DOI:
10.1016/j.ijpharm.2017.07.048

Document Version
Peer reviewed version

Title: Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin

Authors: Safa A. Damiati, Luigi G. Martini, Norman W. Smith, Jayne M. Lawrence, David J. Barlow

PII: S0378-5173(17)30647-6
DOI: http://dx.doi.org/doi:10.1016/j.ijpharm.2017.07.048
Reference: IJP 16863

To appear in: *International Journal of Pharmaceutics*

Received date: 22-5-2017
Revised date: 28-6-2017
Accepted date: 15-7-2017

Please cite this article as: Damiati, Safa A., Martini, Luigi G., Smith, Norman W., Lawrence, Jayne M., Barlow, David J., Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin.International Journal of Pharmaceutics http://dx.doi.org/10.1016/j.ijpharm.2017.07.048

Application of machine learning in prediction of hydrotrope-enhanced solubilisation of indomethacin

Safa A. Damiati[1], Luigi G. Martini [2], Norman W. Smith [3], Jayne M. Lawrence[*,1], and David J. Barlow[*,1]

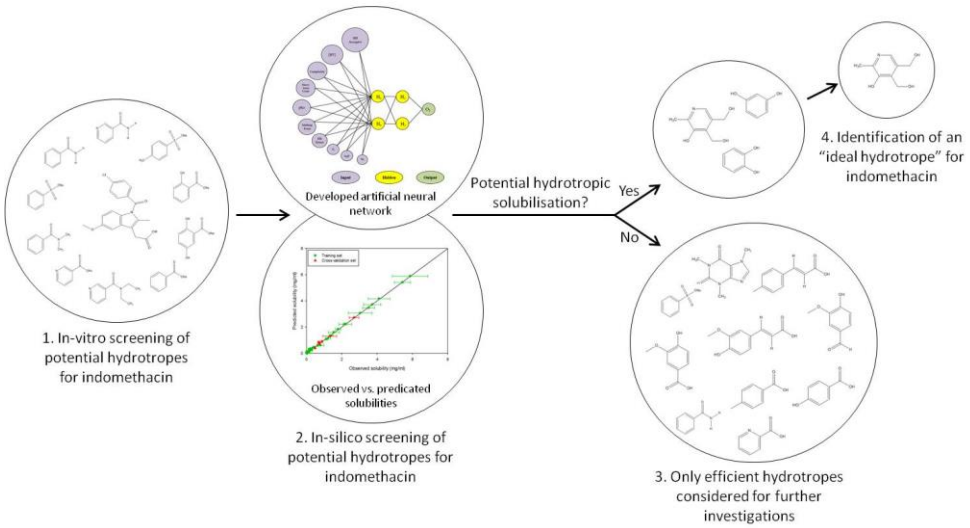
[1] Institute of Pharmaceutical Science, Faculty of Life Sciences & Medicine, Franklin-Wilkins Building, King's College London, 150 Stamford Street, London, SE1 9NH, UK

[2] Current address: Roche Products Ltd. (Pharmaceuticals), Hexagon Place, 6 Falcon Way, Shire Park, Welwyn Garden City, Hertfordshire, AL7 1TW, UK

[3] Analytical and Environmental Science Division, Faculty of Life Sciences & Medicine, King's College London, 150 Stamford Street, London SE1 9NH, UK.

[*] Corresponding authors

Graphical abstract



Abstract

Systematic *in-vitro* studies have been conducted to determine the ability of a range of 10 potential hydrotropes to improve the apparent aqueous solubility of the poorly water soluble drug, indomethacin. Solubilisation of the drug in the presence of the hydrotropes was determined experimentally using high-performance liquid chromatography (HPLC) with ultraviolet (UV) detection. These experimental data, together with various known and computed physicochemical properties of the hydrotropes were thereafter used *in silico* to train an artificial neural network (ANN) to allow for predictions of indomethacin solubilisation. The trained ANN was found to give highly accurate predictions of indomethacin solubilisation in the presence of hydrotropes and was thus shown to provide a valuable means by which hydrotrope efficacy could be screened computationally. Interrogation of the network connection weights afforded a quantitative assessment of the relative importance of the various hydrotrope physicochemical properties in determining the extent of the enhancement in indomethacin solubilisation. It is concluded that *in-silico* screening of drug/hydrotrope systems using artificial neural networks offers significant potential to reduce the need for extensive laboratory testing of these systems, and could thus provide an economy in terms of reduced costs and time in drug formulation development.

[Word count: 189; max = 200]

Keywords: artificial neural networks; indomethacin; drug formulation; hydrotropes; drug solubility

Chemical compounds studied in this article:

Indomethacin (PubChem CID: 3715); sodium nicotinate (PubChem CID: 23687810); sodium benzoate (PubChem CID: 517055); sodium salicylate (PubChem CID: 16760658); sodium gentisate (PubChem CID: 23663629); sodium benzenesulfonate(PubChem CID: 517327); sodium *p*-toluenesulfonate (PubChem CID: 3720192); benzamide (PubChem CID: 2331); *N,N*-dimethylbenzamide (PubChem CID: 11916); nicotinamide (PubChem CID: 936); *N,N*-diethylnicotinamide (PubChem CID: 5497).

1. Introduction

The task of optimising drug solubility remains a major concern for the formulator of pharmaceutical dosage forms. Approximately 40% of drugs with market approval and almost

90% of compounds in the development pipeline fail to reach the market because of problems of poor aqueous solubility (Kalepu and Nekkanti 2015, Zhang et al. 2014). It is well known that poor aqueous solubility can significantly limit a drug's bioavailability and ultimately, therefore, its efficacy, thereby necessitating the administration of higher doses and increasing the chance of side effects as well as costing more per patient (Williams et al. 2013). There are numerous formulation strategies that have been proposed and used to overcome this problem, and one such strategy involves the use of hydrotropes (Kunz et al. 2016).

The term of hydrotropy was used to describe the increase of the apparent aqueous solubility of insoluble or slightly soluble organic substances in water by the addition of large concentrations of some organic acid salts of alkali metal such as sodium benzoate (Neuberg 1916). Hydrotropes are modestly amphiphilic organic compounds that possess only a relatively small hydrophobic moiety which were considered as non-micelle forming compounds (Elworthy et al. 1968). Unlike surfactants, however, – which can also be used to enhance drug solubilisation – hydrotropes achieve their solubilisation effects only at quite high concentration, generally, in the molar rather than the millimolar range (Balasubramanian et al. 1989, Hopkins Hatzopoulos et al. 2011).

The phenomenon of hydrotropy was first recognised by Carl Neuberg back in 1916 (Neuberg 1916) but there has been relatively little work since that time performed to explore – in a detailed and systematic manner – the ability of different hydrotropes to increase the apparent (drug) solubility. A variety of mechanisms have been proposed for hydrotropic solubilisation including: increased drug solubilisation as the result of hydrotrope self-aggregate formation (Badwan et al. 1983, Balasubramanian et al. 1989, Cui et al. 2010), hydrotrope-induced alteration of the water structure (Coffman and Kildsig 1996), formation of solute-hydrotrope complexes (Sanghvi et al. 2007), and more recently, solubilisation driven by an accumulation of hydrotrope around the drug (Abbott et al. 2017, Booth et al. 2015, Shimizu and Matubayasi 2014).

As a consequence of our incomplete understanding of the phenomenon, the task of selecting an appropriate hydrotrope for a given poorly water-soluble drug is generally approached simply by trial-and-error screening of a large number of potential hydrotropic agents. Such selection, however, might be much more conveniently made *in silico* using artificial neural networks (ANNs). ANNs are biologically inspired computational models capable of simulating the brain's ability to learn by example; they provide particularly powerful tools to aid in the modeling of non-linear relationships and have found numerous applications in the pharmaceutical sciences (Gawehn et al. 2016, Marsland 2015, Sutariya et al. 2013).

In the studies reported here our aim was to explore the potential for using ANNs to predict the effects of likely hydrotropic agents on the apparent aqueous solubility of poorly water-soluble drugs. In order for such *in-silico* predictions to provide significant benefit to drug formulators, the ANN ultimately produced would need to cater for *any* poorly water-soluble drug, and be able to provide reliable predictions of the solubility enhancements achieved

using *any* potential hydrotrope. In the initial studies reported here, however, we aimed only to establish proof of principle, seeking only to determine whether an ANN could be trained to give reliable predictions of the solubility enhancements achieved using different hydrotropic agents for just a *single* drug. The drug chosen for the studies was the non-steroidal, anti-inflammatory agent, indomethacin (Figure 1). This particular drug was chosen in part because of the interest in producing aqueous formulations for injection and ophthalmic use (Elsaman and Ali 2016, Halim Mohamed and Mahmoud 2011, Sostres et al. 2013), and partly because there have been some experimental studies reported in which hydrotropes have been explored as a means to improve its aqueous solubility (Jain 2008).

Experimental data were obtained on the solubilisation of the drug using a systematically chosen range of 10 potential hydrotropes, each possessing a simple (phenyl or pyridinyl) aromatic hydrophobe and a hydrophilic group comprising a carboxylate, sulfonate, or amide moiety. The chosen compounds included, sodium nicotinate, sodium benzoate, sodium salicylate, sodium gentisate, sodium benzenesulfonate, sodium *p*-toluenesulfonate, benzamide, *N,N*-dimethylbenzamide, nicotinamide, and *N,N*-diethylnicotinamide (Figure 2). The drug solubilisation data obtained were taken along with various known and computed physicochemical descriptors of the hydrotropes and were used to train an ANN to allow for prediction of new hydrotropes for indomethacin.

2. Materials and methods

2.1. Chemicals and Reagents

Indomethacin (purity ≥99%, confirmed as the λ polymorph, and used as received), sodium nicotinate, sodium benzoate, sodium salicylate, sodium gentisate, sodium benzenesulfonate, sodium *p*-toluenesulfonate, benzamide, *N,N*-dimethylbenzamide, nicotinamide, *N,N*-diethylnicotinamide, pyridoxine, and formic acid were all purchased from Sigma-Aldrich Ltd. (Dorset, U.K.). Caffeine was supplied from Alfa Aesar (Lancashire, U.K.). Both water and acetonitrile were HPLC-grade and were obtained from Fisher Scientific UK Ltd. (Leicestershire, U.K.). All chemical reagents were ≥95% pure and used as received. Water at 18.2 mΩ.cm (25 ^{o}C) was purified by an ultra-pure water system (ELGA, UK).

2.2. Instrumentation and chromatographic conditions

Chromatographic analysis of indomethacin was performed using an HP 1050 series liquid chromatography system (Agilent Technologies, UK), equipped with HP 1050 quaternary pump, DAD detector, autosampler, and column oven, and a G1379A vacuum degasser. The system was controlled by ChemStation Software (version A10.02). Chromatography was performed using an XTerra 3.5 µm C18 column, 2.1 mm ID X 150 mm length (Waters, Milford, USA). The mobile phase was 45:55% (v/v) acetonitrile:water with 0.1% formic acid at a flow rate of 0.3 mL/min. The injection volume was 5 µL and the wavelength set at 320 nm. All measurements were made in triplicate and averaged for all experiments. The total run time

for each analysis was 10 min. Under the conditions described, the measured retention time for indomethacin was 7 min and not influenced by the presence of hydrotrope.

2.3. Working standard solution and sample preparation

A standard solution of indomethacin dissolved in the mobile phase was prepared at a concentration of 100 µg/mL. The standard solution was further diluted in the mobile phase to obtain 2.5, 5, 7.5, 10, and 12.5 µg/mL solutions which were used for calibration. Each dilution was prepared in triplicate and Beer's law was obeyed in this concentration range (r^2=0.999).

Different concentrations of the 10 hydrotropes were prepared in ultrapure water (with uncertainties in the measured concentrations $\leq$ 1%). Excess indomethacin was added to microfuge tubes containing different concentrations of each hydrotrope (n = 9, i.e., 3 concentrations tested on 3 different occasions). To reach equilibrium, the aqueous solutions were prepared at ambient temperature (22 +/- 2 ^{o}C) and mixed by affixing the tubes to a circular rotating wheel with rotation carried out for 24, 48 and 72 h. Tubes were covered with aluminium foil to avoid any possibility of photochemical reaction. At each time interval, 3 tubes/concentration were removed and centrifuged for 15 minutes at 13000 rpm (using an Heraeus Biofuge Pico D-35720; Germany) and the supernatant then transferred into fresh tubes and re-centrifuged under the same conditions, to ensure removal of the excess indomethacin. Note, that preliminary studies indicated that no difference in the apparent aqueous solubility of indomethacin was obtained if the excess drug was removed by filtration through 0.22 µm polyethersulfone filters (Jet Biofiltration Co. Ltd, China). Furthermore, the form of the excess indomethacin in the solubility experiment was confirmed to be the gamma form by means of melting point determination. The pH of the indomethacin, hydrotrope, and indomethacin-hydrotrope solutions was measured in triplicate at ambient temperature (22 +/- 2 ^{o}C) by means of pH meter (Jenway-3505, UK).

The prepared mixtures of indomethacin/hydrotrope were diluted with the mobile phase to a concentration falling within the calibration range and the dissolved indomethacin quantified by HPLC-UV as a function of each hydrotrope concentration. In addition, as equilibrium was achieved by 24 h, all results after 24, 48 and 72 h were averaged and the values reported below represent the mean ± standard deviation (n ≥ 8).

2.4. Artificial neural network structure, training, and usage

Implementation and training of a feed-forward back-propagation artificial neural network (ANN) (Rumerhart et al. 1986) were carried out using the Windows© version (Nazir et al. 2002) of the PUDDLE software (Richardson and Barlow 1996). A network consisting of an input, two hidden, and an output layer, with 10, 2, 2, 1 neurons, respectively, was found adequate for predicting the aqueous solubility of indomethacin as a function of various hydrotrope properties and varying hydrotrope concentrations. The network topology and

learning parameters (i.e., network momentum, noise level, weight and bias range) were varied to optimise training performance and prediction accuracy. ANN training was carried out using batch-processing of the input data, and employing a sigmoid activation function, with the time-invariant noise algorithm (TINA) (Robert et al. 1992) to avoid trapping in local minima.

The parameters chosen as inputs to the ANN were selected so that all were readily interpretable in structural terms, and all determined by simple inspection of the chemical structures of the hydrotropes or else reliably predicted using readily available computational tools and/or web resources. Pilot studies were carried out using a set of 12 physicochemical descriptors, and two of these (molecular weight and aqueous solubility) were found to make little contribution in the ANN training. The descriptors that were ultimately used in ANN training included: hydrotrope concentration, melting point, logP, pKa, hydrogen bond donor count, hydrogen bond acceptor count, heavy atom count, complexity, fraction ionised (pH 6.6), and acid/base species (-1, 0, and +1 for acidic, neutral and basic compounds, respectively). The logP values used for the acidic hydrotropes that were studied experimentally as sodium salts were taken as those of the parent acids.

Values of the melting point, logP, and pKa were determined utilizing EPI suite™ software (EPIsuite™ 2008) unless otherwise stated. The remaining properties of hydrogen bond donor count, hydrogen bond acceptor count, and complexity were obtained from the NCBI PubChem Database (Kim et al. 2015). All input parameters were normalised to the range 0-1, and the ANN training and cross-validation performed using 35 and 8 datasets, respectively (Table 1a-b).

The partitioning of the input between the training and cross-validation datasets was made in such a way that the examples used in training defined the limits of the parameter space to be explored, with the cross-validation input patterns then chosen so that they sampled within and across this space, involving a selection of hydrotropes from the same series used in training (and with varying hydrotrope concentrations), and with the addition of caffeine – a widely-used hydrotrope (Cui 2010, Lee et al. 2003) – serving as an external test compound by which to judge the ability of the trained ANN to generalize.

3. Results and discussion

3.1. In-vitro screening of potential hydrotropes for indomethacin

The baseline aqueous solubility of indomethacin was determined as 2.8 ± 0.4 μg/mL (at pH 6.6; n = 18), and its apparent aqueous solubility in the presence of the various hydrotropes found as summarised in Table 1a-b (the coefficient of variation on the measured drug concentrations in all cases $\leq$ 0.5%). As has been noted previously (Jain 2008, Kim et al. 2010), the magnitude of the drug's solubility enhancement is seen to increase with increasing hydrotrope concentration but the pattern of this increase varies from one hydrotrope to

another (see Figure 3). The range of the drug's solubility enhancements obtained using 0.5 M hydrotrope vary from a modest 3-fold increase obtained with the addition of sodium *p*-toluenesulfonate, through to a 470-fold increase obtained in the presence of 0.5 M sodium nicotinate.

For some hydrotropes, the solubility enhancement that could be achieved was limited by the low aqueous solubility of the hydrotrope. Benzamide, for example, could only be tested at 0.1 M, while sodium benzenesulfonate and sodium gentisate could be tested only up to 0.9 M.

A subjective assessment of the solubility enhancement data shown in Table 1a-b, coupled with a simple inspection of the structures of the experimentally screened hydrotropes (Figure 2) served to confirm some of the observations recorded by earlier researchers, but also revealed some anomalous behaviour.

Kim *et al*. (Kim et al. 2010) have observed that the hydrogen bonding capacity of a hydrotrope is key to its efficacy in solubilising drugs, and Lee *et al.(Lee et al. 2003)* have observed too that the polarity of a hydrotrope is an important determinant of solubility enhancement, with compounds that have a greater proportion of polar groups affording a lesser enhancement of drug solubility. Concordantly, we see here that the solubility of indomethacin is increased roughly 260-fold increase in the presence of sodium benzoate, but only increased by around 60-fold in the presence of the more polar sodium salicylate – these two compounds differing in polarity as the result of an additional hydroxyl group in sodium salicylate (see Figure 2). Contrary to this view, however, we find that sodium gentisate – which is more polar than sodium salicylate (by virtue of a second hydroxyl group on the aromatic ring; Figure 2) exhibits a 90-fold increase in indomethacin solubility, which is therefore *greater* than, rather than *less* than the 60-fold increase seen for sodium salicylate.

Lee *et al.* (Lee et al. 2003) also observed that the nature of a hydrotrope's ring system can influence its drug solubility enhancement, and this too is noted here. We see, for example, that sodium nicotinate, which has a pyridine ring, shows a 470-fold increase in indomethacin solubility at a concentration of 0.5 M, whereas sodium benzoate, which has a phenyl ring, shows only a 260-fold at the same concentration. It would thus appear that, for indomethacin, at least, an aromatic heterocycle serves greater benefit than a carbocycle.

From the data presented here we note too that – consistent with the data presented in Lee *et al.* (Lee et al. 2003) – sulfonates seem to be generally less effective as hydrotropes than carboxylates, with the solubility enhancement achieved with sodium benzenesulfonate much less than that achieved using the same concentration of sodium benzoate. In addition, we find that an increased fraction of sp^3 carbon in the hydrophobe is beneficial (as *per* the findings reported by Kim *et al.* (Kim et al. 2010) and Lee *et al*. (Lee et al. 2003), with the solubility enhancement achieved with 0.1 M *N,N*-diethylnicotinamide greater than that achieved using the same concentration of the *N,N*-dimethyl compound, and this in turn greater than that achieved using the parent benzamide.

Given that the aqueous solubility of indomethacin varies with pH (Jain, 2008), and the fact that many of the hydrotropes studied here will change the solvent pH, one might suppose that the measured solubility enhancements of the drug would follow the pattern of pH change. A plot of the measured equilibrium drug solubility against hydrotrope solution pH, however, shows no clear pattern (see Supplementary Information, Figure S1). Whilst the solvent pH, therefore, will undoubtedly affect the drug's aqueous solubility, as too will the use of counterions, the net enhancement to solubility is patently the result of several contributing factors.

Taking the experimentally determined solubility enhancement data for each hydrotrope and interpolating the fitted curve will of course furnish predicted solubility enhancements for concentrations of hydrotrope that have not been tested but such modelling has limited utility given that it is confined to prediction for a single hydrotrope, and affords no insight into the properties of hydrotropes that influence drug solubilisation. An ANN trained to predict the solubility enhancements using hydrotrope chemical properties as inputs, however, provides the perfect means to determine the nature of the key determinants, and also allows a quantitative assessment of their relative importance.

3.2 In-silico screening of potential hydrotropes for indomethacin

Experimental data from the drug solubility studies described above were used as the target output to train an artificial neural network (ANN), with input data furnished by the corresponding normalised hydrotrope concentrations, together with the physicochemical descriptors for the hydrotropes as given in Table 1a-b. The trained ANN was subsequently interrogated to explore the relative importance of the input parameters in determining drug solubility enhancement, and then also used for *in-silico* screening of untested hydrotrope systems in a search for compounds that might afford a greater enhancement of indomethacin solubility.

The ANN developed was found to give highly accurate predictions of indomethacin solubility in the presence of hydrotropes at different concentrations, with residuals randomly scattered, and typically lying in the range ± 0.1 $mg.mL^{-1}$ (see Supplementary Information, Figure S2). The correlation of the observed and predicted drug solubility data (Figure 4) was high (r^2 = 0.998), with the slope of the regression line insignificantly different from unity and the intercept insignificantly different from zero (p < 0.001 in both cases).

A quantitative assessment of the relative importance of the various hydrotrope physicochemical properties in determining the extent of the solubility enhancement for indomethacin was afforded through an interrogation of the sum of squared connection weights for the outgoing connections from the input neurons (Richardson et al. 1997). By this means, it was found that the order of importance of the input parameters was: hydrogen bond acceptor count > hydrotrope concentration > melting point > complexity > pKa > heavy

atom count > hydrogen bond donor count > F (ionised) > log P > species (acidic, basic or neutral).

As a means to explore the landscape of hydrotrope properties more fully – and by such means then to determine the set of features required in the "ideal" hydrotrope for indomethacin – the trained ANN was subsequently used in *in-silico* screening of a set of 16 new compounds which were manually selected as potential hydrotropes that might afford a greater enhancement of indomethacin solubility (see Table 2 and Figure 5). The same set of 10 physicochemical descriptors as used in the ANN training was employed here for prediction of the solubility enhancement of these 16 compounds.

The results of the solubility enhancement of indomethacin varied quite widely among these *in-silico* screened hydrotropes, with 0.5 M *p*-toluic acid and 0.5 M *p*-coumaric acid predicted to give only 93-fold and 106–fold enhancements, and 0.5 M picolinic acid predicted to give an impressive 800-fold enhancement. Caffeine – the external test compound included to test the ANN's generalisation – was predicted to give an extremely poor solubility enhancement for indomethacin (0.02 mg/mL at 0.1 M) and the level of enhancement was subsequently determined by experiment to be gratifyingly close to this value, at 0.032 ± 0.001 µg/mL.

Of the 16 compounds included in this initial test screening, it is pertinent to note that resorcinol – a compound that has been extensively studied and shown to be a very good hydrotropic agent (Abdelbary et al. 2016, Lee et al. 2003, Roy and Moulik 2002) – was predicted here to give a solubility enhancement for indomethacin of the order of 2000-fold.

On the basis of the ANN connection weight interrogation and the *in-silico* screening of the 16 test hydrotropes, the "ideal" hydrotrope for indomethacin was identified as a low complexity compound having a pyridine ring as hydrophobe, with an alkyl-substituted amide moiety, and a low hydrogen bond acceptor count.

With these criteria, a search was then made for compounds that were commercially available, and which could be tested empirically as potential new hydrotropes for indomethacin. A suitable candidate was found in pyridoxine (vitamin B6; Figure 6) – for which our trained ANN predicted 7.78 mg/mL indomethacin dissolved using a hydrotrope concentration of 0.5 M. Given that the reported aqueous solubility of this compound is just 220 mg/mL (Kim et al. 2015), its ability to enhance the apparent aqueous solubility of indomethacin was determined using concentrations of 0.1 M and 0.5 M only. The apparent aqueous solubility of the indomethacin with these hydrotrope levels was determined as 0.39 ± 0.04 mg/mL, and 2.18 ± 0.27 mg/mL (n = 3), respectively. Fully in accord with our predictions, therefore, we find that pyridoxine does indeed provide a very good hydrotrope for indomethacin, giving a 130-fold solubility enhancement at 0.1 M – which is comparable with the previous best of a 140-fold enhancement achieved using the same concentration of sodium nicotinate - and an impressive 727-fold enhancement at 0.5 M – which exceeds the previous best of a 470-fold enhancement achieved using 0.5 M sodium nicotinate.

4. Conclusions

The *in-vitro* and *in-silico* studies presented here demonstrate that hydrotropy offers promise as a device to increase the apparent aqueous solubility of poorly water-soluble drugs like indomethacin. The ANN trained on the basis of the reported indomethacin/hydrotrope solubilisation data was utilised not only to predict – with high accuracy – the increase in the apparent aqueous solubility of the indomethacin, but also allowed an informative exploration of the relative importance of various physicochemical properties of the hydrotropes in providing such enhancements. The designed ANN was also shown to be valuable in identifying new prospective hydrotropes for indomethacin.

In our ongoing research, we aim to extend these studies by taking the hydrotrope-enhanced aqueous solubilities for a range of different drugs, and then using these data to develop a more general ANN that can be employed to predict solubility data for a range of drug/hydrotrope systems. Generating the experimental data needed to train such an ANN will of course require significant time and effort, but once the ANN has been trained successfully, it will then be possible to reduce the future need for extensive trial-and-error laboratory testing of these systems, and would thereby reduce the costs and time involved in drug formulation development.

Acknowledgments:

This study was funded by the Saudi Arabian government in the form of a scholarship for Safa A. Damiati.

References:

Abbott, S., Booth, J. J. and Shimizu, S. (2017) Practical molecular thermodynamics for greener solution chemistry. *Green Chemistry, 19*, pp. 68-75.

Abdelbary, G. A., Amin, M. M. and Abdelmoteleb, M. (2016) Novel mixed hydrotropic solubilization of Zaleplon: Formulation of oral tablets and in-vivo neuropharmacological characterization by monitoring plasma GABA level. *Journal of Drug Delivery Science and Technology, 33*, pp. 98-113.

Badwan, A., El-Khordagui, L., Saleh, A. and Khalil, S. (1983) The solubility of benzodiazepines in sodium salicylate solution and a proposed mechanism for hydrotropic solubilization. *International Journal of Pharmaceutics, 13*, pp. 67-74.

Balasubramanian, D., Srinivas, V., Gaikar, V. and Sharma, M. (1989) Aggregation behavior of hydrotropic compounds in aqueous solution. *The Journal of Physical Chemistry, 93*, pp. 3865-3870.

Booth, J. J., Omar, M., Abbott, S. and Shimizu, S. (2015) Hydrotrope accumulation around the drug: the driving force for solubilization and minimum hydrotrope concentration for nicotinamide and urea. *Physical Chemistry Chemical Physics, 17*, pp. 8028-8037.

Coffman, R. E. and Kildsig, D. O. (1996) Effect of nicotinamide and urea on the solubility of riboflavin in various solvents. *Journal of Pharmaceutical Sciences, 85*, pp. 951-954.

Cui, Y. (2010) Parallel stacking of caffeine with riboflavin in aqueous solutions: the potential mechanism for hydrotropic solubilization of riboflavin. *International Journal of Pharmaceutics, 397*, pp. 36-43.

Cui, Y., Xing, C. and Ran, Y. (2010) Molecular dynamics simulations of hydrotropic solubilization and self-aggregation of nicotinamide. *Journal of Pharmaceutical Sciences, 99*, pp. 3048-3059.

Elsaman, T. and Ali, M. M. (2016) Nonsteroidal Anti-Inflammatory Drugs (NSAIDs) Derivatives with Anti-Cancer Activity. *American Journal of Research Communication, 4*.

Elworthy, P. H., Florence, A. T. and Macfarlane, C. B. (1968) *Solubilization by surface-active agents and its applications in chemistry and the biological sciences,* Chapman and Hall: London.

EPIsuite™ (2008) Estimation Programs Interface SuiteTM for Microsoft® Windows. in,Washington DC, USA: United States Environmental Protection Agency.

Gawehn, E., Hiss, J. A. and Schneider, G. (2016) Deep learning in drug discovery. *Molecular Informatics, 35*, pp. 3-14.

Halim Mohamed, M. A. and Mahmoud, A. A. (2011) Formulation of indomethacin eye drops via complexation with cyclodextrins. *Current Eye Research, 36*, pp. 208-216.

Hopkins Hatzopoulos, M., Eastoe, J., Dowding, P. J., Rogers, S. E., Heenan, R. and Dyer, R. (2011) Are hydrotropes distinct from surfactants? *Langmuir, 27*, pp. 12346-12353.

Jain, A. K. (2008) Solubilization of indomethacin using hydrotropes for aqueous injection. *European Journal of Pharmaceutics and Biopharmaceutics, 68*, pp. 701-714.

Kalepu, S. and Nekkanti, V. (2015) Insoluble drug delivery strategies: review of recent advances and business prospects. *Acta Pharmaceutica Sinica B, 5*, pp. 442-453.

Kim, J. Y., Kim, S., Papp, M., Park, K. and Pinal, R. (2010) Hydrotropic solubilization of poorly water-soluble drugs. *Journal of Pharmaceutical Sciences, 99*, pp. 3953-3965.

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S. and Shoemaker, B. A. (2015) *PubChem substance and compound databases,* [online], available: https://pubchem.ncbi.nlm.nih.gov [accessed 06 February 2017].

Lee, J., Lee, S. C., Acharya, G., Chang, C.-j. and Park, K. (2003) Hydrotropic solubilization of paclitaxel: analysis of chemical structures for hydrotropic property. *Pharmaceutical Research, 20*, pp. 1022-1030.

Marsland, S. (2015) *Machine learning: an algorithmic perspective,* 2nd ed*.,* CRC press.

Nazir, J., Barlow, D. J., Lawrence, M. J., Richardson, C. J. and Shrubb, I. (2002) Artificial neural network prediction of aerosol deposition in human lungs. *Pharmaceutical Research, 19*, pp. 1130-1136.

Neuberg, C. (1916) Hydrotropy. *Biochem. Z, 76*, pp. 107-176.

Richardson, C. and Barlow, D. (1996) Neural network computer simulation of medical aerosols. *Journal of Pharmacy and Pharmacology, 48*, pp. 581-591.

Richardson, C., Mbanefo, A., Aboofazeli, R., Lawrence, M. and Barlow, D. (1997) Prediction of phase behavior in microemulsion systems using artificial neural networks. *Journal of Colloid and Interface Science, 187*, pp. 296-303.

Robert, J., Burton, M. and Mpitsos, G. (1992) Event-dependent control of noise enhances learning in neural networks. *Neural Networks, 5*, pp. 627-637.

Roy, B. and Moulik, S. (2002) Functions of hydrotropes (sodium salicylate, proline, pyrogallol, resorcinol and urea) in solution with special reference to amphiphile behaviors. *Colloids and Surfaces A: Physicochemical and Engineering Aspects, 203*(1), pp. 155-166.

Rumerhart, D., Hinton, G. and Williams, R. (1986) Learning representations by back-propagation errors. *Nature, 323*, pp. 533-536.

Sanghvi, R., Evans, D. and Yalkowsky, S. H. (2007) Stacking complexation by nicotinamide: A useful way of enhancing drug solubility. *International Journal of Pharmaceutics, 336*, pp. 35-41.

Shimizu, S. and Matubayasi, N. (2014) Hydrotropy: Monomer–Micelle Equilibrium and Minimum Hydrotrope Concentration. *The Journal of Physical Chemistry B, 118*, pp. 10515-10524.

Sostres, C., Gargallo, C. J. and Lanas, A. (2013) Nonsteroidal anti-inflammatory drugs and upper and lower gastrointestinal mucosal damage. *Arthritis Research & Therapy, 15*, pp. 1-8.

Sutariya, V., Groshev, A., Sadana, P., Bhatia, D. and Pathak, Y. (2013) Artificial neural network in drug delivery and pharmaceutical research. *The Open Bioinformatics Journal, 7*, pp. 49-62.

Williams, H. D., Trevaskis, N. L., Charman, S. A., Shanker, R. M., Charman, W. N., Pouton, C. W. and Porter, C. J. (2013) Strategies to address low drug solubility in discovery and development. *Pharmacological Reviews, 65*, pp. 315-499.

Zhang, P., Forsgren, J. and Strømme, M. (2014) Stabilisation of amorphous ibuprofen in Upsalite, a mesoporous magnesium carbonate, as an approach to increasing the aqueous solubility of poorly soluble drugs. *International journal of Pharmaceutics, 472*, pp. 185-191.

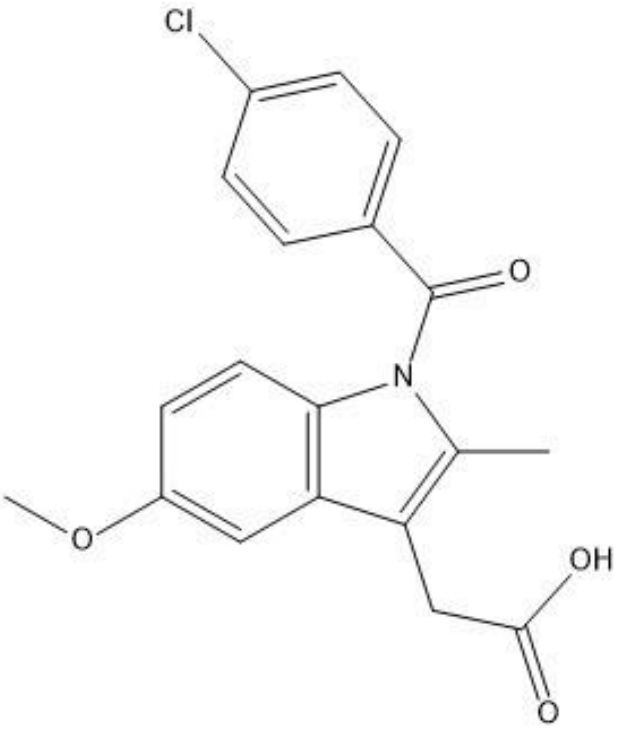


Figure 1: Chemical structure of indomethacin

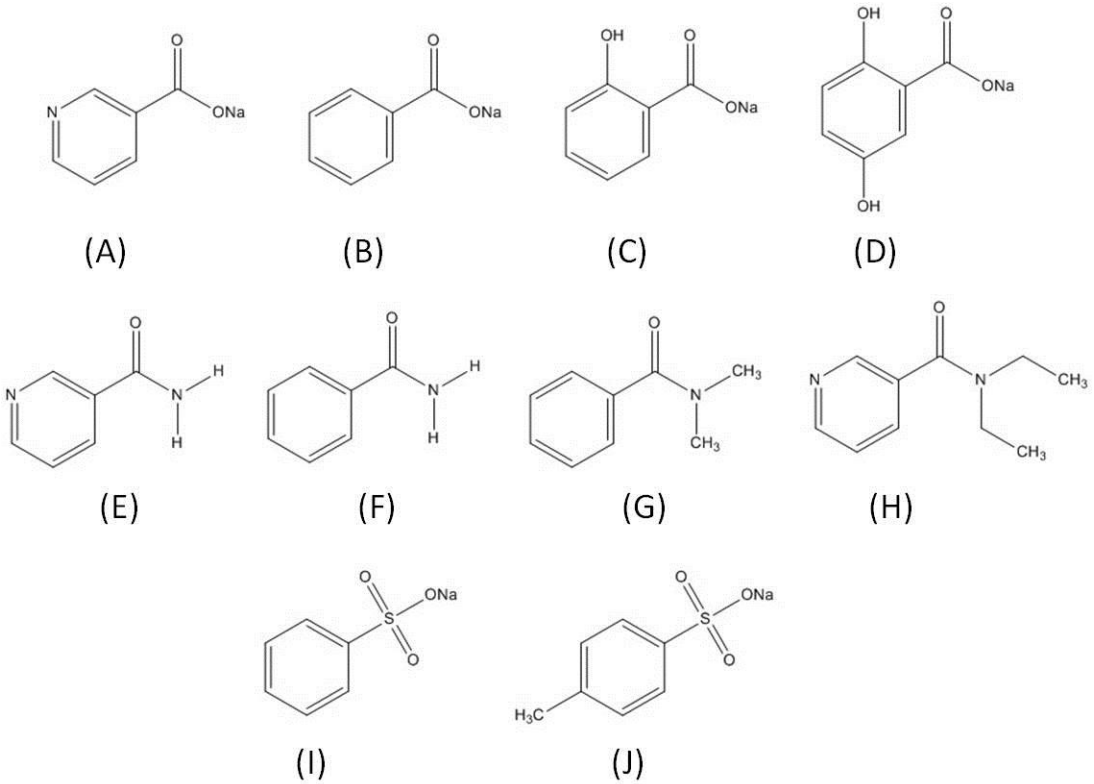


Figure 2: Hydrotropes tested experimentally: sodium nicotinate (A), sodium benzoate (B), sodium salicylate (C), sodium gentisate (D), nicotinamide(E), benzamide (F), *N,N*-dimethylbenzamide (G) *N,N*-diethylnicotinamide (H), sodium benzenesulfonate (I) sodium *p*-toluenesulfonate (J)

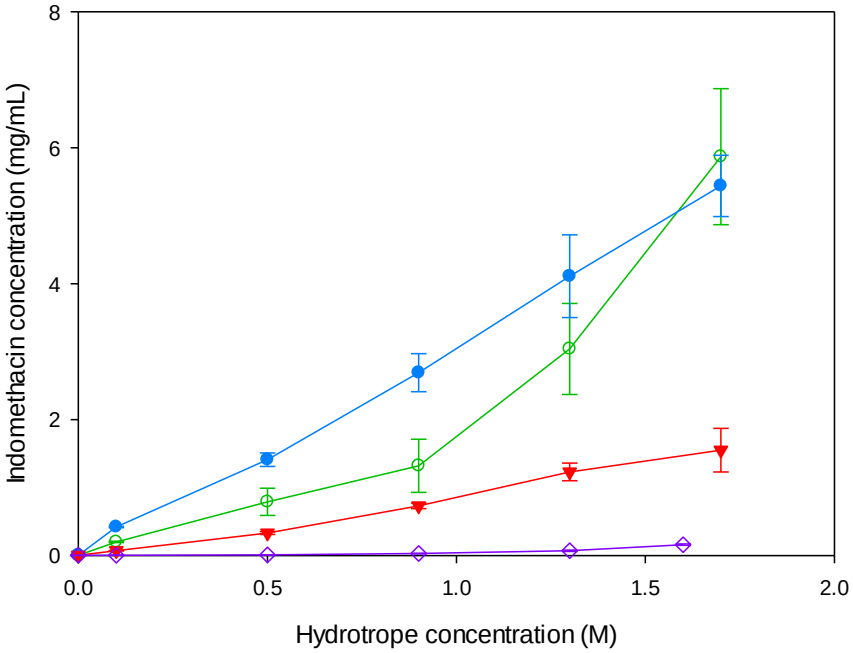


Figure 3: Apparent aqueous solubility of indomethacin as a function of hydrotrope concentration (the data for sodium nicotinate, sodium benzoate, nicotinamide and sodium *p*-toluenesulfonate are plotted as filled circles, open circles, filled triangles and open diamonds, respectively)

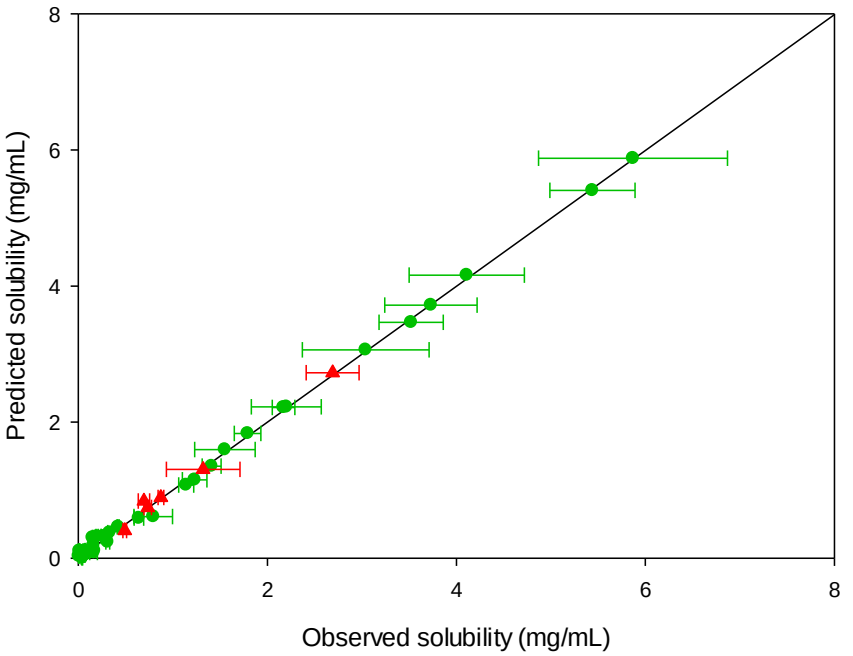


Figure 4: Correlation between the observed and predicted hydrotrope enhanced apparent aqueous solubility of indomethacin (training set data plotted as circles and the cross-validation data as triangles)

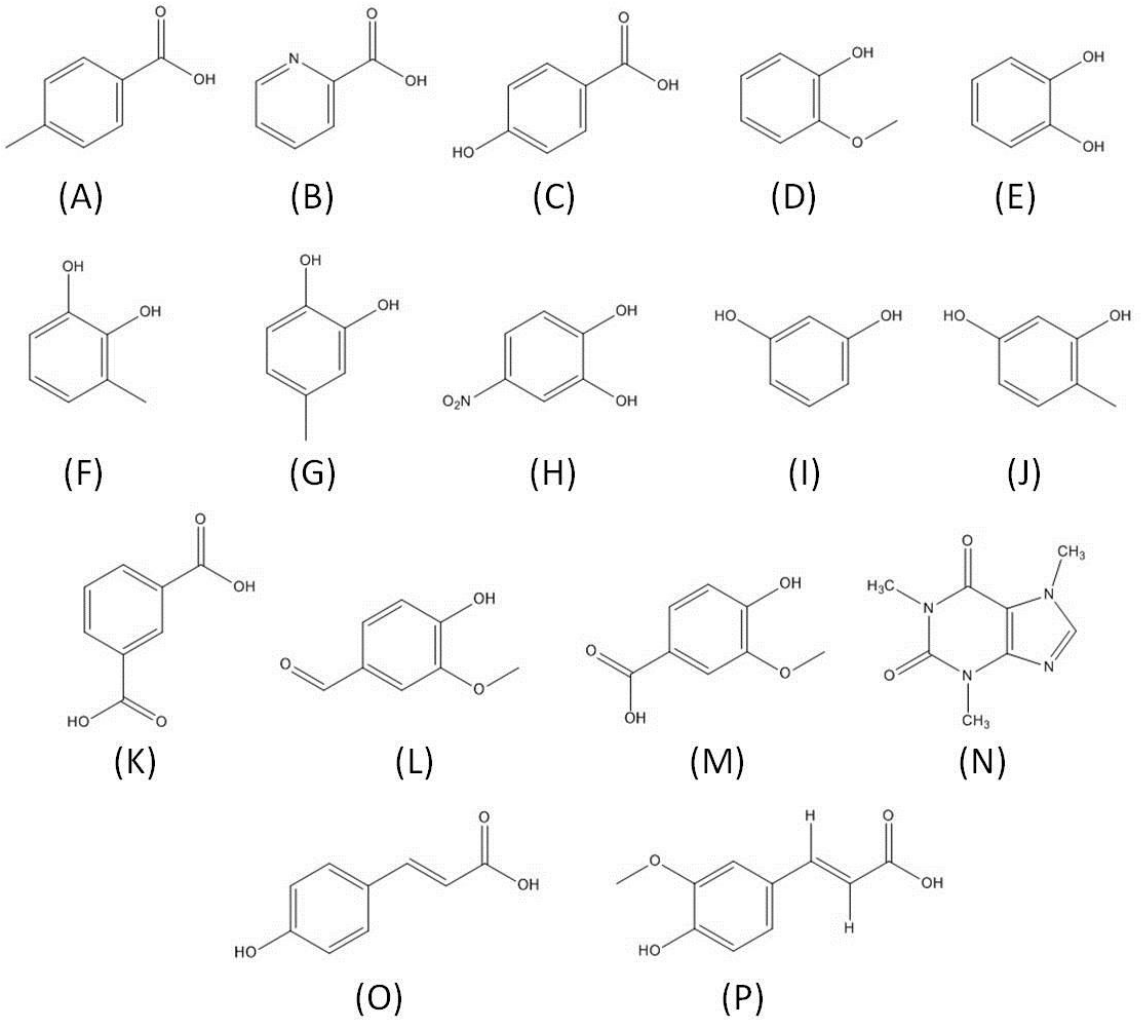


Figure 5: *In-silico* tested potential hydrotropes: *p*-toluic acid (A), picolinic acid (B), *p*-hydroxybenzoate (C), guaiacol (D), catechol (E), 3-methylcatechol (F), 4-methylcatechol (G), 4-nitrolcatechol (H), resorcinol (I), 4-methylresorcinol (J), isophthalic acid (K), vanillin (L), vanillic acid (M), caffeine (N), *p*-coumaric acid (O), and ferrulic acid (P).

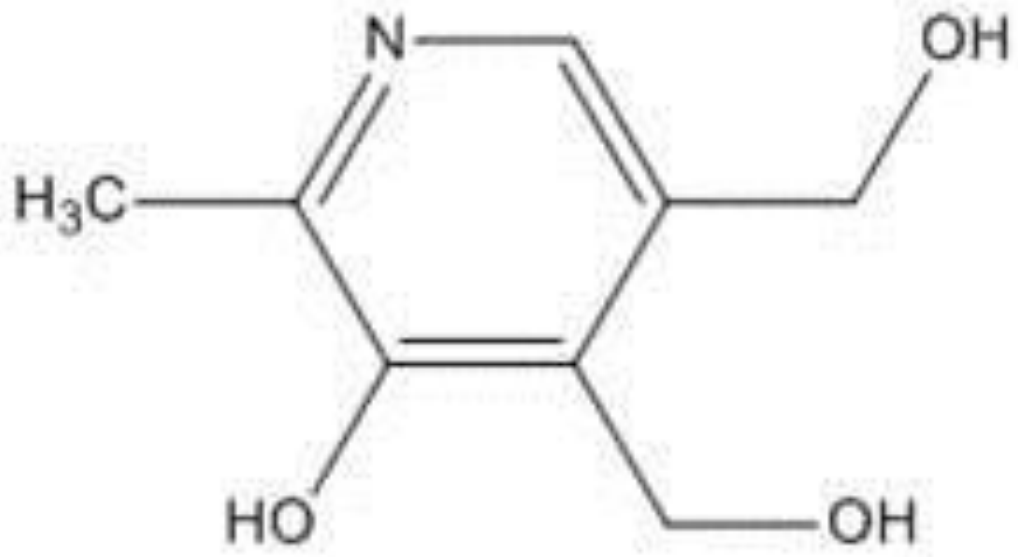


Figure 6: Pyridoxine

Table 1a: ANN training and cross-validation datasets.

Hydrotrope name		Input										Output/ target	
		[HT] (M)	MP† (°C)	log P†	pKa†	HBD††	HBA††	HAC††	Complexity††	F (ionised)	Species	[IND] (mg/mL)	% CV
Training set	**Salicylic acid**	0.1	158	2.26	2.97	2	3	10	133	1	-1	**0.04 ± 0.00**	**0.13**
		0.5										**0.17 ± 0.01**	**0.06**
		1.3										**1.14 ± 0.08**	**0.07**
		1.7										**2.17 ± 0.12**	**0.06**
	Gentisic acid	0.1	199.5	1.74	2.95	3	4	11	157	1	-1	**0.05 ± 0.00**	**0.08**
		0.5										**0.26 ± 0.02**	**0.08**
		0.9										**0.64 ± 0.05**	**0.08**
	Benzoic acid	0.1	122.4	1.87	4.19	1	2	9	104	0.996	-1	**0.20 ± 0.01**	**0.05**
		0.5										**0.79 ± 0.20**	**0.25**
		1.3										**3.04 ± 0.67**	**0.22**
		1.7										**5.87 ± 1.00**	**0.17**
	Nicotinic acid	0.1	236.6	0.36	4.75	1	3	9	114	0.986	-1	**0.42 ± 0.01**	**0.02**
		0.5										**1.41 ± 0.10**	**0.07**
		1.3										**4.11 ± 0.61**	**0.15**
		1.7										**5.44 ± 0.45**	**0.08**
	Nicotinamide	0.1	130	-0.37	3.35	1	2	9	114	0.001	1	**0.07 ± 0.00**	**0.06**
		0.5										**0.33 ± 0.02**	**0.06**
		1.3										**1.23 ± 0.13**	**0.01**
		1.7										**1.55 ± 0.32**	**0.21**
	***N,N*-diethylnicotinamide**	0.1	25	0.33	3.50	0	2	12	166	0.001	1	**0.07 ± 0.00**	**0.03**
		0.5										**0.31 ± 0.02**	**0.06**
		1.5										**2.20 ± 0.37**	**0.17**
		2.0										**3.73 ± 0.49**	**0.13**

Table 1b: ANN training and cross-validation datasets.

Hydrotrope name		Input										Output/ target	
		[HT] (M)	MP† (ºC)	log P†	pKa†	HBD††	HBA††	HAC††	Complexity††	F (ionised)	Species	[IND] (mg/mL)	% CV
Training set	*N,N*-dimethylbenzamide	0.1	44	0.62	0.00	0	1	11	137	0	0	**0.01 ± 0.00**	**0.40**
		0.5										**0.15 ± 0.00**	**0.03**
		1.2										**1.79 ± 0.14**	**0.08**
		1.5										**3.52 ± 0.34**	**0.10**
	P-toluenesulfonic acid	0.1	104.5	1.70†	-1.34	1	3	11	206	1	-1	**0.00 ± 0.00**	**0.10**
		0.5										**0.01 ± 0.00**	**0.10**
		1.3										**0.07 ± 0.01**	**0.14**
		1.6										**0.16 ± 0.01**	**0.06**
	Benzenesulfonic acid	0.1	65	0.00††	2.55	1	3	10	184	1	-1	**0.01 ± 0.00**	**0.50**
		0.5										**0.14 ± 0.03**	**0.21**
		0.9										**0.27 ± 0.01**	**0.04**
	Benzamide	0.1	129.1	0.64	0.00	1	1	9	106	0	0	**0.00 ± 0.00**	**0.25**
Cross-validation set	Salicylic acid	0.9	158	2.26	2.97	2	3	10	133	1	-1	**0.49 ± 0.02**	**0.04**
	Benzoic acid	0.9	122.4	1.87	4.19	1	2	9	104	0.996	-1	**1.32 ± 0.39**	**0.30**
	Nicotinic acid	0.9	236.6	0.36	4.75	1	3	9	114	0.986	-1	**2.69 ± 0.28**	**0.10**
	Nicotinamide	0.9	130	-0.37	3.35	1	2	9	114	0.001	1	**0.73 ± 0.04**	**0.05**
	N,N-diethylnicotinamide	1.0	25	0.33	3.50	0	2	12	166	0.001	1	**0.87 ± 0.03**	**0.03**
	N,N-*d*imethylbenzamide	0.9	44	0.62	0.00	0	1	11	137	0	0	**0.69 ± 0.06**	**0.09**
	p-toluenesulfonic acid	0.9	104.5	1.70	-1.34	1	3	11	206	1	-1	**0.03 ± 0.00**	**0.03**
	Caffeine	0.1	236.2	-0.10	10.40	0	3	14	293	1	1	**0.03 ± 0.00**	**0.04**

Hydrotrope concentration [HT], melting point (MP), hydrogen bond donor count (HBD), hydrogen bond acceptor count (HBA), heavy atom count (HAC), indomethacin concentration ± standard deviation [IND ± SD], %CV, coefficient of variation

† Obtained from EPI suite™ (version 4.00) software

†† obtained from the NCBI PubChem Database (Kim et al. 2015)

Tables 2: Physicochemical properties of *in-silico* tested hydrotropes and the corresponding predicted apparent aqueous solubility of indomethacin (IND)*

Potential hydrotrope (0.5 M concentration)	MP (°C)	logP	pKa	HBD	HBA	HAC	Complexity	F (ionised) pH 6.6	Species	Predicted IND solubility (mg/mL)
***p*-toluic acid**	179.6	2.27	4.37	1	2	10	123	0.994	-1	0.28
Picolinic acid	136.5	0.72	5.39	1	3	9	114	0.942	-1	2.44
Isophthalic acid	347.0	1.66	3.7	2	4	12	179	0.999	-1	0.45
Vanillin	81.5	1.21	7.4	1	3	11	135	0.137	-1	7.73
Vanillic acid	211.5	1.43	4.51	2	4	12	168	0.992	-1	0.74
Guiacol	32.0	1.32	9.98	1	2	9	83	0.000	-1	7.61
Resorcinol	111.0	0.8	9.32	2	2	8	64.9	0.002	-1	7.76
4-Methylresorcinol	106.0	1.6	9.77	2	2	9	92.9	0.001	-1	7.73
Catechol	105.0	0.88	9.45	2	2	8	62.9	0.001	-1	7.75
4-Methylcatechol	65.0	1.37	9.55	2	2	9	92.9	0.001	-1	7.68
3-Methylcatechol	68.0	1.8	9.59	2	2	9	92.9	0.001	-1	6.64
4-Nitrolcatechol	175.0	1.66	9.17	2	4	11	155	0.212	-1	7.70
Ferrulic acid	173.0	1.51	4.58	2	4	14	224	0.991	-1	0.42
***p*-coumaric acid**	214.0	1.46	4.64	2	3	12	178	0.989	-1	0.32
***p*-hydroxybenzoate**	214.5	1.58	4.54	1	3	10	125	0.991	-1	0.71

* Abbreviations used are as given in Table 1a-b