Predicting anti-cancer drugs’ dissolution in supercritical carbon dioxide

Billey Winslet 1, Victor Paxton 2, Alexandrea Rose 3, Tricia East 4

1. Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam, Netherlands.
2. Division of Hematology/Oncology, Department of Medicine, University of California, Los Angeles, CA, USA.
3. Institute de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, UMR 7154 CNRS, 75238 Paris, France.
4. Center for Health and Wellbeing, Princeton University, Princeton, NJ 08542, USA.

Corresponding author email: triciaeast1984@gmail.com

Abstract. There are many different kinds of chemotherapy or chemo drugs are used for treating cancer. All of these medicine from chemical composition, how they are prescribed and given, how useful they are in treating certain types of cancer, and the side effects they might have. Nobody can deny that all medicines to treat cancer work in the different ways. The solubility of Azathioprine, as an immunosuppressive and anti-cancer drug, in supercritical carbon dioxide (SC-CO2) was measured for the first time. Under the applied conditions in terms of pressure (120–270 bar) and temperature (308–338 K), mole fractions were obtained in the range of 0.27 × 10⁻⁵ to 1.83 × 10⁻⁵. Three types of methods including (1) two equations of states (EoSs), namely Peng-Robinson (PR) and Soave–Redlich–Kwong (SRK) with vdW2 mixing rule (2) expanded liquid theory (3) nine semi-empirical density-based models were selected to correlate the solubility data of drug. Our developed correlation presents the absolute average relative deviation (AARD) of 9.54% for predicting 316 experimental measurements. After all we show that, the most accurate correlation in the literature presents the AARD = 14.90% over the same database. Furthermore, 56.2% accuracy improvement in the solubility prediction of the anti-cancer drugs in supercritical CO2 is the primary outcome of the current study.

Keywords: solubility temperature; drug; polymer; support vector regression; kernel type; tuning techniques, anti-cancer, supercritical carbon dioxide, biomass sample; heat capacity; empirical correlation; biomass crystallinity; feature reduction

Introduction:

The SCF (Skp1-cullin-F-box proteins) complex is the largest family of E3 ubiquitin ligases that mediate multiple specific substrate proteins degradation. Two ring-finger family members RBX1/ROC1 and RBX2/RNF7/SAG are small molecular proteins necessary for ubiquitin ligation activity of the multimeric SCF complex. (5)
In order to resolve these operating and economic problems, different thermodynamic-based models (known as the equation of state) (6-9), intelligent paradigms, predictive model (10-11) and empirical correlations 15–21 is proposed to simulate different phenomena, including estimating solids solubility in SCCO2. Sondheimian et al. compared the accuracy of the Peng-Robinson (PR), Soave–Redlich–Kwong (SRK), and available empirical correlations for predicting solubility of sorafenib tosylate19, sunitinib malate20, and azathioprine(21) anti-cancer drugs in SCCO2. Performances of the PR equation of state, statistical associating fluid theory of variable range (SAFT-VR), and six empirical correlations for predicting tamsulosin solubility in supercritical CO2 have also been compared. Generally, the estimation methods of drug solubility in the SCCO2 using the equations of state (EoS) are often mathematically complicated (2), require high computations efforts, need relatively high entry information (20-22), provide high levels of uncertainty (19), and may sometimes fail(20). More precisely, they need the operating conditions, critical properties, and also drug characteristics to deliver their predictions. The least-squares support vector machines, artificial neural networks, quantitative structure–property relationships, adaptive neuro-fuzzy inference systems, wavelet transform, and dynamic simulation58–60 is some of the approaches may be used for estimating the solid solubility in supercritical carbon dioxide. Utilizing these intelligent paradigms is only possible when their structure, adjusted hyper-parameters, and performed pre-processing and post-processing stages be completely available. Despite an acceptable accuracy of these intelligent methods, some parts of their information are often missed to present, and it is hard or even impossible to be used by other researchers. The empirical correlations that only need temperature, pressure, and pure SCCO2 density to predict solid (21-27).

The ubiquitin–proteasome system (UPS) is the major proteolytic system that degrades accumulated or misfolded proteins for cellular homeostasis (27,28). It operates through the presentation of ubiquitin to the substrate proteins using a covalent modification pattern, which involves a series of multienzymes, i.e., Ubiquitin (Ub)-activating enzyme (E1), Ub-conjugating enzyme (E2) and Ub ligase (E3) (29). Among the three enzymes, the E3 ubiquitin ligases play a pivotal role in determining specificity of substrate proteolysis (30,31). Based on the structural characteristics, E3 enzyme can be divided into four categories: RING E3s, HECT E3s, U-box E3s and RBR E3s (32). The SCF multiunit complex, the most common RING E3s composing of a scaffold protein cullin1, a Ring protein (RBX1 or RBX2), an adaptor protein and a substrate receptor protein, is the largest family of E3s that promote the degradation of about 20% of UPS-regulated proteins (32,33).

The least-squares support vector regression (LS-SVR) designs to approximate the solubility temperature of drugs in polymers from polymer and drug types and drug loading in polymers. The structure of this machine learning model is well-tuned by conducting trial and error on the kernel type (i.e., Gaussian, polynomial, and linear) and methods used for adjusting the LS-SVR coefficients (i.e., leave-one-out and 10-fold cross validation scenarios). (34)

The current research briefly reviewed ten well-known and reliable empirical correlations for estimating solid solubility in supercritical CO235. After that, a universal approach based on the modified Arrhenius model is introduced to relate the anti-cancer drug solubility in SCCO2. This universal approach added a departure function to the Arrhenius-shape term to estimate the anti-cancer drug solubility in SCCO2. The predictive performance of the modified Arrhenius model and available correlations in the literature is compared using all available experimental data for solubility of anti-cancer drugs in SCCO2. 316 experimental data for solubility of sorafenib tosylate, sunitinib malate, azathioprine, busulfan, tamoxifen, letrozole, tamsulosin, capecitabine, paclitaxel, 5-fluorouracil, thymidine, and decitabine in SCCO2 are used to perform this comparison (45-46). The results show that the modified Arrhenius model improves the previously achieved accuracy in the literature by more than 56.2%. (4)
Materials and methods

The first part of this section presents the available experimental measurements for the solubility of anti-cancer drugs in supercritical CO2. The second part reviews the most well-known empirical models for correlating the solid solubility in SCCO2 to the independent variables (pressure, temperature, and pure supercritical CO2 density).

Anti-cancer drugs. As mentioned earlier, cancer is approved as the leading cause of human death worldwide15. Therefore, all aspects of anti-cancer drugs, including their solubility in the supercritical CO2 are an exciting research topic for both academic and manufacturing purposes. Based on our best knowledge, the solubility of only twelve anti-cancer drugs in the supercritical carbon dioxide were measured up to now. These anti-cancer drugs are sorafenib tosylate, sunitinib malate, azathioprine, busulfan, tamoxifen, letrozole, tamsulosin, capecitabine, paclitaxel, 5-fluorouracil, thymidine, and decitabine. Table 1 separately reports the range of pressure, temperature, supercritical CO2 density, and anti-cancer drug solubility for all the laboratory-scale studies. Furthermore, the numbers of available measurements in each research are also shown in this table. (42-46)

<table>
<thead>
<tr>
<th>CO2 (1)</th>
<th>Temperature</th>
<th>Pressure</th>
<th>CO2 density</th>
<th>Drug</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sorafenib</td>
<td>308–338</td>
<td>12–27</td>
<td>388–914</td>
<td>0.68–12.57</td>
<td>24</td>
</tr>
<tr>
<td>Sunitinib</td>
<td>308–338</td>
<td>12–27</td>
<td>388–914</td>
<td>5–85.6</td>
<td>24</td>
</tr>
<tr>
<td>Azathioprine</td>
<td>308–338</td>
<td>12–27</td>
<td>388–914</td>
<td>2.7–18.3</td>
<td>24</td>
</tr>
<tr>
<td>Busulfan</td>
<td>308–338</td>
<td>12–40</td>
<td>383–971</td>
<td>32.7–865</td>
<td>32</td>
</tr>
<tr>
<td>Tamoxifen</td>
<td>308–338</td>
<td>12–40</td>
<td>383–971</td>
<td>18.8–989</td>
<td>32</td>
</tr>
<tr>
<td>Letrozole</td>
<td>318–348</td>
<td>12–36</td>
<td>319–922</td>
<td>1.6–85.1</td>
<td>20</td>
</tr>
<tr>
<td>Tamsulosin</td>
<td>308–338</td>
<td>12–27</td>
<td>384–914</td>
<td>0.18–10.13</td>
<td>24</td>
</tr>
<tr>
<td>Capecitabine</td>
<td>308–348</td>
<td>15.2–35.4</td>
<td>477–955</td>
<td>2.7–158.8</td>
<td>40</td>
</tr>
<tr>
<td>Paclitaxel</td>
<td>308–328</td>
<td>10–27.5</td>
<td>654–915</td>
<td>1.2–6.2</td>
<td>21</td>
</tr>
<tr>
<td>5-</td>
<td>308–328</td>
<td>12.5–25</td>
<td>541–901</td>
<td>3.8–14.6</td>
<td>18</td>
</tr>
<tr>
<td>Thymidine</td>
<td>308–328</td>
<td>10–30</td>
<td>325–928</td>
<td>1.2–8</td>
<td>25</td>
</tr>
<tr>
<td>Decitabine</td>
<td>308–338</td>
<td>12–40</td>
<td>383–971</td>
<td>28.4–1070</td>
<td>32</td>
</tr>
</tbody>
</table>

Since the collected experimental data covers different polymer–drug systems, molecular weights (Mw) of the drug and polymer are selected to help LS-SVRs discriminate between the behavior of various drugs and polymers during the modeling phase. In summary, the current study includes 16 drugs, 13 polymers, a drug load of 1 to 100 weight percent (wt%) in polymers, and a solubility temperature of 30 to 252.7 ◦C. It should be mentioned that the composition of polymer–drug mixtures can vary from low (~0 wt%) to high (~100 wt%) dosages of a drug. Indeed, these points indicate the pure polymer and pure drug, respectively. When the drug loading is 100 wt% (i.e., pure drug), the melting temperature is considered as the solubility temperature (47).

Machine learning, deep learning, feature selection, and decision-making techniques have a broad range of applications for implementing either classification or approximation tasks in different fields of daily life, science, and technology. Support vector machine and its derivation (i.e., least-squares support vector regression have recently gained great attention.(48-53)
Results and discussion

This section presents the idea of developing the modified Arrhenius correlation, adjusts its unknown coefficients, and compares its accuracy with other available correlations. The next part of this section is devoted to the performance analysis of the modified Arrhenius correlation using different graphical methods. Finally, the modified Arrhenius correlation is employed to monitor the effect of operating conditions on the anti-cancer drug solubility in SCCO2.

Developing the modified Arrhenius correlation. The massive data processing stages are performed on the experimental values of solubility of each drug in SCCO2 to reach a general form of the proposed correlation as follows:

\[ y_2 = \text{Arrhenius term} + \text{departure function} \]  

Equation (1) states that the anti-cancer drug solubility in the SCCO2 can be accurately estimated by combining an Arrhenius term and a departure function.

At this stage, it is necessary to clarify how the pre-exponential and exponential parts of the Arrhenius term are related to the influential variables. Then, the departure function incorporates to reduce the deviation between the Arrhenius term predictions and experimental measurements.

Spearman and Pearson are two well-known relevancy discovery scenarios in the field of data processing\textsuperscript{62}. They introduce the relevancy between a pair of feature-response variables by a factor in the range of \(-1\) to \(+1\). The minus, zero, and positive factors correspond with indirect dependency, no-relation, and direct dependency, respectively. The strength of either direct or indirect relevancy increases by increasing the magnitude of factors. Furthermore, the higher absolute value of the Spearman than the Pearson factor confirms that the non-linear relationship is stronger than the linear one and vice versa. (54-58)

Figure 1 exhibits the values of relevancy factor between anti-cancer drug solubility and pressure, temperature, and pure SCCO2 density. This figure confirms that direct relationships exist between the response and all feature.
Figure 1. Relevancy between the solubility of anti-cancer drugs in supercritical CO₂ and temperature, pressure, and carbon dioxide density.

Figure 2. Dependency of sorafenib tosylate solubility in the supercritical CO₂ on the isobaric variation of temperature (the cartesian coordinate).
Comparative analysis. This section compares the uncertainty in the predictions of the modified Arrhenius model and available correlations in the literature for solubility of anti-cancer drugs in SCCO2. The prediction uncertainty of all considered empirical correlations is measured in terms of the AARD% and reported in Table 4. First of all, it is better to clarify that the highlighted cells (gray color) are calculated in the present study, and the clean cells are those reported in the literature. As mentioned earlier, the associated coefficients for calculating this AARD% are presented in Supplementary file. The cells shown by the bold font are the smallest AARD% (the best results) obtained for estimating a specific anti-cancer drug in supercritical CO2. It is obvious that the modified Arrhenius correlation provides the most accurate results for solubility of six out of twelve anti-cancer drugs in SCCO2 (i.e., sorafenib tosylate, sunitinib malate, azathioprine, tamsulosin, 5-fluorouracil, thymidine). On the other hand, the derived correlation by Bian et al. predicts the solubility of busulfan, tamoxifen, and decitabine in supercritical CO2 with the highest accuracy. Finally, the Garlapati and Madras, Sodeifian et al., and Tan et al. correlations provide the most accurate predictions for only one anti-cancer drug. It can be readily deduced that the proposed correlation in the current study not only presents the most accurate predictions for six anti-cancer drugs, it also has two second and three third ranks. The worst accuracy of the modified Arrhenius correlation is associated with capecitabine solubility in the SCCO2 (i.e., the fourth rank). The proposed correlation by Bian et al. with the three first, two second, four third, one fourth, and one ninth ranks is the next reliable model for the given task. On the other hand, the proposed correlations by Gordillo, Jouyban et, and Tan et al. have the highest levels of uncertainty, respectively. (49-59)

The variation of solubility temperature of D-Mannitol in PVP K12, PVP K15, and VP dimer has been plotted in Figures. This figure covers both experimental and modeling Tsoldrug load profiles. It can be viewed that the PVP K12/D-Mannitol and VP dimer/D-Mannitol systems have the highest and lowest
solubility temperature, respectively. The acceptable performance of the LS-SVR for predicting the equilibrium behavior of polymer/D-Mannitol systems can be approved by the observed MARDP value of 1.65 (PVP K12/D-Mannitol), 2.84 (PVP K15/D-Mannitol), and 2.34 (VP dimer/D-Mannitol). In addition, both the experimental data and LS-SVR curves in Figures show that the effect of the polymer type on solubility temperature is minor. We recall from Figure 1 that the molecular weight of polymers has a minor effect on the drug solubility temperature. On the other hand, PVP K12, PVP K15, and VP dimer are polyvinyl pyrrolidone-based polymers with different molecular weights. Therefore, it is expected that the solubility temperature of D-Mannitol in these polymers is almost equal. (34)

**Conclusion:**

A combination of the Arrhenius-shape and departure functions is proposed to correlate the anti-cancer drug solubility in the supercritical carbon dioxide. The pre-exponential part of the Arrhenius-shape term is linearly related to the temperature and carbon dioxide density, and its exponential part inversely relates to the pressure. The departure function is directly related to the natural logarithm of the carbon dioxide density to the temperature ratio. The developed correlation outperformed all well-known literature equations for predicting the solute solubility in supercritical carbon dioxide. The modified Arrhenius correlation provided the AARD = 9.54% and R2 = 0.98479 for estimating all experimental datasets in the literature. In contrast, the most accurate correlation in the literature (i.e., Bian et al. correlation) showed the AARD = 14.90% for predicting the considered database. It is possible to improve predicting accuracy of anti-cancer drug solubility in supercritical CO2 by more than 56% using the developed correlation in this study. The relevancy analysis exhibited that anti-cancer drug solubility in supercritical CO2 increases by increasing either pressure and temperature. Furthermore, it is found that less than 7.5% of the literature data are suspect information, and the remaining 92.5% are valid measurements. (4)

**References:**