International Journal of Pharmaceutics 540 (2018) 185–193

Contents lists available at ScienceDirect

International Journal of Pharmaceutics

journal homepage: www.elsevier.com/locate/ijpharm

Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting

Christel A.S. Bergström*, Per Larsson

Department of Pharmacy, Uppsala University, Biomedical Centre P.O. Box 580, SE-751 23 Uppsala, Sweden

ARTICLE INFO

Keywords:
Computational prediction
Solubility
Solid state
Intestinal fluid
Quantitative structure property relationships
Molecular dynamics simulations

ABSTRACT

In this review we will discuss recent advances in computational prediction of solubility in water-based solvents. Our focus is set on recent advances in predictions of biorelevant solubility in media mimicking the human intestinal fluids and on new methods to predict the thermodynamic cycle rather than prediction of solubility in pure water through quantitative structure property relationships (QSPR). While the literature is rich in QSPR models for both solubility and melting point, a physicochemical property strongly linked to the solubility, recent advances in the modelling of these properties make use of theory and computational simulations to better predict these properties or processes involved therein (e.g. solid state crystal lattice packing, dissociation of molecules from the lattice and solvation). This review serves to provide an update on these new approaches and how they can be used to more accurately predict solubility, and also importantly, inform us on molecular interactions and processes occurring during drug dissolution and solubilisation.

1. Background

Poor drug solubility is one of the main obstacles in the drug discovery and development process and was recently identified to be strongly related to the choice of target explored. (Bergstrom et al., 2016) Solubility is the driving force for absorption and acceptable solubility in the intestinal fluid is a prerequisite for achieving sufficiently high drug blood concentrations to obtain a therapeutic effect when systemic effects are warranted. The solubility of a compound affects its absorption, distribution, metabolism, excretion and toxicity (ADMET) profile. Only when the ADMET properties of a drug-like compound are of a sufficiently high quality, and when the target has been validated, can the compound be developed into a new medication (Cook et al., 2014; Morgan et al., 2012). Since the molecular requirements of some targets inevitably result in poor solubility of the ligands, early awareness of this fact by the medicinal chemistry team is crucial for them to make the right decisions on which analyses and assays to perform. Understanding the risk of poor solubility is also important for analysing the results of ADMET assays, since there is potential to identify false readouts as an effect of precipitation or aggregation of the drug compound (Coan and Shoichet, 2008; Pohjala and Tammela, 2012). If the compound is precipitating, it is easy to identify why the *in vitro* screen has failed. However, if there is no visible sign of aggregation and/or precipitation it is much more difficult to interpret the results. This may lead to false conclusions being drawn from the assay and the "false positives" of the assay pushing the compound forward in the discovery process. In the worst case scenario, this could lead to a poor pharmacological and/or ADMET profile of the chosen compound. Indeed, based on Pfizer clinical trial data, Morgan and colleagues have identified that, to a large extent, failure of drugs in clinical trials could be related to poor efficacy (Morgan et al., 2012). More importantly, the authors questioned whether the target had been explored and validated correctly during the drug discovery stage. As an example, promiscuous compounds may aggregate in *in vitro* buffers and by this mechanism then cause a non-competitive inhibition, whereas they are diluted in the blood stream and the much lower concentration of the free fraction does not result in target engagement.

Since solubility has a profound impact on all the factors that are important for decision making with respect to the fate of the compound, much effort has been directed to developing tools applicable to predicting drug solubility (Delaney, 2005). Various *in vitro* assays have been developed, ranging from high throughput assays based on titrations of DMSO stock solutions and identification of the precipitations as a qualitative screen providing yes/no answers to highly accurate small scale thermodynamic measurements (Bergstrom et al., 2014). Typically these studies have focussed on solubility in pure water or in non-

* Corresponding author.
E-mail address: christel.bergstrom@farmaci.uu.se (C.A.S. Bergström).

https://doi.org/10.1016/j.ijpharm.2018.01.044
Received 10 November 2017; Received in revised form 20 January 2018; Accepted 22 January 2018
Available online 06 February 2018

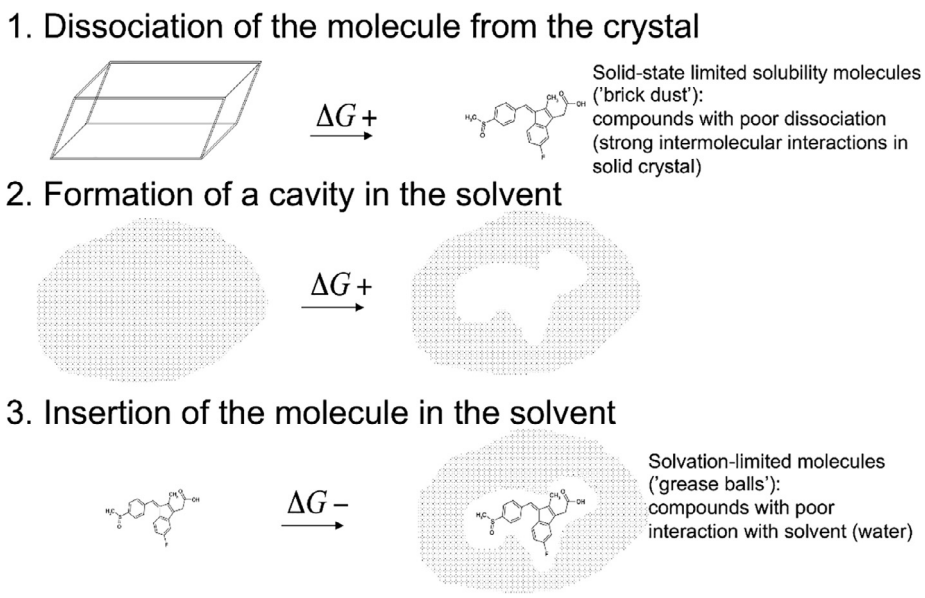

Fig. 1. The thermodynamics behind the solubility process. The drug molecule needs to dissociate from its solid form (step 1) and the tight structure of the water needs to form a cavity large enough to incorporate the drug molecule (step 2). Finally, the drug molecule is inserted into the water where it interacts with the surrounding water molecules (step 3).

complex buffers. However, over the last two decades a large number of modified media have also been developed, with the aim of better predicting the intestinal solubility of new compounds *in vivo* (Fuchs et al., 2015; Galia et al., 1998; Jantratid et al., 2008). Recently also media mimicking the interindividual variability of the composition of intestinal fluids in fasted and fed state have been explored for their influence on drug solubility (Khadra et al., 2015; Madsen et al., 2018; Perrier et al., 2017). These more biorelevant solvents include additives such as bile salts, phospholipids, cholesterol and lipids to reflect fasted and fed intestinal states. Biorelevant dissolution media have so far mainly been used in the early phases of development and have not to any major extent been investigated for their potential as a base for computational predictions. Instead, *in silico* models developed for prediction of solubility are based on the solubility of the neutral (non-ionised) compound in pure water (see e.g. examples provided in Norinder and Bergstrom (2006)). There are several reasons for taking this approach, including the complexity of the solubility process, since both dissociation from the solid state and solvation of the molecule by the solvent studied influence the final solubility (this is further discussed in Section 2, and Fig. 1). This, together with the difficulty of predicting ionisation constants for complex protolytes, and the difficulties associated with forecasting the influence of additives (such as the bile salts, phospholipids and cholesterol included in simulated intestinal fluids) on the solubility, has made pure water adjusted to a pH allowing the non-ionised species to be determined the first choice for computational modelling. Unfortunately, several of the datasets used, or large fractions thereof, do not reflect the drug-like chemical space. These datasets are repeatedly used and it is not always possible to evaluate the experimental quality of the data. For instance, one of the most repeatedly used datasets includes experimental data ranging from -11.62 to 2.77 on a log molar scale, corresponding to 2 pM and 589 M (Kühne et al., 1995). The exact quality of these data can be questioned, since the pM concentrations need a very sensitive analytical method to be trustworthy and the high end of the range corresponds to a solubility value greater than that of water in water (which is 55 M). Hence, if data like these are used in computational modelling, the accuracy of the predictions could be improved by weighting the influence of the observation by the accuracy of the experimental data, in order not to let poor experimental data influence the final model.

In this review we will discuss computational models and modelling approaches that have identified the molecular features resulting in poor aqueous solubility. We will discuss how these findings can be applied in the drug discovery and development settings both as tools for predicting solubility and as indicators of whether the compound will make it to the market after extensive formulation strategies are applied. The review will not focus on statistical approaches to solubility predictions; the interested reader is referred to several other reviews within this area, e.g. (Delaney, 2005; Johnson and Zheng, 2006; Norinder and Bergstrom, 2006; Skyner et al., 2015). Instead, we will discuss new approaches recently presented for modelling properties of importance for solubility. These include models revealing molecular features that indicate solid-state-limited versus solvation-limited solubility, models that include the current status of prediction of biorelevant solubility reflecting the intestinal environment, models for predicting crystal structure and solid-state properties such as melting point (T_m) and models that make use of the thermodynamic cycle theory. The extent to which solubility and solubility changes can be calculated from first principles is also addressed.

2. Molecular properties resulting in poor aqueous solubility

The thermodynamics behind solubility are shown in Fig. 1. In order for the molecule to dissolve in the aqueous solvent, it must be able to dissociate from its crystal lattice. This process is dependent on the intermolecular interactions between the molecules in the crystal lattice. Compounds with strong intermolecular bonds and/or complex interaction patterns with a large number of interaction points between the molecules in the crystal lattice often show a limited capacity to dissociate from the solid form. These compounds are sometimes referred to as 'brick dust' molecules, to demonstrate the poor solubility of a strong (stone-like) solid structure. Typically T_m is used to identify whether a compound shows solid-state-limited solubility (i.e. is a brick dust compound). A T_m of 200 °C has been identified as the cut-off value; for compounds that melt at higher temperatures, the crystal lattice will have a strong influence on the solubility (Bergstrom et al., 2016). For these compounds, any formulation strategy that changes the solid crystal form (e.g. using salts, cocrystals or amorphous systems) will be useful for increasing the dissolution rate and achieving a greater apparent solubility (Edueng et al., 2017; Elder et al., 2013; Kuminek et al., 2016; Taylor and Zhang, 2016). While the compound needs to dissociate from its crystal lattice, the surrounding solvent also needs to prepare for incorporating a new molecule. The larger the cavity needs