

Article

Solvation Enthalpies and Free Energies for Organic Solvents through a Dense Neural Network: A Generalized-Born Approach †

Sergei F. Vyboishchikov 

Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, Carrer Maria Aurèlia Capmany 69, 17003 Girona, Spain; vyboishchikov@googlemail.com

† Dedicated to Professor William E. Acree, Jr.

Abstract: A dense artificial neural network, ESE- ΔH -DNN, with two hidden layers for calculating both solvation free energies $\Delta G^\circ_{\text{solv}}$ and enthalpies $\Delta H^\circ_{\text{solv}}$ for neutral solutes in organic solvents is proposed. The input features are generalized-Born-type monatomic and pair electrostatic terms, the molecular volume, and atomic surface areas of the solute, as well as five easily available properties of the solvent. ESE- ΔH -DNN is quite accurate for $\Delta G^\circ_{\text{solv}}$, with an RMSE (root mean square error) below 0.6 kcal/mol and an MAE (mean absolute error) well below 0.4 kcal/mol. It performs particularly well for alkane, aromatic, ester, and ketone solvents. ESE- ΔH -DNN also exhibits a fairly good accuracy for $\Delta H^\circ_{\text{solv}}$ prediction, with an RMSE below 1 kcal/mol and an MAE of about 0.6 kcal/mol.

Keywords: solvation; solvation free energy; solvation enthalpy; artificial neural networks; generalized-Born method



Citation: Vyboishchikov, S.F. Solvation Enthalpies and Free Energies for Organic Solvents through a Dense Neural Network: A Generalized-Born Approach. *Liquids* **2024**, *4*, 525–538. <https://doi.org/10.3390/liquids4030030>

Academic Editors: Enrico Bodo, Juan Ortega Saavedra and William E. Acree, Jr.

Received: 22 May 2024

Revised: 3 July 2024

Accepted: 2 August 2024

Published: 12 August 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Solvation is a major effect in chemistry and has to be accounted for in chemical computations. Typically, in high demand is the solvation *free energy* $\Delta G^\circ_{\text{solv}}$, since it makes a significant contribution to the total free energy of chemical reactions in solution and is also crucially important for the estimation of partition coefficients. Nevertheless, solvation *enthalpy* $\Delta H^\circ_{\text{solv}}$ also plays a significant role, as it gives a direct measure of the heat of solvation and provides, in conjunction with $\Delta G^\circ_{\text{solv}}$, access to the solvation entropy $\Delta S^\circ_{\text{solv}}$. Hence, the development of reliable and efficient computational tools for prediction $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$ is an important task in computational chemistry.

While highly accurate explicit methods for solvation energy evaluation, such as umbrella sampling [1,2] and free energy perturbation [3], are available, they are quite costly. Consequently, most practical calculations of solvation energy rely on the continuum solvation (CS) model, in which the electrostatic E_{elst} and nonelectrostatic contributions to $\Delta G^\circ_{\text{solv}}$ are evaluated separately. The two most prominent CS approaches are the polarizable continuum model (PCM) [4–15] and the generalized Born (GB) method [16,17], including SMx [18–21]. In the PCM family of methods, the solute placed in a cavity interacts with the solvent, represented by a continuum. The cavity surface needs to be constructed and induced charges placed on it have to be computed.

In contrast, in the GB method, E_{elst} is calculated directly from atomic charges $\{Q_I\}$:

$$E_{\text{elst}}^{\text{GB}} = -\frac{1}{2} \sum_I E_I^{\text{self}} - \sum_{I < J} E_{IJ}^{\text{pair}} = -\frac{1}{2} \left(1 - \frac{1}{\epsilon}\right) \sum_I \frac{Q_I^2}{R_I} - \left(1 - \frac{1}{\epsilon}\right) \sum_{I < J} \frac{Q_I Q_J}{f_{IJ}} \quad (1)$$

where ϵ is the dielectric constant of the solvent and f_{IJ} is a function of the atomic radii R_I (in this context referred to as the Born radii) and the interatomic distance r_{IJ} . The monatomic terms (self-terms) $E_I^{\text{self}} = (1 - 1/\epsilon)Q_I^2/R_I$ in Equation (1) correspond to the solvation energy in the Born theory [22] for spherical ions. The function f_{IJ} in the pair term $E_{IJ}^{\text{pair}} = (1$

$-1/\epsilon)Q_I Q_J / f_{IJ}$ is designed to interpolate between the Coulombic limit $(1 - 1/\epsilon)Q_I Q_J / r_{IJ}$ at large r_{IJ} and the Born or Onsager limits at small r_{IJ} , when the diatomic fragment IJ collapses either into a single ion of charge $Q_I + Q_J$ or into a neutral.

The choice of effective Born radii R_I in the traditional GB methods is crucial for achieving an acceptable accuracy [23]. A typical form of f_{IJ} is the following [16]:

$$f_{IJ} = \sqrt{r_{IJ}^2 + R_I R_J \exp(-r_{IJ}^2 / (4R_I R_J))}, \quad (2)$$

although alternative expressions for f_{IJ} are also in use [18,23–26]. The GB approach has been implemented within the framework of numerous solvation energy schemes [17–19,21]. The GB-type methods are generally more efficient than the PCM-type schemes, since they avoid an explicit cavity construction and induced-charge calculation. However, the GB methods are in general less accurate than PCM.

Nonelectrostatic correction is necessary for both PCM- and GB-type methods in order to achieve reasonably accurate results. Typically, it is made of linear or nonlinear terms depending on atomic surfaces S_I , with the simplest form of the correction being $\sum_I \kappa_I S_I$ [16]. In addition, the correction term often involves the molecular volume, induced charges, and possibly other characteristics of the solute and typically contains some adjustable parameters that must be fitted on a suitable database. Thus, the CS methods inevitably have some degree of empiricity.

In our previous works [27–32], we developed a family of non-iterative methods for evaluating $\Delta G^\circ_{\text{solv}}$ (uESE—*universal Easy Solvation Energy*). The electrostatics is included by means of COSMO (*COnductor-like Screening MOdel*) [14,15], while the correction term depends on induced charges and surfaces. The atomic charges necessary for the COSMO calculation can be evaluated by means of various techniques [27,33–36], including ab initio or DFT-based, semiempirical, and even electronegativity-equalization (EE) [31] charges. However, changing the charge scheme requires re-parameterization of the correction term. Although the semiempirical and EE charges provide $\Delta G^\circ_{\text{solv}}$ estimates of reasonable quality, the accuracy of the EE-based method (ESE-EE) [31] is not as high as that of the uESE method [29] employing DFT charges. In order to improve the quality of an EE-based scheme, artificial neural networks (ANN) were introduced as a computational framework [37,38]. In the ESE-EE-DNN method [37,39], the ANN input features are the COSMO electrostatic energy, atomic cavity surface areas, total cavity volume, and induced surface charges. In contrast, the ESE-GB-DNN method [38,40] is based on GB-style input features. In particular, the number of atoms in the solute molecule, the total charge, the molecular volume, the atomic surface areas as well as GB-type self-terms and pair terms (E_I^{self} and E_{IJ}^{pair} summed over the elements) are used to represent the solute, while the solvent is described by just three parameters: the dielectric constant ϵ , the boiling point (BP), and the number of non-hydrogen atoms. The full list of solvents and their properties is given in the Supporting Information (Table S1). Since no cavity construction is needed, the ESE-GB-DNN method is substantially faster than ESE-EE-DNN (and dramatically more efficient than any DFT-based approach), allowing for a virtually instantaneous $\Delta G^\circ_{\text{solv}}$ calculation for solutes up to 100 atoms in size.

In the present paper, I further develop the GB-based approach proposed in [38]. The purpose of the present work is to create a consistent ANN-based predictor of both $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$ at 298 K for organic solvents.

Neural networks for $\Delta G^\circ_{\text{solv}}$ evaluation were developed previously. Chen et al. [41] presented a graph ANN with atomistic embedding. Vermeire and Green's approach [42] employed a directed message-passing ANN with SMILES (*Simplified Molecular-Input Line-Entry System*) and InChI (*International Chemical Identifier*) as input features. Another successful, but sophisticated, message-passing ANN for neutral solutes was presented by Low et al. [43]. Lim and Jung [44] developed a recurrent ANN and a graph convolutional ANN based on atomic vectors. Alibakhshi and Hartke [45] achieved quite accurate results within their ANN that uses a self-consistent C-PCM input. The works by Bernazzani et al. [46],

Hutchinson and Kobayashi [47], Wang et al. [48], Jaquis et al. [49], and Chung et al. [50] are also worth mentioning. The latter two address approaches that target solvation enthalpy in addition to Gibbs free energy.

2. Methods

The atomic charges used for the calculations are computed using an EE-based scheme described in detail in my previous paper [38]. This is similar but not identical to the version by Svobodová Vařeková et al. [51].

The original input features used to describe a solute closely follow the approach used in my previous paper [38]. They are as follows:

- (1) the number of atoms in the solute molecule N ;
- (2) the molecular volume V_{tot} : $V_{tot} = \sum_I V_I$;
- (3) the total surface area S_{tot} composed of atomic surfaces: $S_{tot} = \sum_I S_I$;
- (4–12) atomic surface areas summed over all the atoms of a given element L : $S_L = \sum_{I \in L} S_I$ for $L = \text{H, C, N, O, F, S, Cl, Br, and I}$. The atomic volumes V_I and surfaces S_I are efficiently calculated by simple formulas based on geometric considerations. The details are given in [38];
- (13–21) the Born-type self-terms, also summed over all the atoms of a given element L :

$$E_1^{\text{Born}}(L) = \sum_{I \in L} E_I^{\text{self}} = (1 - 1/\epsilon) \sum_{I \in L} Q_I^2 / R_I \quad (3)$$

for $L = \text{H, C, N, O, F, S, Cl, Br, I}$ calculated from the EE charges Q_I ;

- (22–51) the Born-type pair terms:

$$E_2^{\text{Born}}(L_1, L_2) = \sum_{I \in L_1} \sum_{J \in L_2} E_{IJ}^{\text{pair}} = (1 - 1/\epsilon) \sum_{I \in L_1} \sum_{J \in L_2} Q_I Q_J / f_{IJ} \quad (4)$$

The thirty L_1 – L_2 pair terms are as follows: H–H, C–C, C–H, N–N, N–H, N–C, O–O, O–H, O–C, O–N, F–F, F–H, F–C, F–O, S–H, S–C, S–N, S–O, Cl–H, Cl–C, Cl–N, Cl–O, Cl–F, Br–H, Br–C, Br–N, Br–O, Br–Cl, I–H, and I–C.

The radii R_I used in Equations (2)–(4) are unmodified Bondi [52] radii;

- (52–56) five solvent-related input features: in addition to the dielectric constant, boiling point, and the number of nonhydrogen atoms employed in my previous work [38], in this paper, the molar volume and the number of hydrogen-bond centers (the sum of the donor and acceptor centers) are also used.

Thus, the initial input feature set consists of 56 parameters. Some of them exhibit a strong correlation on the training dataset used (vide infra). For instance, V_{tot} and S_{tot} have a correlation coefficient of 0.99; S_{tot} and N have a correlation coefficient of 0.97; $E_1^{\text{Born}}(\text{Cl})$ and $E_2^{\text{Born}}(\text{Cl, H})$ have a correlation coefficient of -0.99 , etc. In total, 16 pairs of features with correlation coefficients greater than 0.96 were identified. To reduce the number of ANN parameters, principal component analysis using the `sklearn.decomposition.PCA` class from the Python Scikit-learn package [53] was applied. This process allowed truncating the 16 most correlated features, resulting in a 56×40 transformation matrix. Consequently, a vector of 40 input features is produced and fed into the ANN.

I employed a dense ANN with 40 input neurons and 2 hidden layers with 14 and 6 neurons, respectively, and an output layer with two neurons (corresponding to $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$). The *ReLU* (*Rectified Linear Unit*) activation function for the hidden layers and the linear activation for the output layer were used. Other ANN configurations were also tested, including an ANN with a single hidden layer, but the abovementioned $40 \times 14 \times 6 \times 2$ network (Figure 1) turned out to be the most accurate. It contains 678 adjustable parameters ($40 \times 14 + 14 \times 6 + 6 \times 2 = 656$ weights plus $14 + 6 + 2 = 22$ biases). The input data are min-max scaled and fed into the dense ANN described above. The ANN fitting was performed on a suitable database (vide infra) using the Nesterov-accelerated [54] Adaptive Moment Estimation algorithm [55] as implemented in the `tensorflow.keras.optimizers.Nadam` class [56], with mean squared error as the loss function and L_2 regularization with a strength $\lambda = 0.01$.

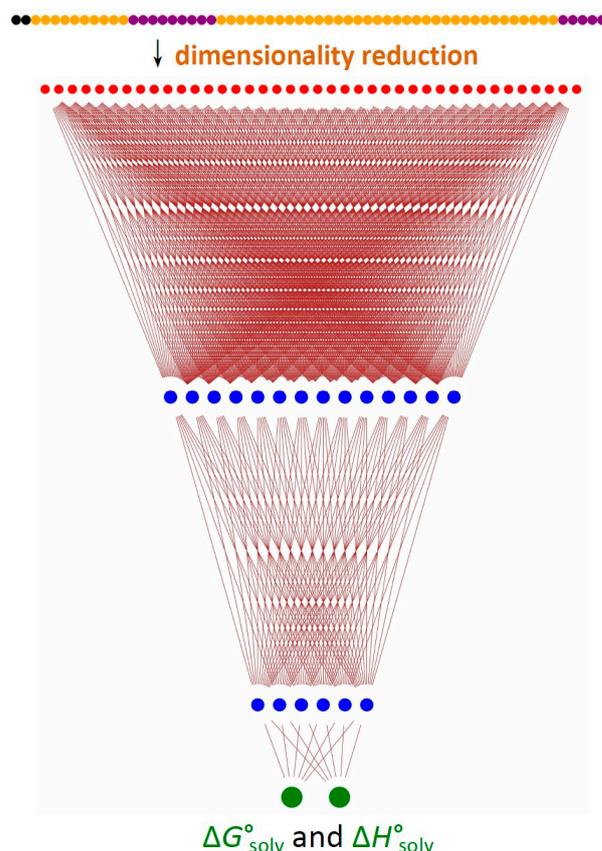


Figure 1. DNN architecture used in the present work. The first row (various colors) denotes the original 56 input features. The dimensionality reduction is achieved via a 56×40 linear transformation. The second row (40 red circles) represents the DNN input layer (40 linear combinations of the 56 initial features). The following blue circles denote two hidden layers (14 and 6 neurons, respectively). The green circles at the bottom are neurons in the output layer, corresponding to $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$.

3. Results and Discussion

3.1. Database

The datasets used for the training and testing of ESE- ΔH -DNN are based on the database of $\Delta H^\circ_{\text{solv}}$ and partition coefficients at $T = 298$ K by Prof. W. E. Acree [57,58]. The logarithms of partition coefficients $\log_{10}p$ were converted to solvation free energies: $\Delta G^\circ_{\text{solv}} = -RT\ln(10)\log_{10}p$. Subsequently, chemical names of the solutes were automatically translated into SMILES codes using the website of the National Institutes of Health [59]. The resulting SMILES codes were then automatically converted to Cartesian coordinates, and the geometries were optimized with MMFF94 (*Merck Molecular Force Field*) [60] using the rdkit.Chem.AllChem module [61]. Subsequently, the geometries were manually controlled and corrected when necessary. Atomic charges, volumes, and surfaces were computed based on these geometries. The solvent parameters (boiling points, dielectric constants, molar volume, and the number of hydrogen-bond centers) were retrieved semi-automatically from other public databases [62,63]. Finally, the resulting database contained 5201 $\Delta H^\circ_{\text{solv}}$ values and 3789 $\Delta G^\circ_{\text{solv}}$ values. It was randomly divided into a training/validation set (80%) and testing set (20%). The training/validation set was further split into a training set and a validation set, with 20% assigned for validation. Once trained, the optimized ESE- ΔH -DNN parameters (neuron weights and biases) were incorporated into a user-friendly Fortran code that reads the molecular geometry, computes the EE charges and input features, and finally evaluates $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$ via ESE- ΔH -DNN.

3.2. Training

The Cartesian coordinates of the solutes obtained as described above were used to calculate the EE charges and atomic surfaces and volumes as explained in the Section 2. Along with the five solvent-related features (52–56), they were submitted first to the dimension-reducing 56×40 linear transformation and subsequently to the ANN training. The mean-square error was used as the loss function. The learning rate was initially set to 1×10^{-3} mol²/kcal² and then reduced to 1×10^{-4} mol²/kcal² within the same training run. Since ESE- ΔH -DNN was trained simultaneously for both $\Delta H^\circ_{\text{solv}}$ and $\Delta G^\circ_{\text{solv}}$, the loss function was defined as the sum of two residual sums of squares calculated from the predicted and reference $\Delta H^\circ_{\text{solv}}$ and $\Delta G^\circ_{\text{solv}}$ values. After a number of training runs, ESE- ΔH -DNN with the fitted parameters was evaluated on the testing sets that include various classes of solvents.

3.3. Performance of ESE- ΔH -DNN

The statistical data for predicted $\Delta H^\circ_{\text{solv}}$ and $\Delta G^\circ_{\text{solv}}$ values for various classes of solvents are reported in Tables 1 and 2. Table 1 shows errors in predicted values, while Table 2 contains information about the correlation of predicted and reference values. Individual data for the training, validation, and testing sets are given in the Supporting Information (Tables S2 and S3).

Table 1. Mean signed error (MSE), mean absolute error (MAE), and root mean square error (RMSE) of the predicted solvation free energy and solvation enthalpy for various testing subsets of ESE- ΔH -DNN.

Solvent Class ^a	$\Delta G^\circ_{\text{solv}}$, kcal/mol			$\Delta H^\circ_{\text{solv}}$, kcal/mol		
	MSE	MAE	RMSE	MSE	MAE	RMSE
Alkanes (245/186)	−0.10	0.24	0.34	−0.27	0.52	0.75
Alkoxyalcohols (44/7)	−0.13	1.03	1.55	−0.02	0.40	0.43
Aromatic (81/71)	0.04	0.19	0.25	0.11	0.61	0.85
Amides (56/39)	−0.17	0.47	0.62	0.19	0.68	1.06
Ethers (64/65)	−0.21	0.34	0.70	0.22	0.58	0.80
Esters (45/8)	0.02	0.25	0.33	−0.54	0.59	0.80
Haloalkanes (123/85)	0.18	0.41	0.55	0.16	0.60	0.80
Haloaromatic (50/38)	−0.17	0.48	0.69	−0.31	0.65	0.93
Ketones (60/41)	0.05	0.26	0.36	0.41	0.73	0.90
Miscellaneous (154/472)	0.01	0.39	0.57	−0.09	0.63	1.00
ALL (922/1036)	−0.03	0.36	0.59	−0.04	0.62	0.96

^a The numbers in parentheses are the number of entries in the $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$ databases, correspondingly.

Table 2. Slope, intercept (in kcal/mol), and coefficient of determination R^2 of the predicted solvation free energy and solvation enthalpy for various testing subsets of the ESE- ΔH -DNN.

Solvent Class ^a	$\Delta G^\circ_{\text{solv}}$			$\Delta H^\circ_{\text{solv}}$		
	Slope	Intercept	R^2	Slope	Intercept	R^2
Alkanes (245/186)	0.98	−0.16	0.978	0.99	−0.33	0.972
Alkoxyalcohols (44/7)	0.55	−1.61	0.384	1.21	1.85	0.972
Aromatic (81/71)	0.96	−0.07	0.989	0.91	0.03	0.965
Amides (56/39)	0.84	−0.67	0.900	0.88	−1.13	0.969
Ethers (64/65)	0.94	−0.42	0.933	0.97	−0.09	0.976
Esters (45/8)	0.92	−0.27	0.976	0.81	−2.33	0.853
Haloalkanes (123/85)	0.92	−0.24	0.956	0.97	−0.13	0.962
Haloaromatic (50/38)	1.02	−0.11	0.944	1.08	0.46	0.949
Ketones (60/41)	0.94	−0.13	0.978	0.96	0.07	0.979
Miscellaneous (154/472)	0.89	−0.36	0.926	0.99	−0.23	0.953
ALL (922/1036)	0.93	−0.29	0.939	0.98	−0.29	0.960

^a The numbers in parentheses are the number of entries in the $\Delta G^\circ_{\text{solv}}$ and $\Delta H^\circ_{\text{solv}}$ databases, correspondingly.

3.3.1. Performance for Solvation Gibbs Free Energies

The overall quality of ESE- ΔH -DNN for $\Delta G^\circ_{\text{solv}}$ over the entire testing set is characterized by an RMSE (root mean square error) below 0.6 kcal/mol and an MAE (mean absolute error) well below 0.4 kcal/mol. This performance is very convincing: even DFT-based methods often yield an RMSE between 0.6 and 1.5 kcal/mol. For instance, for nonaqueous solvents, our uESE [29] yields an RMSE between 0.6 and 1 kcal/mol depending on the solvent class, while the SCCS model by Hille et al. [12] yields an average RMSE of about 0.8 kcal/mol. The recent methods from W. H. Green's group [50] exhibit an RMSE range from about 0.7 to 1.5 kcal/mol. It should be noted, however, that the databases used are not identical and the comparison is therefore illustrative only. Considering the ESE- ΔH -DNN results separately for various solvent classes (Table 1), very good results (with RMSE < 0.5 kcal/mol, MAE < 0.3 kcal/mol) were obtained for nonpolar solvents (alkanes and aromatic solvents), as well as for esters and ketone solvents. For other solvent classes, the results are also compelling, with an RMSE mostly below 0.7 kcal/mol. The only clearly unfavorable exception is alkoxyalcohols, for which a significant error (RMSE > 1.5 kcal/mol) is detected. The failure with alkoxyalcohols is also reflected by poor slope and R^2 values (Table 2).

The ESE- ΔH -DNN results for the entire testing set are illustrated in Figure 2a. There are 420 entries (46%) with an error below 0.2 kcal/mol and 564 entries (61%) with an error below 0.3 kcal/mol, thus confirming the overall good performance of ESE- ΔH -DNN. Still, there are some 60 (6.5%) outliers with a deviation $\Delta\Delta G^\circ_{\text{solv}} > 1$ kcal/mol. The worst failure is benzene dissolved in 2-butoxyethanol and pentane in anisole, for which ESE- ΔH -DNN fails to reproduce experimental positive $\Delta G^\circ_{\text{solv}}$. For nitromethane in triethylene glycol and SO_2F_2 in tributyl phosphate, the predicted $\Delta G^\circ_{\text{solv}}$ values are much too negative. On the contrary, for C_3F_8 dissolved in hexafluorobenzene, a too positive $\Delta G^\circ_{\text{solv}}$ is incorrectly predicted.

The results for amide solvents (Figure 2b) may also appear somewhat disturbing based on a slope value of 0.84. Indeed, ESE- ΔH -DNN tends to slightly overestimate $|\Delta G^\circ_{\text{solv}}|$ for amides, but in fact the accuracy is still quite good: even for the worst-case amide solvent, methylformamide, the RMSE is just 0.81 kcal/mol and there are only 9 (16%) outliers ($\Delta\Delta G^\circ_{\text{solv}} > 1$ kcal/mol) out of 56 entries. Other examples shown in Figure 2 are alkane and aromatic solvents. Figure 2c,d demonstrate a very good agreement between the predicted and reference data for these solvent classes. There are only five (6%) outliers in each case.

The statistical $\Delta G^\circ_{\text{solv}}$ data for all the solvents tested are summarized in Table 3. ESE- ΔH -DNN performs consistently well for all the alkane, aromatic, and ketone solvents. The results for the ethers are also very encouraging, except for anisole, where RMSE is about 1.2 kcal/mol. Ether and amide solvents are discussed above. The results for haloalkanes are also very good. Only chloroform and methylene iodide exhibit slightly larger deviations (RMSE about 0.7 kcal/mol). The haloaromatics are mostly very good (RMSE around 0.5 kcal/mol), with the only exception being perfluorobenzene. The problematic solvents within the alkoxyalcohol class are butoxyethanol, diethylene glycol, and triethylene glycol, the former being the worst case with an RMSE of almost 2.9 kcal/mol. The broad class of various solvents that are grouped together in Table 3 as *Miscellaneous* demonstrate a reliable accuracy, with an RMSE ranging from 0.2 to 1 kcal/mol. Eleven of the 24 solvents exhibit an RMSE below 0.4 kcal/mol, and eight more exhibit an RMSE below 0.8 kcal/mol. Only tributyl phosphate is slightly problematic, with an RSME of about 1 kcal/mol, which is still an acceptable accuracy for many applications.

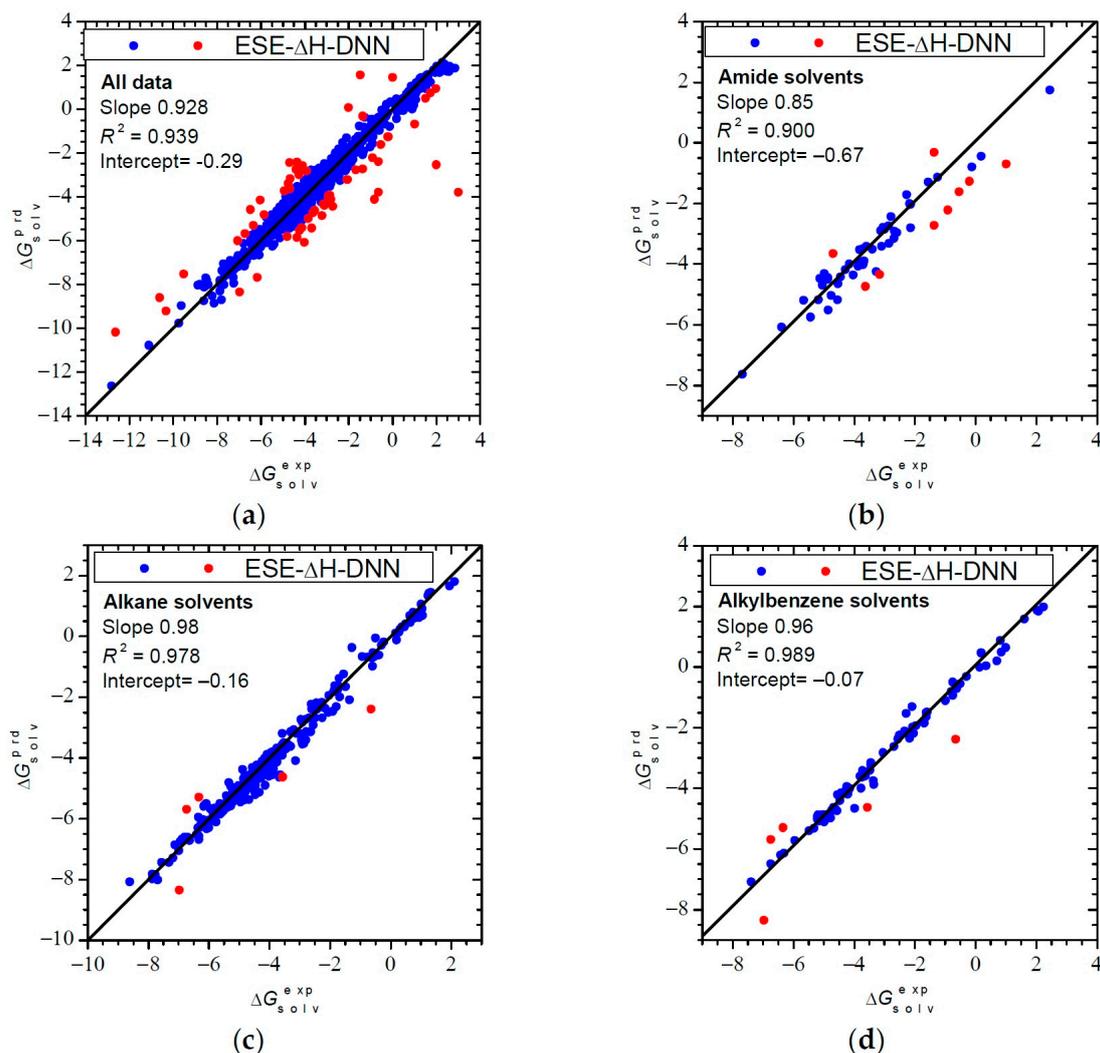


Figure 2. Solvation free energies (in kcal/mol) calculated by ESE- Δ H-DNN: (a) the entire testing set (922 entries); (b) amide solvents (56 entries); (c) alkane solvents (245 entries); (d) aromatic solvents (81 entries) versus reference values. Red points denote outliers with a deviation greater than 1 kcal/mol. The slanting line represents the identity line.

Table 3. MSE, MAE, and RMSE of the predicted solvation free energy and solvation enthalpy for various solvents (testing subsets).

Solvent ^a	$\Delta G_{\text{solv}}^{\circ}$, kcal/mol			$\Delta H_{\text{solv}}^{\circ}$, kcal/mol		
	MSE	MAE	RMSE	MSE	MAE	RMSE
<i>Alkane solvents:</i>						
Pentane (6/3)	−0.03	0.13	0.20	−0.04	0.41	0.43
Hexane (27/32)	−0.15	0.33	0.49	−0.08	0.49	0.61
Heptane (27/49)	−0.02	0.27	0.36	−0.31	0.54	0.80
Octane (27/12)	−0.08	0.16	0.23	0.01	0.31	0.38
Nonane (14/2)	−0.08	0.20	0.25	−0.36	0.36	0.46
Decane (15/11)	0.08	0.17	0.22	−0.28	0.50	0.71
Undecane (10/2)	−0.20	0.32	0.42	−0.33	0.34	0.48
Dodecane (11/8)	−0.07	0.17	0.19	−0.29	0.41	0.57
Hexadecane (65/17)	−0.06	0.22	0.31	−0.25	0.35	0.42
Tetradecane (2/3)	−0.12	0.24	0.27	0.50	0.68	0.87
Pentadecane (1/0)	−0.32	0.32	0.32			

Table 3. Cont.

Solvent ^a	$\Delta G^{\circ}_{\text{solvr}}$, kcal/mol			$\Delta H^{\circ}_{\text{solvr}}$, kcal/mol		
	MSE	MAE	RMSE	MSE	MAE	RMSE
Methylcyclohexane (5/0)	−0.29	0.29	0.50			
Cyclohexane (33/51)	−0.23	0.32	0.42	−0.48	0.67	0.96
Cyclooctane (2/0)	−0.10	0.10	0.11			
<i>Alkoxyalcohol solvents:</i>						
2-methoxyethanol (6/7)	0.23	0.27	0.36	−0.02	0.40	0.43
2-ethoxyethanol (4/0)	0.55	0.55	0.59			
2-butoxyethanol (6/0)	−1.16	1.39	2.79			
diethylene glycol (17/0)	−0.42	0.99	1.16			
triethylene glycol (11/0)	0.43	1.49	1.74			
<i>Aromatic solvents:</i>						
Benzene (13/46)	−0.09	0.17	0.23	−0.14	0.59	0.84
Toluene (25/25)	−0.03	0.21	0.29	0.56	0.65	0.86
Ethylbenzene (12/0)	0.09	0.19	0.20			
<i>o</i> -xylene (8/0)	0.08	0.12	0.15			
<i>m</i> -xylene (11/0)	0.14	0.18	0.20			
<i>p</i> -xylene (12/0)	0.17	0.24	0.31			
<i>Amide solvents:</i>						
Formamide (15/18)	−0.53	0.63	0.73	0.39	0.91	2.33
<i>Methylformamide (7/8)</i>						
<i>N</i> -methylacetamide (15/0)	−0.24	0.42	0.56			
<i>N</i> -methyl-2-pyrrolidone (19/14)	0.25	0.35	0.46	0.53	0.80	1.30
<i>Ether solvents:</i>						
diethyl ether (7/7)	−0.28	0.30	0.51	0.09	0.22	0.29
dipropyl ether (5/0)	−0.14	0.18	0.28			
diisopropyl ether (4/0)	−0.23	0.44	0.47			
dibutyl ether (6/20)	−0.17	0.24	0.41	0.22	0.53	0.72
methyl <i>tert</i> -butyl ether (5/0)	−0.24	0.39	0.45			
bis(2-ethoxyethyl) ether (1/0)	0.51	0.51	0.51			
Tetrahydrofuran (17/32)	−0.02	0.25	0.43	0.19	0.69	0.93
Tetrahydropyran (4/0)	−0.04	0.16	0.18			
anisole(15/0)	−0.52	0.56	1.22			
<i>Ester solvents:</i>						
methyl acetate (10/8)	−0.04	0.37	0.47	−0.54	0.58	0.80
ethyl acetate (13/24)	0.00	0.27	0.36	−0.15	0.56	0.74
propyl acetate (6/0)	0.11	0.16	0.17			
butyl acetate (8/0)	0.09	0.17	0.21			
pentyl acetate (7/0)	−0.02	0.23	0.26			
hexyl acetate (1/0)	−0.08	0.08	0.08			
<i>Haloalkane solvents:</i>						
Dichloromethane (9/24)	0.02	0.26	0.29	−0.59	0.68	0.87
Chloroform (48/21)	0.34	0.51	0.67	0.77	0.86	1.08
carbon tetrachloride (45/40)	0.21	0.30	0.40	0.29	0.42	0.54
1-chlorobutane (8/0)	0.02	0.35	0.39			
Dibromomethane (2/0)	0.32	0.42	0.52			
Bromoethane (2/0)	0.13	0.39	0.41			
methylene iodide (9/0)	−0.57	0.63	0.71			
<i>Haloaromatic solvents:</i>						
Fluorobenzene (3/0)	−0.54	0.54	0.58			
Chlorobenzene (21/30)	−0.30	0.41	0.53	−0.28	0.64	0.96
Bromobenzene (14/0)	−0.35	0.44	0.53			
Iodobenzene (8/0)	−0.08	0.30	0.45			
Hexafluorobenzene (4/0)	1.24	1.24	1.69			

Table 3. Cont.

Solvent ^a	$\Delta G^{\circ}_{\text{solv}}$, kcal/mol			$\Delta H^{\circ}_{\text{solv}}$, kcal/mol		
	MSE	MAE	RMSE	MSE	MAE	RMSE
<i>Ketone solvents:</i>						
Acetone (20/27)	0.01	0.35	0.50	0.32	0.78	0.95
2-butanone (10/7)	0.18	0.21	0.28	0.35	0.41	0.52
3-pentanone (1/0)	0.21	0.21	0.21			
2-hexanone (2/0)	0.23	0.23	0.24			
4-methyl-2-pentanone (1/0)	0.29	0.29	0.29			
Cyclohexanone (13/7)	0.05	0.20	0.25	0.78	0.85	1.00
Acetophenone (8/0)	0.10	0.19	0.23			
Cyclopentanone (1/0)	−0.06	0.06	0.06			
2-methylcyclohexanone (4/0)	−0.36	0.36	0.39			
<i>Miscellaneous solvents:</i>						
Benzonitrile (9/0)	0.08	0.21	0.27			
tributyl phosphate (16/0)	−0.14	0.57	1.01			
propylene carbonate (12/17)	0.04	0.20	0.30	−0.19	0.61	0.74
carbon disulfide (5/0)	−0.36	0.43	0.51			
Triethylamine (4/10)	0.18	0.19	0.35	0.65	0.93	1.58
Ethoxybenzene (3/0)	−0.14	0.66	0.77			
2-methylpyridine (4/0)	0.40	0.41	0.47			
benzyl ether (1/0)	−0.57	0.57	0.57			
3-methylphenol (1/0)	1.87	1.87	1.87			
acetic acid (10/14)	−0.08	0.48	0.60	0.12	0.66	0.94
Nitroethane (1/0)	0.60	0.60	0.60			
benzyl alcohol (2/0)	−0.30	0.30	0.35			
Butyronitrile (9/0)	−0.11	0.16	0.20			
Aniline (11/0)	−0.44	0.61	0.69			
Nitromethane (4/9)	0.09	0.21	0.31	−0.49	0.55	0.66
Nitrobenzene (3/0)	0.68	0.68	0.76			
dimethyl sulfoxide (11/37)	0.22	0.65	0.79	−0.58	0.75	1.08
Propionitrile (4/0)	−0.03	0.34	0.39			
Acetonitrile (6/41)	0.15	0.25	0.33	−0.23	0.89	1.23
ethyl benzoate (3/0)	−0.03	0.24	0.26			
Sulfolane (18/0)	0.24	0.36	0.43			
Pyridine (12/14)	0.05	0.22	0.29	0.21	1.37	1.97
diethyl carbonate (5/15)	−0.12	0.33	0.51	0.10	0.29	0.33

^a The numbers in parentheses are the number of entries in the $\Delta G^{\circ}_{\text{solv}}$ and $\Delta H^{\circ}_{\text{solv}}$ databases, correspondingly.

3.3.2. Performance for Solvation Enthalpies

Tables 1–3 also provide statistical data for predicted solvation enthalpies $\Delta H^{\circ}_{\text{solv}}$. Overall, ESE- ΔH -DNN yields somewhat larger errors for $\Delta H^{\circ}_{\text{solv}}$ than for $\Delta G^{\circ}_{\text{solv}}$, with a global RMSE of 0.96 kcal/mol and an MAE of 0.62 kcal/mol. While it is difficult to pinpoint the exact reasons of the lower absolute accuracy of ESE- ΔH -DNN for $\Delta H^{\circ}_{\text{solv}}$, it is tempting to assume that a large total range of $\Delta H^{\circ}_{\text{solv}}$ contributes to it. In fact, $\Delta H^{\circ}_{\text{solv}}$ is virtually always more negative than $\Delta G^{\circ}_{\text{solv}}$ due to negative solvation entropy. An inspection of Figure 3 shows that larger deviations often occur for more negative $\Delta H^{\circ}_{\text{solv}}$. The worst result is obtained for benzo-15-crown-5 in propanol, with $|\Delta H^{\circ}_{\text{solv}}|$ overestimated by 6.7 kcal/mol and for glycerol in methanol and *tert*-butanol, for which $|\Delta H^{\circ}_{\text{solv}}|$ is underestimated by 5.5 and 4.9 kcal/mol, respectively. Nevertheless, it is encouraging that 859 of 1036 values exhibit a deviation within 1 kcal/mol and 603 within 0.5 kcal/mol.

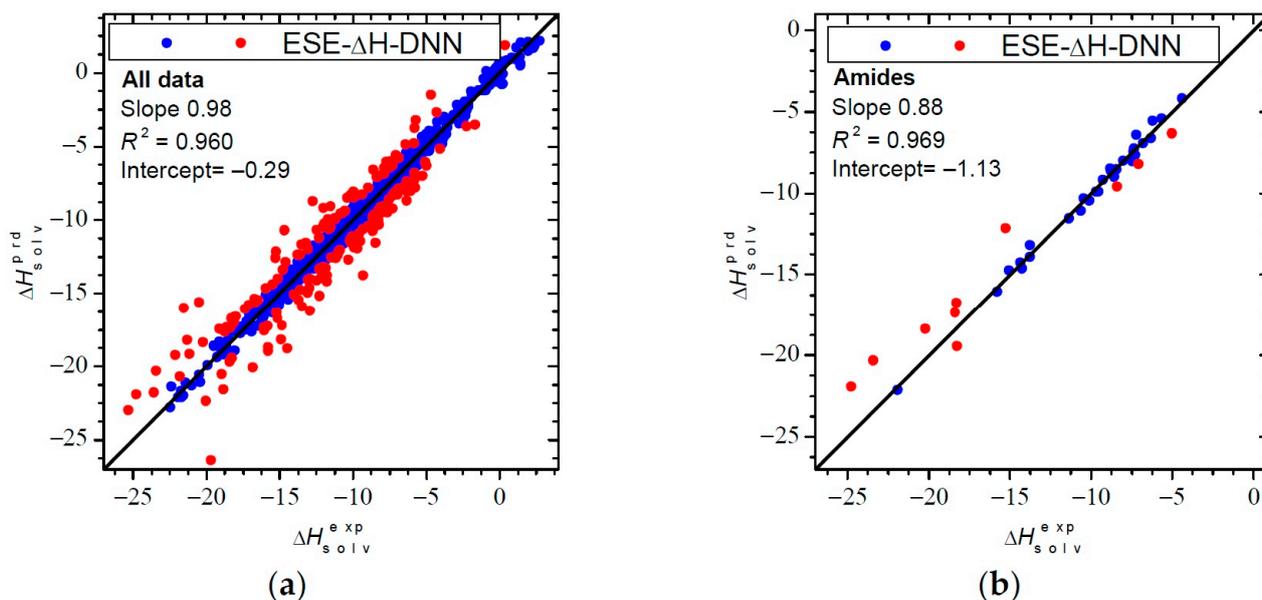


Figure 3. Solvation enthalpies (in kcal/mol) calculated by ESE- Δ H-DNN: (a) the entire testing set (1036 entries); (b) amide solvents (39 entries) versus reference values. Red points denote outliers with a deviation greater than 1 kcal/mol. The slanting line represents the identity line.

Various classes of solvent demonstrate a fairly homogeneous distribution of errors for $\Delta H^\circ_{\text{solv}}$, with RSME ranging from approximately 0.4 kcal/mol for alkoxyalkohols to about 1.1 kcal/mol for amides (see the rightmost column of Table 1). There are no solvent classes with notably large errors. It is worth noting that the performance of ESE- Δ H-DNN for $\Delta H^\circ_{\text{solv}}$ does not correlate with that for $\Delta G^\circ_{\text{solv}}$: the lowest $\Delta H^\circ_{\text{solv}}$ RSME (0.43 kcal/mol) is observed for alkoxyalkohols, which is the most problematic in terms of $\Delta G^\circ_{\text{solv}}$. The largest RMSE observed is for amides, but it does not remarkably deviate from the global average.

As for individual solvents, relatively large errors are detected for some amides and amines: formamide, pyridine, triethylamine, methylformamide, and *N*-methylpyrrolidone. While methylformamide is also problematic for $\Delta G^\circ_{\text{solv}}$, pyridine and *N*-methylpyrrolidone show quite accurate results in terms of $\Delta G^\circ_{\text{solv}}$.

In addition to the tests conducted on our randomly selected testing dataset, I performed evaluations using the structures retrieved from the Minnesota Solvation Database (MNSol) [64]. The statistical results for nonpolar solvents are presented in Table 4. Individual data are given in the Supporting Information (Table S4). Overall, these data confirm a good performance of ESE- Δ H-DNN, with a global RMSE of about 0.7 kcal/mol. An inferior performance was observed only for ethyl acetate, tributyl phosphate, and diethyl ether solvents.

Table 4. MSE, MAE, and RMSE of the solvation free energy predicted by ESE- Δ H-DNN for various nonpolar solvents (MNSol database).

Solvent ^a	$\Delta G^\circ_{\text{solv}}$, kcal/mol			Solvent ^a	$\Delta G^\circ_{\text{solv}}$, kcal/mol		
	MSE	MAE	RMSE		MSE	MAE	RMSE
acetic acid (7)	0.43	0.78	0.87	fluorooctane (6)	−0.11	0.16	0.19
aniline (9)	0.00	0.53	0.63	heptane (69)	−0.08	0.36	0.59
anisole (8)	−0.06	0.22	0.28	hexadecane (198)	−0.12	0.29	0.56
benzene (74)	−0.04	0.42	0.75	hexadecyl iodide (9)	0.44	0.44	0.51
bromobenzene (27)	−0.47	0.50	0.61	hexane (59)	−0.23	0.36	0.56
bromoform (12)	0.10	0.26	0.32	iodobenzene (20)	−0.31	0.41	0.54
bromooctane (5)	0.25	0.25	0.28	isooctane (32)	−0.37	0.38	0.46

Table 4. Cont.

Solvent ^a	$\Delta G^{\circ}_{\text{solvr}}$ kcal/mol			Solvent ^a	$\Delta G^{\circ}_{\text{solvr}}$ kcal/mol		
	MSE	MAE	RMSE		MSE	MAE	RMSE
butyl acetate (21)	0.10	0.38	0.50	isopropylbenzene (19)	0.03	0.22	0.26
butylbenzene (10)	0.42	0.42	0.47	isopropyltoluene (6)	0.25	0.25	0.30
carbon disulfide (14)	−0.51	0.58	0.67	mesitylene (7)	0.31	0.31	0.37
carbon tetrachloride (78)	0.01	0.25	0.41	nonane (26)	−0.11	0.20	0.25
chlorobenzene (38)	−0.59	0.60	0.72	nonanol (10)	0.27	0.42	0.50
chloroform (108)	0.09	0.65	0.98	octane (38)	−0.23	0.28	0.37
chlorohexane (11)	−0.07	0.17	0.20	pentadecane (9)	0.12	0.16	0.20
cyclohexane (92)	−0.49	0.53	0.81	pentane (26)	−0.41	0.43	0.47
decalin (27)	−0.26	0.36	0.56	perfluorobenzene (15)	0.94	0.94	1.02
decane (39)	−0.11	0.23	0.34	phenyl ether (6)	−0.55	0.55	0.62
decanol (11)	0.38	0.50	0.58	sec-butylbenzene (5)	0.31	0.31	0.35
dibromoethane (10)	−0.17	0.26	0.31	tert-butylbenzene (14)	0.12	0.18	0.21
dibutyl ether (14)	−0.22	0.47	0.65	tetrachloroethene (10)	0.04	0.23	0.30
dichloromethane (11)	−0.26	0.39	0.55	tetrahydrofuran (7)	0.18	0.27	0.30
diethyl ether (71)	−0.03	0.74	1.30	tetralin (9)	−0.92	0.92	1.15
diisopropyl ether (21)	−0.10	0.45	0.66	toluene (50)	0.01	0.20	0.31
dimethylpyridine (6)	−0.16	0.34	0.54	tributyl phosphate (16)	1.43	1.49	1.79
dodecane (8)	−0.29	0.35	0.45	triethylamine (7)	0.25	0.26	0.38
ethoxybenzene (7)	−0.08	0.22	0.30	trimethylbenzene (11)	0.27	0.27	0.29
ethyl acetate (23)	0.72	0.95	1.99	undecane (13)	−0.09	0.29	0.40
ethylbenzene (29)	−0.06	0.24	0.33	xylene (48)	0.08	0.25	0.34
fluorobenzene (7)	−0.66	0.66	0.79	ALL (1543)	−0.08	0.42	0.71

^a The numbers in parentheses are the number of entries in the MNSol database.

4. Conclusions

In this paper, I proposed ESE- ΔH -DNN, a dense neural network for calculating the $\Delta G^{\circ}_{\text{solvr}}$ and $\Delta H^{\circ}_{\text{solvr}}$ based on physically sound input features such as generalized-Born-type electrostatic terms, the molecular volume, and the atomic surface area of the solute as well as five characteristics of the solvent. ESE- ΔH -DNN is defined for neutral solutes in a wide range of organic solvents. It requires only the molecular geometry as the input and provides accurate results, with an average RMSE below 0.6 kcal/mol. It is particularly accurate for alkane, aromatic, ester, and ketone solvents. The limitations (which also present opportunities for further development) are that ESE- ΔH -DNN is trained only on neutral solutes and is thus not automatically applicable to ions. It also has lower accuracy for specific alkoxyalcohol solvents (butoxyethanol, triethylene glycol, and diethylene glycol), as well as for hexafluorobenzene.

ESE- ΔH -DNN also demonstrates a fairly good accuracy for $\Delta H^{\circ}_{\text{solvr}}$ prediction, although the errors are slightly larger, with an average RMSE still below 1 kcal/mol. One positive aspect is that the error is quite homogeneous, with the RMSE showing small variance among diverse solvent classes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/liquids4030030/s1>: Part 1: ESE- ΔH -DNN neural network buildup: 56 initial input vector components; 56×40 dimensionality reduction linear transformation; ESE- ΔH -DNN weights and biases; Part 2: Solvents and their properties. Part 3: Predicted and reference solvation free energies and solvation enthalpies for the training, validation, and testing sets; predicted and reference solvation free energies for nonpolar solvents from MNSol database. Reference [64] appear in the Supplementary material.

Funding: This research received no external funding.

Data Availability Statement: The executable ESE- ΔH -DNN program and a user guide are openly available for download: <https://github.com/vyboishchikov/ESE-DeltaH-DNN>. Accessed on 3 July 2024.

Acknowledgments: I am very grateful to W. E. Acree for kindly providing me with the large database of experimental $\Delta G^\circ_{\text{soln}}$ and $\Delta H^\circ_{\text{soln}}$ used in this work.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199. [[CrossRef](#)]
2. You, W.; Tang, Z.; Chang, C.A. Potential mean force from umbrella sampling simulations: What can we learn and what is missed? *J. Chem. Theory Comput.* **2019**, *15*, 2433–2443. [[CrossRef](#)] [[PubMed](#)]
3. Chipot, C.; Pohorille, A. Calculating free energy differences using perturbation theory. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer Series in Chemical Physics; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2007; Volume 86, p. 33. ISBN 540-38447-2.
4. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094. [[CrossRef](#)] [[PubMed](#)]
5. Skyner, R.E.; McDonagh, J.L.; Groom, C.R.; van Mourik, T.; Mitchell, J.B.O. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174–6191. [[CrossRef](#)] [[PubMed](#)]
6. Barone, V.; Cossi, M.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **1997**, *107*, 3210–3221. [[CrossRef](#)]
7. Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041. [[CrossRef](#)]
8. Mennucci, B.; Cancès, E.; Tomasi, J. Evaluation of solvent effects in isotropic and anisotropic dielectrics, and in ionic solutions with a unified integral equation method: Theoretical bases, computational implementation and numerical applications. *J. Phys. Chem. B* **1997**, *101*, 10506–10517. [[CrossRef](#)]
9. Cossi, M.; Barone, V.; Robb, M.A. A direct procedure for the evaluation of solvent effects in MC-SCF calculations. *J. Chem. Phys.* **1999**, *111*, 5295–5302. [[CrossRef](#)]
10. Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct. (Theochem)* **1999**, *464*, 211–226. [[CrossRef](#)]
11. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396. [[CrossRef](#)]
12. Pomogaeva, A.; Chipman, D.M. Hydration energy from a composite method for implicit representation of solvent. *J. Chem. Theory Comput.* **2014**, *10*, 211–219. [[CrossRef](#)]
13. Hille, C.; Ringe, S.; Deimel, M.; Kunkel, C.; Acree, W.E.; Reuter, K.; Oberhofer, H. Generalized molecular solvation in non-aqueous solutions by a single parameter implicit solvation scheme. *J. Chem. Phys.* **2019**, *150*, 041710. [[CrossRef](#)]
14. 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.* **1993**, *2*, 799–805. [[CrossRef](#)]
15. Klamt, A. The COSMO and COSMO-RS solvation models. *WIREs Comput. Mol. Sci.* **2011**, *1*, 699–709. [[CrossRef](#)]
16. Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T.J. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129. [[CrossRef](#)]
17. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Generalized Born solvation model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609–620. [[CrossRef](#)] [[PubMed](#)]
18. Cramer, C.J.; Truhlar, D.G. General parameterized SCF model for free energies of solvation in aqueous solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305–8311. [[CrossRef](#)]
19. Cramer, C.J.; Truhlar, D.G. An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **1992**, *256*, 213–217. [[CrossRef](#)]
20. Cramer, C.J.; Truhlar, D.G. A universal approach to solvation modeling. *Acc. Chem. Res.* **2008**, *41*, 760–768. [[CrossRef](#)]
21. Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Uniform treatment of solute–solvent dispersion in the ground and excited electronic states of the solute based on a solvation model with state-specific polarizability. *J. Chem. Theory Comput.* **2013**, *9*, 3649–3659. [[CrossRef](#)]
22. Born, M. Volumen und Hydratationswärme der Ionen. *Z. Physik* **1920**, *1*, 45–48. (In German) [[CrossRef](#)]
23. Onufriev, A.V.; Case, D.A. Generalized Born implicit solvent models for biomolecules. *Annu. Rev. Biophys.* **2019**, *48*, 275–296. [[CrossRef](#)] [[PubMed](#)]
24. Grycuk, T. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. *J. Chem. Phys.* **2003**, *119*, 4817–4826. [[CrossRef](#)]
25. Grant, J.; Pickup, B.; Sykes, M.; Kitchen, C.; Nicholls, A. The Gaussian Generalized Born model: Application to small molecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4913–4922. [[CrossRef](#)] [[PubMed](#)]
26. Lange, A.W.; Herbert, J.M. Improving Generalized Born models by exploiting connections to Polarizable Continuum Models. I. An improved effective Coulomb operator. *J. Chem. Theory Comput.* **2012**, *8*, 1999–2011. [[CrossRef](#)] [[PubMed](#)]

27. Voityuk, A.A.; Vyboishchikov, S.F. A simple COSMO-based method for calculation of hydration energies of neutral molecules. *Phys. Chem. Chem. Phys.* **2019**, *21*, 18706–18713. [[CrossRef](#)] [[PubMed](#)]
28. Voityuk, A.A.; Vyboishchikov, S.F. Fast and accurate calculation of hydration energies of molecules and ions. *Phys. Chem. Chem. Phys.* **2020**, *22*, 14591–14598. [[CrossRef](#)]
29. Vyboishchikov, S.F.; Voityuk, A.A. Fast non-iterative calculation of solvation energies for water and nonaqueous solvents. *J. Comput. Chem.* **2021**, *42*, 1184–1194. [[CrossRef](#)] [[PubMed](#)]
30. Vyboishchikov, S.F.; Voityuk, A.A. Solvation free energies for aqueous and nonaqueous solutions computed using PM7 atomic charges. *J. Chem. Inf. Model.* **2021**, *61*, 4544–4553. [[CrossRef](#)]
31. Vyboishchikov, S.F. A quick solvation energy estimator based on electronegativity equalization. *J. Comput. Chem.* **2023**, *44*, 307–318. [[CrossRef](#)]
32. Vyboishchikov, S.F.; Voityuk, A.A. Noniterative solvation energy method based on atomic charges. In *Chemical Reactivity: Approaches and applications*; Kaya, S., von Szentpály, L., Serdaroğlu, G., Guo, K., Eds.; Elsevier: Amsterdam, The Netherlands, 2023; Volume 2, pp. 399–427. [[CrossRef](#)]
33. Voityuk, A.A.; Stasyuk, A.J.; Vyboishchikov, S.F. A simple model for calculating atomic charges in molecules. *Phys. Chem. Chem. Phys.* **2018**, *20*, 23328–23337. [[CrossRef](#)] [[PubMed](#)]
34. Vyboishchikov, S.F.; Voityuk, A.A. Iterative atomic-charge partitioning of valence electron density. *J. Comp. Chem.* **2019**, *40*, 875–884. [[CrossRef](#)] [[PubMed](#)]
35. Marenich, A.V.; Jerome, S.V.; Cramer, C.J.; Truhlar, D.G. Charge Model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 527–541. [[CrossRef](#)] [[PubMed](#)]
36. Kříž, K.; Řezáč, J. Reparametrization of the COSMO solvent model for semiempirical methods PM6 and PM7. *J. Chem. Inf. Model.* **2019**, *59*, 229–235. [[CrossRef](#)] [[PubMed](#)]
37. Vyboishchikov, S.F. Dense neural network for calculating solvation free energies from electronegativity-equalization atomic charges. *J. Chem. Inf. Model.* **2023**, *63*, 6283–6292. [[CrossRef](#)] [[PubMed](#)]
38. Vyboishchikov, S.F. Predicting solvation free energies using electronegativity-equalization atomic charges and a dense neural network: A generalized-Born approach. *J. Chem. Theory Comput.* **2023**, *19*, 8340–8350. [[CrossRef](#)]
39. The Program Executable Is Available Free of Charge. Available online: <https://github.com/vyboishchikov/ESE-EE-DNN> (accessed on 3 July 2024).
40. The Program Executable Is Available Free of Charge. Available online: <https://github.com/vyboishchikov/ESE-GB-DNN> (accessed on 3 July 2024).
41. Chen, Y.; Krämer, A.; Charron, N.E.; Husic, B.E.; Clementi, C.; Noé, F. Machine learning implicit solvation for molecular dynamics. *J. Chem. Phys.* **2021**, *155*, 084101. [[CrossRef](#)]
42. Vermeire, F.H.; Green, W.H. Transfer learning for solvation free energies: From quantum chemistry to experiments. *Chem. Engin. J.* **2021**, *418*, 129307. [[CrossRef](#)]
43. Low, K.; Coote, M.L.; Izgorodina, E.I. Explainable solvation free energy prediction combining graph neural networks with chemical intuition. *J. Chem. Inf. Model.* **2022**, *62*, 5457–5470. [[CrossRef](#)]
44. Lim, H.; Jung, I. MLSolvA: Solvation free energy prediction from pairwise atomistic interactions by machine learning. *J. Cheminform.* **2021**, *13*, 56. [[CrossRef](#)]
45. Alibakhshi, A.; Hartke, B. Improved prediction of solvation free energies by machine-learning polarizable continuum solvation model. *Nat. Commun.* **2021**, *12*, 3584. [[CrossRef](#)] [[PubMed](#)]
46. Bernazzani, L.; Duce, C.; Micheli, A.; Mollica, V.; Tiné, M.R. Quantitative structure–property relationship (QSPR) prediction of solvation Gibbs energy of bifunctional compounds by recursive neural networks. *J. Chem. Eng. Data* **2010**, *55*, 5425–5428. [[CrossRef](#)]
47. Hutchinson, S.T.; Kobayashi, R. Solvent-specific featurization for predicting free energies of solvation through machine learning. *J. Chem. Inf. Model.* **2019**, *59*, 1338–1346. [[CrossRef](#)] [[PubMed](#)]
48. Wang, B.; Wang, C.; Wu, K.; Wei, G.W. Breaking the polar–non-polar division in solvation free energy prediction. *J. Comput. Chem.* **2018**, *39*, 217–233. [[CrossRef](#)] [[PubMed](#)]
49. Jaquis, B.J.; Li, A.; Monnier, N.D.; Sisk, R.G.; Acree, W.E.; Lang, A.S.I.D. Using machine learning to predict enthalpy of solvation. *J. Solut. Chem.* **2019**, *48*, 564–573. [[CrossRef](#)]
50. Chung, Y.; Vermeire, F.H.; Wu, H.; Walker, P.J.; Abraham, M.H.; Green, W.H. Group contribution and machine learning approaches to predict Abraham solute parameters, solvation free energy, and solvation enthalpy. *J. Chem. Inf. Model.* **2022**, *62*, 433–446. [[CrossRef](#)] [[PubMed](#)]
51. Svobodová Vařeková, R.; Jiroušková, Z.; Vaněk, J.; Suchomel, Š.; Koča, J. Electronegativity equalization method: Parameterization and validation for large sets of organic, organohalogen and organometal molecule. *Int. J. Mol. Sci.* **2007**, *8*, 572–582. [[CrossRef](#)]
52. Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441–451. [[CrossRef](#)]
53. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. Available online: <https://scikit-learn.org/stable> (accessed on 1 March 2024).

54. Nesterov, Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Dokl.* **1983**, *27*, 372–376.
55. Dozat, T. Incorporating Nesterov Momentum into Adam. In International Conference on Learning Representations. 2016. Available online: <https://openreview.net/pdf/OM0jvwB8jIp57ZJjtNEZ.pdf> (accessed on 8 November 2023).
56. Martín, A.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: <http://tensorflow.org> (accessed on 30 April 2023).
57. Lu, J.Z.; Acree, W.E.; Abraham, M.H. Updated Abraham model correlations for enthalpies of solvation of organic solutes dissolved in benzene and acetonitrile. *Phys. Chem. Liquids* **2019**, *57*, 84–99. [[CrossRef](#)]
58. Acree, W.E., Jr.; (University of North Texas, Denton, TX, USA). Personal communication, 2024.
59. Available online: <http://cactus.nci.nih.gov/chemical/structure> (accessed on 1 March 2024).
60. Halgren, T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519. [[CrossRef](#)]
61. Landrum, G. Rdkit Documentation, Release 1 September 2019. 2019. Available online: <https://buildmedia.readthedocs.org/media/pdf/rdkit/latest/rdkit.pdf> (accessed on 1 March 2024).
62. Available online: <https://webbook.nist.gov> (accessed on 1 March 2024).
63. Available online: <https://www.stenutz.eu/chem> (accessed on 1 March 2024).
64. Marenich, A.V.; Kelly, C.P.; Thompson, J.D.; Hawkins, G.D.; Chambers, C.C.; Giesen, D.J.; Winget, P.; Cramer, C.J.; Truhlar, D.G. Minnesota Solvation Database—Version 2012, University of Minnesota, 26 November 2012. Available online: https://conservancy.umn.edu/bitstream/handle/11299/213300/MNSolDatabase_v2012.zip (accessed on 10 June 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.