

Predicting sulfanilamide solubility in the binary mixtures using a reference solvent approach

Przewidywanie rozpuszczalności sulfanilamidu w mieszaninach dwuskładnikowych z wykorzystaniem metodyki rozpuszczalnika referencyjnego

Piotr Cysewski^{A–F}, Maciej Przybyłek^{B–F}, Tomasz Jeliński^{B–F}

Department of Physical Chemistry, Faculty of Pharmacy, Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń, Poland

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation; D – writing the article; E – critical revision of the article; F – final approval of the article

Polymers in Medicine, ISSN 0370-0747 (print), ISSN 2451-2699 (online)

Polim Med. 2024;54(1):27–34

Address for correspondence

Piotr Cysewski
E-mail: piotr.cysewski@cm.umk.pl

Funding sources

None declared

Conflict of interest

None declared

Received on November 19, 2023

Reviewed on December 19, 2023

Accepted on January 5, 2024

Published online on February 2, 2024

Cite as

Cysewski P, Przybyłek M, Jeliński T. Predicting sulfanilamide solubility in the binary mixtures using a reference solvent approach. *Polim Med.* 2024;54(1):27–34.
doi:10.17219/pim/178284

DOI

10.17219/pim/178284

Copyright

Copyright by Author(s)

This is an article distributed under the terms of the Creative Commons Attribution 3.0 Unported (CC BY 3.0) (<https://creativecommons.org/licenses/by/3.0/>)

Abstract

Background. Solubility is a fundamental physicochemical property of active pharmaceutical ingredients. The optimization of a dissolution medium aims not only to increase solubility and other aspects are to be included such as environmental impact, toxicity degree, availability, and costs. Obtaining comprehensive solubility characteristics of chemical compounds is a non-trivial and demanding process. Therefore, support from theoretical approaches is of practical importance.

Objectives. This study aims to examine the accuracy of the reference solubility approach in the case of sulfanilamide dissolution in a variety of binary solvents. This pharmaceutically active substance has been extensively studied, and a substantial amount of solubility data is available. Unfortunately, using this set of data directly for theoretical modeling is impeded by noticeable inconsistencies in the published solubility data. Hence, this aspect is addressed by data curation using theoretical and experimental confirmations.

Materials and methods. In the experimental part of our study, the popular shake-flask method combined with ultraviolet (UV) spectrophotometric measurements was applied for solubility determination. The computational phase utilized the conductor-like screening model for real solvents (COSMO-RS) approach.

Results. The analysis of the results of solubility calculations for sulfonamide in binary solvents revealed abnormally high error values for acetone-ethyl acetate mixtures, which were further confirmed with experimental measurements. Additional confirmation was obtained by extending the solubility measurements to a series of homologous acetate esters.

Conclusions. Our study addresses the crucial issue of coherence of solubility data used for many theoretical inquiries, including parameter fitting of semi-empirical models, in-depth thermodynamic interpretations and application of machine learning protocols. The effectiveness of the proposed methodology for dataset curation was demonstrated for sulfanilamide solubility in binary mixtures. This approach enabled not only the formulation of a consistent dataset of sulfanilamide solubility binary solvent mixtures, but also its implementation as a qualitative tool guiding rationale solvent selection for experimental solubility screening.

Key words: *in silico*, solubility, sulfanilamide, binary solvents, COSMO-RS

Streszczenie

Wprowadzenie. Rozpuszczalność jest jedną z najbardziej fundamentalnych charakterystyk fizykochemicznych substancji aktywnych farmaceutycznie. Optymalizacja medium rozpuszczalnikowego jest złożonym problemem obejmującym nie tylko wyznaczenie samej rozpuszczalności ale również takie aspekty jak wpływ na środowisko, toksyczność, dostępność i koszt zakupu oraz utylizacji rozpuszczalników. Ponieważ uzyskanie kompleksowej charakterystyki rozpuszczalności nie jest trywialna, a często wiąże się z kosztownym i czasochłonnym procesem, badania teoretyczne stanowią istotne wsparcie.

Cel pracy. Celem pracy jest ilościowa ocena efektywności obliczania rozpuszczalności w mieszaninach dwuskładnikowych w oparciu o metodologię referencyjnego rozpuszczalnika. Z uwagi na duże zainteresowanie tym związkiem dostępna jest znaczna ilość danych doświadczalnych umożliwiających szczegółową analizę teoretyczną. Niestety, wykorzystanie tego zestawu bezpośrednio do modelowania teoretycznego jest utrudnione przez zaobserwowane niespójności. Ten problem został szczegółowo przeanalizowany w oparciu o rozważania teoretyczną oraz przeprowadzone eksperymenty.

Materiał i metody. Część doświadczalna badań obejmowała równowagowy pomiar rozpuszczalności z wykorzystaniem metody spektrofotometrycznej w zakresie UV. Badania teoretyczne zostały zrealizowane z wykorzystaniem modelu COSMO-RS.

Wyniki. Analiza wyników przeprowadzonych obliczeń rozpuszczalności sulfonamidu w binarnych rozpuszczalnikach ujawniła anomalnie wysokie wartości błędów dla przypadku mieszanin aceton-octan etylu, co zostało zweryfikowane poprzez nowe pomiary eksperymentalne. Dodatkowego potwierdzenia dostarczyło rozszerzenie pomiarów rozpuszczalności o serie homologicznych estrów octanowych. Ponowne obliczenie rozpuszczalności w oparciu o metodykę rozpuszczalnika referencyjnego doprowadziło do znaczącej mniejszej rozbieżności pomiędzy obliczonymi oraz zmierzonymi wartościami ułamków molowych.

Wnioski. Niniejszy raport porusza kluczową kwestię zapewnienia spójności danych rozpuszczalności, które częstokroć są wykorzystywane w rozważaniach teoretycznych, takich jak interpretacja rozpuszczalności w oparciu o modele empiryczne czy półempiryczne oraz termodynamicznej charakterystyce procesów rozpuszczania. Zastosowanie metodyki obliczania rozpuszczalności w mieszaninach dwuskładnikowych na podstawie danych dla czystych rozpuszczalników pozwoliło nie tylko na identyfikację niespójności danych ale również na potencjalne badania przesiewowe.

Słowa kluczowe: *in silico*, rozpuszczalność, sulfonamid, COSMO-RS

Background

Sulfonamides are a highly important class of active pharmaceutical ingredients (APIs). This group of antimicrobial substances has a broad spectrum of applications, including the treatment of various bacterial and fungal infections and diabetes therapy.^{1,2} Sulfanilamide (SA; $C_6H_8N_2O_2S$; CAS:63-74-1, DrugBank: DB00259) is a precursor of sulfonamides used as topical anti-infectives.³ Furthermore, besides its conventional applications, SA has been employed in the synthesis of functional polymers with potential biomedical and pharmaceutical significance.^{4–6}

The solubility of a particular API is a fundamental physicochemical property routinely determined at the early stages of drug development. Moreover, the optimization of the dissolution media is often required to fulfill the criterion of solubilization effectiveness and additional requirements such as ecological issues, toxicity, availability, and costs. However, due to the enormously large solvent space, the practical realization of this step poses serious challenges. Hence, multiple theoretical approaches have been developed for the estimation of the properties of saturated solutions. Among the many available methods, the conductor-like screening model for real solvents (COSMO-RS)^{7,8} is one of the most ambitious theoretical approaches, aiming to determine the comprehensive characteristics of liquids exclusively from molecular structure. This particular approach harnesses the first principle

of quantum chemistry computations augmented with statistical thermodynamics. This 2-step procedure results in full temperature-dependent thermodynamics for bulk systems, even with complex compositions. A unique feature is its theoretical framework that explicitly includes the structural and energetic diversity of the liquid systems by allowing the molecular representation by the set of the most probable conformers. There is considerable evidence of the successful application of this method to a wide range of chemical problems. Some reported examples include solubility modeling of rutin in natural deep eutectic solvents (NADES),⁹ halogenated hydrocarbons in ionic liquids¹⁰ or cellulose in ionic liquids.¹¹ Notably, according to a recent report by Klajmon,¹² the COSMO-RS model is generally significantly more accurate for modeling the solubility of various pharmaceuticals in conventional solvents compared to the popular perturbed-chain statistical associating fluid theory (PC-SAFT) method.

Importantly, COSMO-RS is a theory of bulk liquids, and the characteristics of the solid-liquid equilibria (SLE) require taking into account the Gibbs free energy values of fusion of the solute (ΔG_{fus}). This thermodynamic property considers the transfer of the compound from the crystalline state into an unordered liquid and is typically determined experimentally using differential scanning calorimetry (DSC) measurements, including melting temperature (T_m), heat of fusion (ΔH_{fus}), and heat capacity change upon melting. Although many solids have been fully

characterized in these respects,^{13,14} such thermodynamic data are not available for numerous active pharmaceutical ingredients. The reasons for this are related to the thermal instabilities and decomposition below the melting point. Additionally, the complex polymorphic behavior, which is often observed for many solids, requires further determination of thermodynamic equilibria between the different polymorphs. Occasionally,¹⁵ the polymorph dissolved under the experimental conditions of solubility measurements differs from the crystal form stable at the melting point. Also, sulfanilamide is considered a problematic solid. In the Cambridge Structural Database (CSD), sulfanilamide solid structure is deposited in 4 sets of records documenting distinct polymorphic forms, namely α (Pbca), β (P21/c), γ (P 21/c), and δ (Pbca). It has been reported^{16,17} that both α and β forms during the heating process undergo a transition to the γ form, which is the only existing form under melting conditions. The 4th form is metastable and does not melt. The β -polymorph is the most stable one under ambient conditions and is usually commercially available.¹⁸ This complicates the theoretical computations of solubility since the experimental fusion data are unavailable. Moreover, proper characteristics of the fusion require the determination of the temperature-related heat capacities of the solid and liquid states, which often exceed the melting conditions. Therefore, the melting and fusion terms are often distinguishable. The former is used to depict the solid-liquid phase transition at the melting state, while the latter is reserved for solid transfer into a sub-cooled liquid state at any other temperature. Furthermore, the possibility of solvate formation should be considered. Fortunately, for sulfanilamide, no solvent inclusion occurs after crystallization.^{16,19} The reported DSC thermograms¹⁹ provide a consistent picture where the thermograms' shapes are preserved irrespective of the recrystallization from many solvents. This is evident by the observed small endothermic peak corresponding to the polymorphic transition and the large endothermic peak associated with the melting of the γ form.

The above information suggests that solubility computation is not a straightforward and trivial task, even in conjunction with the first-principle approach. Hence, many simplifications have been proposed by some authors, but criticized by others.²⁰ However, the lack of consensus regarding the appropriate way of including fusion thermodynamics in solubility predictions should not be a prohibiting factor. Indeed, in the COSMOtherm program (Dassault Systèmes, Biovia, San Diego, USA), which practically implements the COSMO-RS theory, it is possible to compute solid solubility by indirectly defining necessary fusion data. This is achieved by providing experimental data for a given solute in a given solvent or mixture at a given temperature as input data for computing solubility in different media. This approach refers to the reference solubility method. Furthermore, using the value of ΔG_{fus} determined in such a manner reproduces the solubility of the solutes

given as the reference. Hence, this method is very flexible, providing the reference solubility at conditions different from those of the solubility predictions. Particularly, it is possible to generate several reference solubility values or a variety of temperatures. The former option is utilized in our study by providing the 2 reference solubility data, which were experimentally determined for neat solvents for computing the solubility in the binary mixture at the given temperature. This reference solvent procedure was previously used with varying effectiveness, depending on the system under consideration.²¹ In general, the proper selection of a reference solvent can provide a reasonable solubility estimate. Such a solvent is commonly referred to as a consonance solvent.^{15,21} It is interesting to note that this approach is not utilized for the practical prediction of solubility in solvent mixtures.

Herein, we analyzed the accuracy of the aforementioned approach for sulfanilamide dissolved in a variety of binary solvent mixtures. This particular solute was extensively studied, and the available solubility data were large enough for theoretical analysis.

Materials and methods

Analytical grade sulfanilamide ($\geq 98\%$, SA, CAS: 63-74-1), ethyl acetate (99.5%, CAS: 141-78-6), n-propyl acetate ($\geq 99.5\%$, CAS: 109-60-4), n-butyl acetate (99.7%, CAS: 123-86-4), and methanol ($\geq 99.8\%$, CAS: 67-56-1) were obtained from Sigma-Aldrich (Saint Louis, USA) and used as received. Nitrogen (99.999%) for DSC analyses was supplied by Linde (Warsaw, Poland).

Solubility measurements

In this study, the shake-flask solubility determination procedure was applied. It is worth noting that the protocol has been previously employed for various organic compounds and validated in literature data.^{19,22–27} To determine the concentration of the saturated solutions of sulfanilamide in the considered set of solvents, the samples containing undissolved excess of active substances were incubated with mixing (60 rpm) at 25°C. The mixtures were prepared by dissolving sulfanilamide in a solvent in glass tubes (10 mL). Then, the test tubes were placed in an incubator (Orbital Shaker ES-20/60; Biosan, Riga, Latvia). After allowing the undissolved precipitate to settle for 1 h at 25°C, the mixture and the sediment were separated by filtration using a syringe equipped with a 0.22 μm PTFE filter. Spectrophotometric measurements were taken using 0.1 mL of the filtered solution, which was immediately diluted with 2 mL of methanol. Rapid dilution of the samples was essential to avoid crystallization during subsequent analytical procedures for concentration determination. For the same reason, i.e., to avoid crystallization, the equipment employed at the filtration and dilution steps

(syringes, filters, test tubes for filtrates, and pipette tips) was preheated at 25°C before use. For density determination, 1 mL of the filtrate was immediately transferred using an automatic pipette (Eppendorf Reference 2; Eppendorf, Hamburg, Germany) with a preheated tip to the tared 10 mL glass flask and weighted using an analytical balance (RADWAG AS 110 R2.PLUS; Radwag, Radom, Poland).

The quantification of SA in filtrates was ascertained using a spectrophotometric method ($\lambda_{\max} = 262 \text{ nm}$). To establish the calibration curve, the absorbance values were determined for a series of dilutions of the stock sulfanilamide solution (1.78×10^{-5} – $6.68 \times 10^{-5} \text{ M}$). All spectrophotometric measurements were performed using the A360 UV-VIS device (AOE Instruments, Shanghai, China).

Thermal analysis

After the shake-flask procedure, SA solid residues were air-dried and subjected to DSC analysis using a DSC 6000 PerkinElmer calorimeter (PerkinElmer, Waltham, USA). Heating rate was set to 5 K/min, while the nitrogen flow was 20 mL/min. The samples were measured in standard aluminum pans. The heat flow and temperature calibration were performed using zinc and indium standards supplied by the DSC device manufacturer.

Solubility dataset

An extensive literature review revealed sulfanilamide solubility of 18 different binary mixtures, including water-propylene glycol,²⁸ 1,4-dioxane-water,²⁹ water-methanol,³⁰ acetone-ethanol,³¹ acetone-methanol,³¹ acetone-toluene,³¹ methanol-toluene,³² ethanol-toluene,³² methanol-chloroform,³² ethanol-chloroform,³² methanol-ethanol,³² 1,4-dioxane-water,¹⁹ ethanol-water,¹⁹ dimethyl sulfoxide (DMSO)-water,¹⁹ dimethylformamide (DMF)-water,¹⁹ acetonitrile-water,¹⁹ and 4-formylmorpholine-water.¹⁹

This collection included 13 solvents used for binary mixtures' preparation in broad ranges of compositions and temperatures. The total number of datapoints was equal to $n = 1171$.

COSMO-RS computations

The COSMO-RS quantum chemistry approach^{7,8} was used for solubility predictions, as implemented in the COSMOtherm software.³³ A standard procedure for conformer generation was adopted, using the COSMOconf program (Dassault Systèmes, Biovia, San Diego, USA)³⁴ at the highest parametrization level available (BP_TZVPD_FINE_21.ctd). The most stable conformer is shown in Fig. 1.

The solubility computations relied on a thermodynamic equation defining the general solid-liquid equilibria in the following form (Equation 1):

$$\ln(y_i \cdot x_1^{\text{id}}) = \frac{\Delta H_{\text{fus}}}{R} \left(\frac{1}{T_m} - \frac{1}{T} \right) - \frac{1}{RT} \int_{T_m}^T \Delta C_{p,\text{fus}} dT + \frac{1}{R} \int_{T_m}^T \frac{\Delta C_{p,\text{fus}}}{T} dT \quad (1)$$

where the left side defines the solute activity $a_i = y_i \cdot x_1^{\text{id}}$ at saturated conditions, T_m stands for melting temperature, ΔH_{fus} represents heat of fusion, $\Delta C_{p,\text{fus}}$ denotes heat capacity change upon melting, and R is the gas constant. The fusion data indispensable for direct solubility computations in the COSMOtherm were taken as the average of available values,¹³ namely heat of fusion $\Delta H_{\text{fus}} = 23.40 \pm 0.38 \text{ kJ/mol}$ (average of reported 23.28 kJ/mol,³⁵ 23.30 kJ/mol,^{35,36} 24.02 kJ/mol,³⁷ 23.0 kJ/mol,³⁸ and 23.42 kJ/mol³²) and melting point $T_m = 437.3 \pm 1.84 \text{ K}$ (mean value of reported 435.35,³⁵ 435.4 K,^{35,36} 439.3K,³⁷ 438.7 K,³⁸ and 437.7 K³²). The application of this equation for any solid solute is often simplified by assuming the temperature independence of $\Delta C_{p,\text{fus}}$. There are 2 common alternatives, which suffer

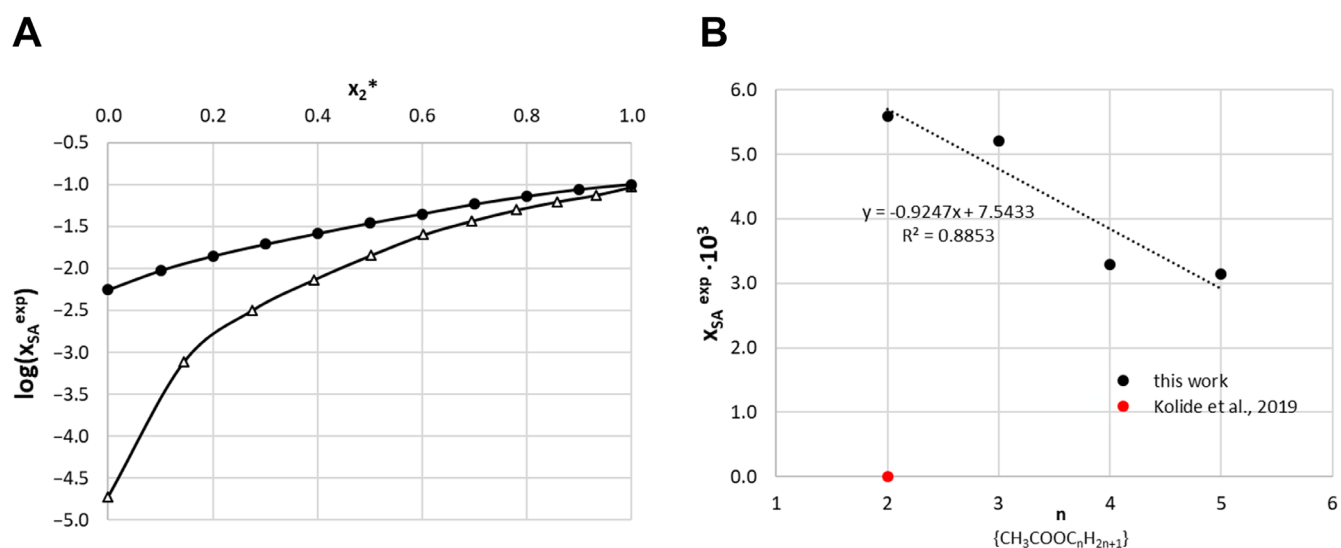


Fig. 1. A graphical representation of the structure and charge density distribution of the most stable conformer of sulfanilamide

from different inaccuracies depending on the studied system. The crudest simplification ignores heat capacity by setting $\Delta C_{p,\text{fus}} = 0$, which is often quite acceptable and was denoted as COSMO-RS in our study. Alternatively, one can assume $\Delta C_{p,\text{fus}} \approx S_{\text{fus}} = \Delta H_{\text{fus}}/T_m$ referred to below as CSOMO-RS(2). It is worth mentioning that these fusion data were used for computing the solubility in binary mixtures only for comparisons with reference solubility computations. The latter did not require fusion thermodynamics, since they were computed using the COSMOtherm program from the provided reference solubility data. The solubility values were computed by completely solving the SLE problem, utilizing a sequence of the following options in the input file “SLESOL solub screening”. For every reference solvent computation, actual values of neat solvent solubility were declared using the “ref_sol_x_log10” option along with the binary mixture composition, “x_ref_sol”, and temperature, “T_ref_sol”, keywords.

Results and discussion

The solubility of sulfanilamide in binary solvent mixtures was computed using the reference solvent approach. The experimental mole fractions of the saturated solutions were collected from the literature for systems listed in the methodology section. The computed values of SA solubility for the whole dataset are shown in Fig. 2, which comprises of 4 sets of results. The 1st set consisted of the prediction of SA solubility in binary mixtures derived directly from solving SLE within the COSMO-RS framework, and was displayed as the distribution of gray crosses. It was visible that this method provided a very raw estimate of the experimental solubility data.

However, reports have stated that COSMO-RS performs poorly in solubility prediction for various systems,^{22,39} although there are cases where the computed values are in good agreement with the experimental ones.²⁴ This was not the case for SA dissolved in the binary solvent mixtures, and we did not rely on such values for screening purposes, due to the back-computed values suffering from serious inaccuracies. The value of the mean absolute percentage error (MAPE) of mole fractions was as high as 1,310% (or 100% of the mean absolute logarithmic error), and cases arose that predicted value deviates by 3 orders of magnitude with respect to the measured ones. This was found for SA dissolution in chloroform, DMSO, DMF, and 4-formylmorpholine. Fortunately, the reference solvent procedure significantly increased the accuracy of the overall performance, since MAPE was reduced to about 40%. Although this was not a qualitative prediction, it offered rational estimates directing solvent selection for further experiments, if screening was needed. This was a very practical result, since only about 10% of the systems were studied experimentally within the whole solubility space formed by all possible combinations of the 13 neat solvents involved in the collected dataset of SA. Additionally, by performing measurements of SA in new neat solvents, the procedure enabled the estimation of solubility in virtually any combinations forming new binary mixtures.

Apart from the above determinations, additional interesting aspects emerged after performing reference solvent computations. Unexpected and anomalous behavior of acetone-ethyl acetate binary systems can be observed.³¹ This occasionally occurs where the reported solubilities were incongruent²³ and should be used with reserve. Herein, the highly inaccurate values of SA solubility in the mentioned binary mixture computed using

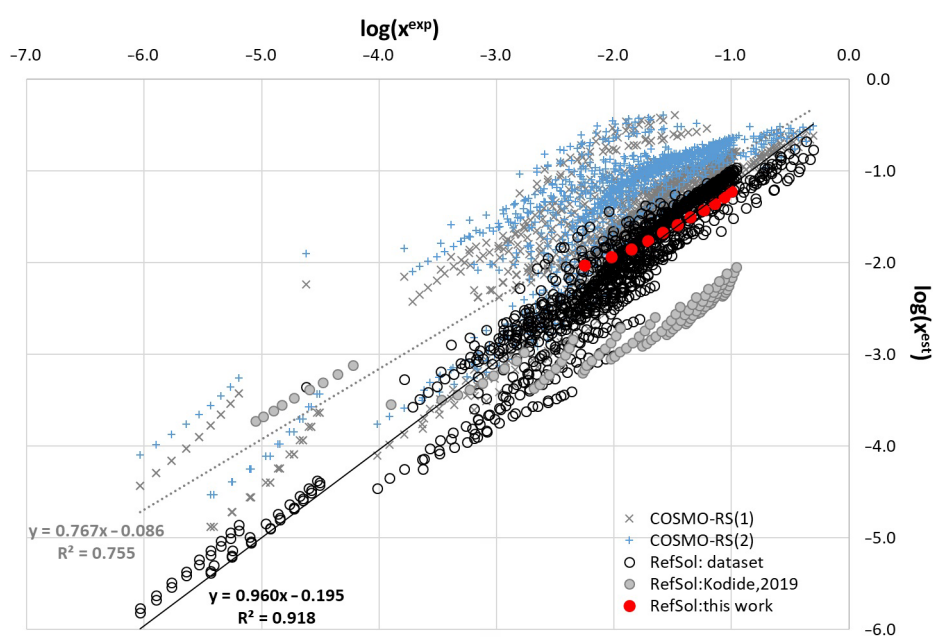


Fig. 2. Distribution of the computed values of sulfanilamide solubility in 18 binary mixtures confronted with experimentally determined ones. The circles represent the results of reference solubility computations and are denoted in the legend as the RefSol series. The distribution obtained for the whole dataset (open black circles) is overlapped with data described by Kodide et al.³¹ (gray circles) and measurements from the present study work (red circles). Additionally, results of the solubility computed using the COSMO-RS approach are provided, assuming $\Delta C_p = 0$ (gray crosses, COSMO-RS(1)) and $\Delta C_p = \Delta H_{\text{fus}}/T_m$ (blue pluses, COSMO-RS(2)). In both cases, the following fusion data are used: $\Delta H_{\text{fus}} = 23.40$ kJ/mol and $T_m = 437.3$ K

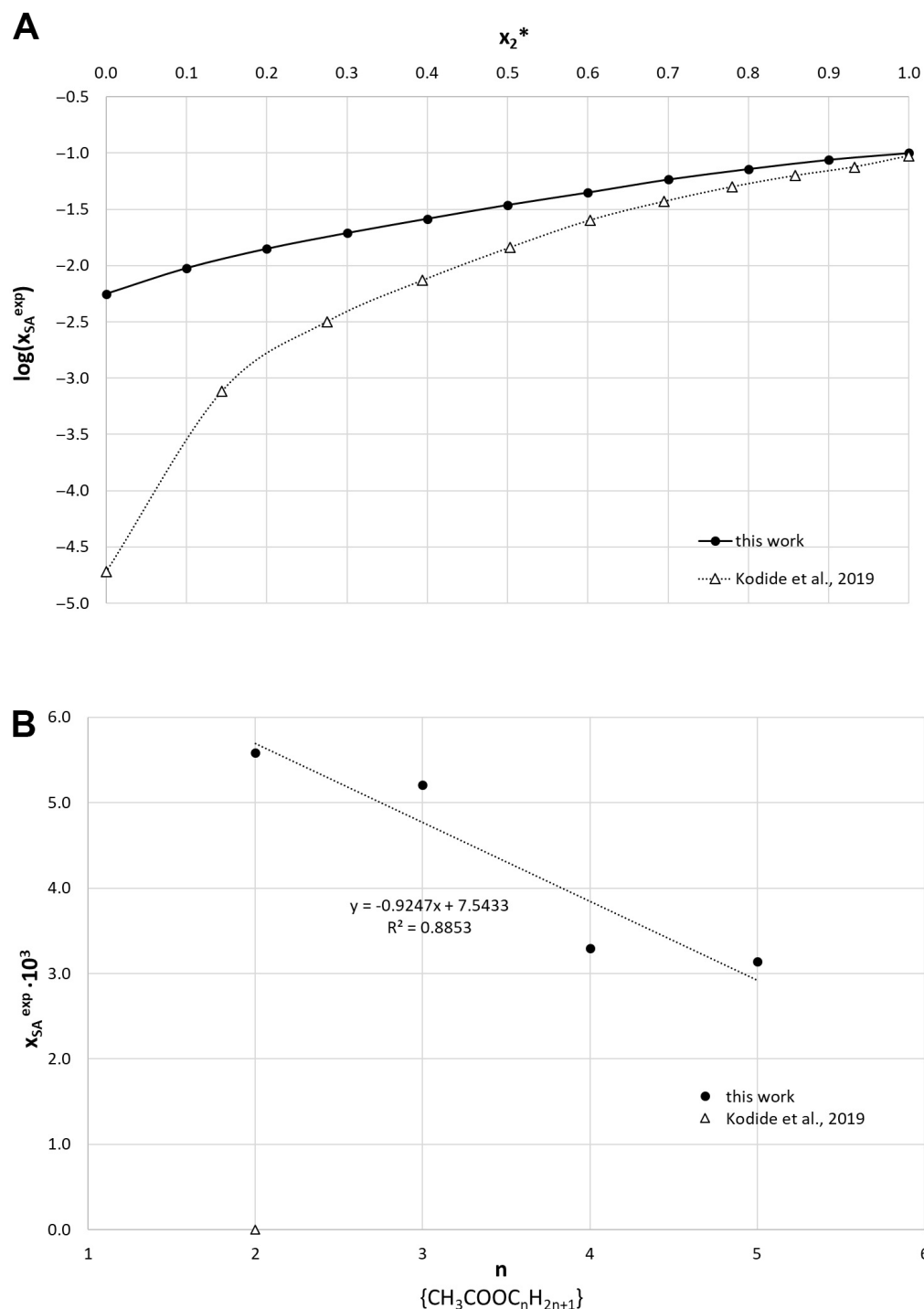


Fig. 3. Results of solubility determination ($T = 298.15K$) of sulfanilamide (SA) in (a) acetone (2) – ethyl acetate (3) as a function of acetone composition in solute-free solutions and (b) homology series of acetate esters, where n denotes carbon atoms of the alcohol part (2 for ethyl acetate, 3 for n-propyl acetate, 4 for n-butyl acetate, and 5 for n-amyl acetate)

the reference solvent approach are highlighted in Fig. 2 as gray circles distribution. Reports suggest a similarity in the solubility of SA in toluene and ethyl acetate.³¹ However, due to the differences in the polarity of both solvents, the opposite is expected. Notably, Asadi et al.⁴⁰ showed relatively high deviations for SA solubility in the acetone-ethyl acetate system if interpreted using different solubility equations. Hence, extremely inaccurate predictions using reference solvent approaches for this system should not be attributed to theoretical issues. To resolve this problem and univocally address the observed discrepancy, 2 series

of new measurements were performed. In the 1st trial, the solubility of SA was measured in 9 compositions of acetone and ethyl acetate at room temperature. To further confirm the solubility trend, the solubility of SA was determined in a series of ester homologs. The results of these 2 series of experiments are provided in Fig. 3. The obtained data highlighted the problem of the reported SA solubility in acetone-ethyl acetate binary mixtures and provided a comprehensive and convincing resolution.

First, the left panel confirms the serious discrepancy between our measurements and those reported

in the literature.³¹ Interestingly, the application of the reference solvent method to these new solubility data resulted in a very accurate prediction. Indeed, as shown in Fig. 2, the red circles representing the values for SA in acetone-ethyl acetate mixtures were the same as the overall trend and similar to the experimental data. Hence, this conducted analysis confirmed the higher reliability of the new measurements collected in Fig. 3A over those previously reported.³¹ To further support our observations, additional confirmation was offered by the results of new experimental solubility measurements (Fig. 3B). It was reasonable to expect that for a homolog series of solvents, a type of smooth trend of solubility should be observed. As shown in Fig. 3B, this was the case for a series of 4 acetate esters. A linear trend was found between the number of carbon atoms in the alcohol part and SA solubility in the corresponding acetate esters. The literature values were beyond that of the measured trend.

Undoubtedly, factors related to the characteristics of the solid, such as purity and polymorphism, have an impact on the results of solubility measurements. Therefore, calorimetric studies could provide valuable insight. Importantly, DSC measurement results of the sediments collected after the shake-flask procedure in neat solvents did not differ significantly from the pure sulfanilamide measured in this study (Fig. 4) and those reported in the literature.³¹ For all thermograms, a small peak corresponding to the polymorphic transformation and an intense γ form melting peak was observed. The onset values for the polymorphic and solid-liquid phase transitions of pure SA were

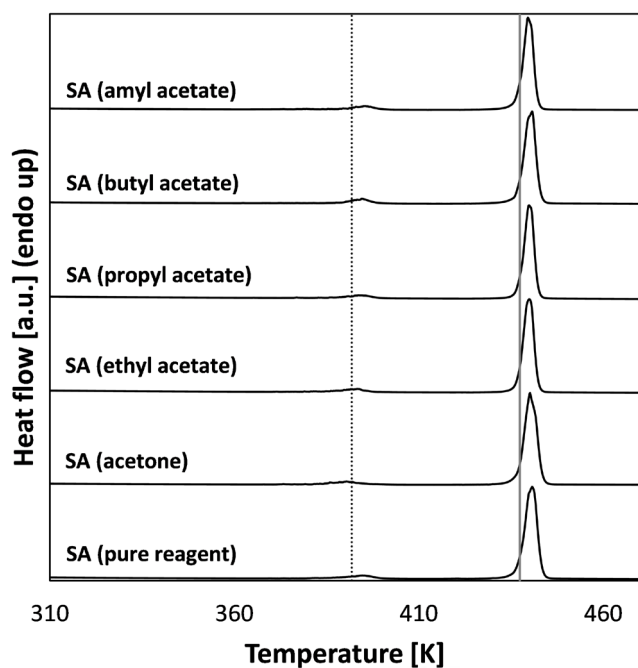


Fig. 4. Differential scanning calorimetry (DSC) analysis of solid residues from selected shake-flask experiments and pure sulfanilamide (SA). The dotted black and solid gray lines denote the onset values of the polymorphic transformation, and γ forms the melting point

391.84 K and 437.41 K, with corresponding enthalpy values of 1.89 kJ/mol and 24.17 kJ/mol. Notably, the melting point obtained in our study was similar to that reported by Kodide et al. for SA powder used for solubility measurements ($T_m = 437.78$ K).³¹

Conclusions

In this report, we addressed an important problem of solubility data coherence. The accuracy of this data is crucial as it is often used for further theoretical investigations, such as fitting parameters of empirical or semi-empirical solubility equations, solubility thermodynamics interpretation or more sophisticated inquiries, such as local composition determinations. Furthermore, the development of non-linear models using machine learning methodology also requires a reliable and consistent dataset for parameter hypertuning. Hence, data curation for solubility collections is a vital and valuable step.

This study displayed a straightforward methodology for the curation of certain subsets of the solubility data available in the published resources. The well-known saying “garbage in – garbage out” was exemplified for the solubility computations using the reference solvent approach. The accuracy of computed values in mixed solvents was inherently dependent on the reliability of solubility in neat solvents, as shown for sulfanilamide dissolved in binary mixtures of acetone and ethyl acetate. The replacement of suspicious solubility in neat solvents promoted an increase in the overall accuracy of solubility computations. This method is a reliable way to test the consistency of solid solutes in mixed dissolution media at various temperatures, including sulfanilamide.


Although qualitatively accurate, the reference solvent solubility values are still not accurate enough and cannot replace actual measurements. However, machine learning protocols may be implemented to further increase the accuracy of predicted solubility. The cured dataset, which encompasses congruent and coherent data, will be used in our next project for this purpose.


Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

ORCID iDs

Piotr Cysewski  <https://orcid.org/0000-0001-6220-2634>

Maciej Przybyłek  <https://orcid.org/0000-0002-3399-6129>

Tomasz Jeliński  <https://orcid.org/0000-0002-2359-6260>

References

1. Ovung A, Bhattacharyya J. Sulfonamide drugs: Structure, antibacterial property, toxicity, and biophysical interactions. *Biophys Rev.* 2021;13(2):259–272. doi:10.1007/s12551-021-00795-9

2. Hegazy WAH, Rajab AAH, Abu Lila AS, Abbas HA. Anti-diabetics and antimicrobials: Harmony of mutual interplay. *World J Diabetes*. 2021; 12(11):1832–1855. doi:10.4239/wjd.v12.i11.1832
3. Drugbank Online. Sulfanilamide. Drugbank Online; 2023. <https://go.drugbank.com/drugs/DB00259>. Accessed November 15, 2023.
4. Odnoroh M, Marty JD, Bourdon V, Coutelier O, Destarac M. Primary sulfonamide-functional polymers with controlled chain architectures by RAFT polymerisation. *Polym Chem*. 2023;14(33):3821–3826. doi:10.1039/D3PY00629H
5. Erol I. Synthesis and characterization of novel sulfonamide functionalized maleimide polymers: Conventional kinetic analysis, antimicrobial activity and dielectric properties. *J Mol Struct*. 2022;1255:132362. doi:10.1016/j.molstruc.2022.132362
6. Reddy CU, Amalraj A, Reddy BSR, Arun A. Polymeric drug based on sulfanilamide: Synthesis, antimicrobial and drug releasing studies. *J Pharm Pharmacol*. 2010;59(9):1207–1213. doi:10.1211/jpp.59.9.0004
7. Eckert F, Klamt A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J*. 2002;48(2):369–385. doi:10.1002/aic.690480220
8. Klamt A, Schüürmann G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J Chem Soc Perkin Trans 2*. 1993;5:799–805. doi:10.1039/P29930000799
9. Popović BM, Uka D, Alioui O, Ždero Pavlović R, Benguerba Y. Experimental and COSMO-RS theoretical exploration of rutin formulations in natural deep eutectic solvents: Solubility, stability, antioxidant activity, and bioaccessibility. *J Mol Liq*. 2022;359:119266. doi:10.1016/j.molliq.2022.119266
10. AlSaleem SS, Zahid WM, AlNashef IM, Hadj-Kali MK. Solubility of halogenated hydrocarbons in hydrophobic ionic liquids: Experimental study and COSMO-RS prediction. *J Chem Eng Data*. 2015;60(10):2926–2936. doi:10.1021/acs.jced.5b00310
11. Kahlen J, Masuch K, Leonhard K. Modelling cellulose solubilities in ionic liquids using COSMO-RS. *Green Chem*. 2010;12(12):2172. doi:10.1039/c0gc00200c
12. Klajmon M. Purely predicting the pharmaceutical solubility: What to expect from PC-SAFT and COSMO-RS? *Mol Pharm*. 2022;19(11):4212–4232. doi:10.1021/acs.molpharmaceut.2c00573
13. Acree W, Chickos JS. Phase transition enthalpy measurements of organic and organometallic compounds: Sublimation, vaporization and fusion enthalpies from 1880 to 2015. Part 1: C₁–C₁₀. *J Phys Chem Ref*. 2016;45(3):033101. doi:10.1063/1.4948363
14. Acree W, Chickos JS. Phase transition enthalpy measurements of organic and organometallic compounds and ionic liquids: Sublimation, vaporization, and fusion enthalpies from 1880 to 2015. Part 2: C₁₁–C₁₉₂. *J Phys Chem Ref*. 2017;46(1):013104. doi:10.1063/1.4970519
15. Cysewski P, Przybyłek M, Jeliński T. Intermolecular interactions as a measure of dapsone solubility in neat solvents and binary solvent mixtures. *Materials (Basel)*. 2023;16(18):6336. doi:10.3390/ma16186336
16. Toscani S, Thorén S, Agafonov V, Céolin R, Dugué J. Thermodynamic study of sulfanilamide polymorphism: (I) Monotropy of the α -variety. *Pharm Res*. 1995;12(10):1453–1456. doi:10.1023/A:1016271017802
17. Trontelj Z, Lužnik J, Pirnat J, Jazbinšek V, Lavrič Z, Srčič S. Polymorphism in sulfanilamide: 14N nuclear quadrupole resonance study. *J Pharm Sci*. 2019;108(9):2865–2870. doi:10.1016/j.xphs.2019.05.015
18. Toscani S, Dzyabchenko A, Agafonov V, Dugué J, Céolin R. Polymorphism of sulfanilamide: (II) Stability hierarchy of α -, β - and γ - forms from energy calculations by the atom–atom potential method and from the construction of the p,T phase diagram. *Pharm Res*. 1996; 13(1):151–154. doi:10.1023/A:1016058123659
19. Jeliński T, Bugalska N, Koszucka K, Przybyłek M, Cysewski P. Solubility of sulfanilamide in binary solvents containing water: Measurements and prediction using Buchowski–Ksiazczak solubility model. *J Mol Liq*. 2020;319:114342. doi:10.1016/j.molliq.2020.114342
20. Abbott S. *Solubility: Principles and Practice*. Ipswich, UK: Steven Abbott TCNF Ltd; 2017. https://www.stevenabbott.co.uk/_downloads/Solubility%20Science%20Principles%20and%20Practice.pdf. Accessed December 1, 2023.
21. Cysewski P. Application of the consonance solvent concept for accurate prediction of Buckminster solubility in 180 net solvents using COSMO-RS approach. *Symmetry*. 2019;11(6):828. doi:10.3390/sym11060828
22. Cysewski P, Jeliński T, Cymerman P, Przybyłek M. Solvent screening for solubility enhancement of theophylline in neat, binary and ternary NADES solvents: New measurements and ensemble machine learning. *Int J Mol Sci*. 2021;22(14):7347. doi:10.3390/ijms22147347
23. Cysewski P, Jeliński T, Przybyłek M. Application of COSMO-RS-DARE as a tool for testing consistency of solubility data: Case of coumarin in neat alcohols. *Molecules*. 2022;27(16):5274. doi:10.3390/molecules27165274
24. Przybyłek M, Miernicka A, Nowak M, Cysewski P. New screening protocol for effective green solvents selection of benzamide, salicylamide and ethenzamide. *Molecules*. 2022;27(10):3323. doi:10.3390/molecules27103323
25. Cysewski P, Przybyłek M, Rozalski R. Experimental and theoretical screening for green solvents improving sulfamethizole solubility. *Materials (Basel)*. 2021;14(20):5915. doi:10.3390/ma14205915
26. Cysewski P, Przybyłek M, Kowalska A, Tymorek N. Thermodynamics and intermolecular interactions of nicotinamide in neat and binary solutions: Experimental measurements and COSMO-RS concentration dependent reactions investigations. *Int J Mol Sci*. 2021; 22(14):7365. doi:10.3390/ijms22147365
27. Przybyłek M, Kowalska A, Tymorek N, Dziarn T, Cysewski P. Thermodynamic characteristics of phenacetin in solid state and saturated solutions in several neat and binary solvents. *Molecules*. 2021; 26(13):4078. doi:10.3390/molecules26134078
28. Delgado DR, Romdhani A, Martínez F. Thermodynamics of sulfanilamide solubility in propylene glycol + water mixtures. *Lat Am J Pharm*. 2011;30(10):2024. http://www.latamjpharm.org/resumenes/30/10/LAJOP_30_10_1_23.pdf. Accessed November 15, 2023.
29. Delgado DR, Peña Fernández MÁ, Martínez F. Preferential solvation of some sulfonamides in 1,4-dioxane + water co-solvent mixtures at 298.15 K according to the inverse Kirkwood–Buff integrals method. *Rev la Acad Colomb ciencias exactas, físicas y Nat*. 2014;38(146): 104–114. <https://raccefyn.co/index.php/raccefyn/article/view/44>. Accessed November 15, 2023.
30. Cárdenas ZJ, Jiménez DM, Almanza OA, Jouyban A, Martínez F, Acree WE. Solubility and preferential solvation of sulfanilamide, sulfamethizole and sulfapyridine in methanol+water mixtures at 298.15 K. *J Solution Chem*. 2016;45(10):1479–1503. doi:10.1007/s10953-016-0513-7
31. Kodide K, Asadi P, Thati J. Solubility and thermodynamic modeling of sulfanilamide in 12 mono solvents and 4 binary solvent mixtures from 278.15 to 318.15 K. *J Chem Eng Data*. 2019;64(12):5196–5209. doi:10.1021/acs.jced.9b00411
32. Asadi P, Kodide K, Kota M, Thati J. Determination and correlation of solubility and solution thermodynamics of 4-aminobenzenesulfonamide in five binary solvent mixtures from 278.15 to 318.15 K. *J Mol Liq*. 2020;303:112670. doi:10.1016/j.molliq.2020.112670
33. Dassault Systèmes. COSMOtherm. Dassault Systèmes, Biovia, San Diego, USA; 2022. <https://www.3ds.com/products/biovia/cosmo-rs/cosmotherm>. Accessed November 15, 2023.
34. Dassault Systèmes. COSMOconf. Dassault Systèmes, Biovia, San Diego, USA; 2022. <https://www.3ds.com/products/biovia>. Accessed November 15, 2023.
35. Martínez F, Gómez A. Thermodynamic study of the solubility of some sulfonamides in octanol, water, and the mutually saturated solvents. *J Solution Chem*. 2001;30(10):909–923. doi:10.1023/A:1012723731104
36. Martínez F, Gomez A. Estimation of the solubility of sulfonamides in aqueous media from partition coefficients and entropies of fusion. *Phys Chem Liq*. 2002;40(4):411–420. doi:10.1080/0031910021000017735
37. Ciocazanu I, Meltzer V. Etude thermodynamique de certains médicaments. *J Ther Anal*. 1996;47(6):1755–1758. doi:10.1007/BF01980921
38. Ohm A, Lippold BC. Surface tension of solid drugs and its correlation with solubility. *Int J Pharm Technol Prod Manuf*. 1985;6(4):1–6.
39. Cysewski P. Prediction of ethenzamide solubility in organic solvents by explicit inclusions of intermolecular interactions within the framework of COSMO-RS-DARE. *J Mol Liq*. 2019;290:111163. doi:10.1016/j.molliq.2019.111163
40. Asadi P, Kodide K, Thati J, Busi MR. Predicting sulfanilamide solubility in mixed solvents: A comparative analysis of computational models. *Fluid Phase Equilibria*. 2024;577:113966. doi:10.1016/j.fluid.2023.113966