://www.cecam.org/workshop-0-1497.html
Dates: May 14, 2018 to May 15, 2018
Organizers: Eduardo Sanz (Physical Chemistry Department, Complutense University of Madrid, Spain), Carine Michel (CNRS, Ecole Normale Supérieure de Lyon, France), Daan Frenkel (University of Cambridge, United Kingdom), Robert Docherty (Pfizer Limited, United Kingdom)

1 State of the art

Based on the Biopharmaceutics Classification System (BCS), over 75% of drug development candidates have low solubility, which is a major issue for drug development as formulation of low solubility compounds can be problematic. Other industries rely on value-added formulation and thus on solubility issues. Despite tremendous efforts, a definitive accurate and comprehensive approach to predicting solubility has proven elusive. This workshop has focused on the different approaches to predict solubility trends.

Recent work includes a systematic experimental approach to examine key thermodynamic functions such as sublimation and hydration properties as a function of structural modifications and a comprehensive computational approach to lattice energy estimation from molecular descriptors. A recent review has analysed simple predictive methods for the estimation of aqueous solubility and the specific use of a chemical informatics and theory to predict the solubility of drug like molecules.

Algorithms for solubility calculations have been carried out by two different general approaches: (i) the search of the concentration where the solute and the solid chemical potentials are equal and (ii) direct simulations of the solid and the solution at contact. Both these approaches have been discussed at length in the workshop.

In the complementary area of structure activity relationships, automatic model generation process for building QSAR models using Gaussian Processes were discussed, a powerful machine learning modeling method. The stages of the process that ensure models are built and validated within a rigorous framework were examined: descriptor calculation, splitting data into training, validation and test sets, descriptor filtering, application of modeling techniques and selection of the best model. The effectiveness of the automatic model generation process for two types of data sets commonly encountered in building ADME QSAR models was explored.

2 Major outcomes

In this workshop were discussed the tools that allow an unprecedented deconstruction of the relative importance of molecular solvation and crystal packing on solubility. Recent work includes a systematic experimental approach to examine key thermodynamic functions such as sublimation and hydration properties as a function of structural modifications and a comprehensive computational approach to solubility, from chemical informatics approaches to advanced molecular simulations.

Accumulating reliable experimental data of hydration and sublimation is essential to benchmark established as well as novel simulation tools. However, it has been underlined how challenging it is to compare experimental data with simulations results. Indeed, computing values are obtained for an ideal system, which is not what we measure experimentally. In addition, it is difficult to get a complete comprehensive coherent picture from experiments. Typically, the solid in equilibrium with a saturated solution is not necessarily in the same crystallographic phase or even in the chemical state than the one that was initially introduced. Re-precipitation can occur and the solid phase in equilibrium with the solution should be systematically fully characterized. Nevertheless, to ensure the development of novel simulation methodologies, hydration free energies are currently collected in an online database FreeSolv, ready to be used in a simulation benchmark.

Algorithms for solubility calculations have been carried out by two different general approaches. The thermodynamic approach seeks the concentration at which the electrolyte chemical potential, in solution, is equal to that of the pure solid. To compute the Gibbs free energy cost of the insertion of particles in a liquid, the thermodynamic integration is the workhorse and an alternative is the Wang Landau approach. A direct simulation of the solubility equilibrium can be modelled using the slab method that put in contact the saturated solution with the surface of the solid. The electrolyte concentration in the solution phase sufficiently far from the crystal surface is taken to be the solubility. It has been largely applied on the NaCl case. A controversy was relayed in the literature demonstrating that this approach is very demanding computationally, prone to size effect issues and requires very long simulation time to be able to reach the equilibrium.

Both strategies rely strongly on the quality of the force field used. OPLS has been developed to compute hydration energies of small organic molecules. It provides a good balance between the description of the liquid and the solid description. However, with the shift towards larger molecules (MW>500g/mol), this force field reaches its limitations. In addition, the typical Lennard-Jones potential used to deal with dispersion forces for practical reasons but it is too repulsive at short distance. Polarisability may also be key to reach a better quality in simulation data. A proposition is to use the Yukama potential with smeared charge and charge on spring for polarisability. Neural network appears as an alternative, but the large number of parameters implies the use of extensive experimental database eventually complemented with ab initio data.

Another major challenge that has been put forward is a good prediction of the crystal phases that can adopt an organic molecule. In AstraZeneca, this has been circumvented using systematically an amorphous phase to describe the solid phase. Recently, several approaches have been proposed to predict polymorphisms, either combining new order parameters and string method, either benefiting from conceptual DFT to understand better the crystal packing.

3 Community needs

To further progress, the scientific community needs to strengthen interaction between specialists in modelling but also with other communities (experimentalists, other field of applications than pharmaceutical). There is a clear need for open data, with the publication of raw data obtained by simulation as well as experiments to be able to re-investigate the influence of the scheme chosen to split data into training/validating set. Some semi-empirical approaches are currently not shared, impeding their further development in a participative scheme. Others are widely shared (see for instance bottledsaft.org).

There is a clear need for improved force field that are able to describe large molecules (MW > 500 g/mol) in solution as well as in solid phase. This is also true for ions, since for instance the force field to describe Na+ and Cl- should differ in water and in the NaCl phase. In addition, innovative algorithms to predict polymorphisms are a necessity. To assess the quality of the novel methods that are to be invented, good experimental references data are compulsory, spanning multiple families of molecules. Those developments would be better performed with dedicated series of CECAM workshop around force field development.

In addition to innovative methods, there is also a strong need in speeding-up the current methodologies and make them available in codes that can be transferred to industry through software engineering.

4 Funding

The funding's were not discussed during the workshop. A clear line that emerged is to intensify collaboration between academia and industry. Integrating research and innovation, with strong societal impact in health as well as sustainability, modelling solvation is clearly in the line of the H2020 priorities. A possibility would be to build an initiative training network around solvation.

An action has recently started in France to structure the scientific community around solvation (http://solvate.cnrs.fr) gathering around hundred researchers

5 Will these developments bring societal benefits?

During this workshop, we have seen that solubility issues can be found across several fields of applications, from pharmaceutical to specialty chemical industry.

In pharmaceutical industry, improved knowledge on solubility will help in designing efficiently production lines of novel drugs (it is key in purification steps). It is also key in the final formulation of a drug limiting aggregation issues, to favour highly concentrated solution and avoid injecting huge volume of solution to patients.

In specialty chemical industry, improved knowledge on solubility in water can drive the formulation of greener lubricants, detergents, etc. using water as a based instead of a oil-derived based (this shift occurred few years ago for painting, it is a general trend to avoid any hazard once those formulated product are at use). This drastic modification necessitates a complete revision of the additives of which properties are directly related to their solvation.

The development and use of modelling in those industrial contexts could be key to speed up the time to market of novel products, with health and sustainable benefits for the society as an end-user. Already at use in some companies, the benefits of a modelling study have to be balanced against the experimental duration and costs. In most cases, trends can be obtained rapidly with chemo-informatics approaches, even on a desktop PC and are already highly valuable. More advanced methods (based on MM-MD) would bring more insight on more tricky cases, that can be also interesting topic of collaboration between industry and academia, challenging the state-of-the-art methodologies and triggering innovative developments.

6 Participant list

Organizers

BARENDSON, Samantha
Centre Blaise Pascal - ENS de Lyon, France
Docherty, Robert
Pfizer Limited, United Kingdom
Frenkel, Daan
University of Cambridge, United Kingdom
Guilleminot, Alexandra
École normale supérieure de Lyon, France
Michel, Carine
CNRS, Ecole Normale Supérieure de Lyon, France
Sanz, Eduardo
Physical Chemistry Department, Chemistry Faculty, University Complutense of Madrid, Spain

Anwar, Jamshed - University of Lancaster, United Kingdom
Borgis, Daniel - Maison de la Simulation, France
Costa Gomes, margarida - ENS, Lyon, France
Dronet, Severin - Michelin, France
Kolafa, Jiri - Institute of Chemical Technology, Prague , Czech Republic
Levesque, Maximilien - École Normale Supérieure, France
Li, Tonglei - Purdue University, USA
Lindfors, Lennart - Astra Zeneca, Sweden
Lozano, Sylvain - Total Marketing Services, France
MacKernan, Donal - University College Dublin, Ireland
Martis, Alessandro - APC Ltd., Ireland
Mitchell, John - University of Saint Andrews, United Kingdom
Nezbeda, Ivo - E. Hala Lab of Thermodyn., Acad. Sci. , Czech Republic
Padua, Agilio - Blaise Pascal University, France
Perlovich, German - Russian academy of sciences, Russian Federation
Rene Espinosa, Jorge - University Complutense of Madrid, Spain
Santiso, Erik - North Carolina State University, USA
Schnell, Benoît - Michelin, France
Smith, William - University of Guelph, Canada
Ukrainczyk, Marko - APC Ltd., Ireland