ChemPhysChem

Research Article
doi.org/10.1002/cphc.202300566

Chemistry Europe
European Chemical Societies Publishing

www.chemphyschem.org

Special Issue

Insights into Hildebrand Solubility Parameters – Contributions from Cohesive Energies or Electrophilicity Densities?**

Ramón Alain Miranda-Quintana,*[a] Lexin Chen,[b] and Jens Smiatek*[c]

We introduce certain concepts and expressions from conceptual density functional theory (DFT) to study the properties of the Hildebrand solubility parameter. The original form of the Hildebrand solubility parameter is used to qualitatively estimate solubilities for various apolar and aprotic substances and solvents and is based on the square root of the cohesive energy density. Our results show that a revised expression allows the replacement of cohesive energy densities by electrophilicity densities, which are numerically accessible by simple DFT calculations. As an extension, the reformulated expression provides a deeper interpretation of the main contributions and, in particular, emphasizes the importance of charge transfer mechanisms. All calculated values of the Hildebrand parameters for a large number of common solvents are compared with experimental values and show good agreement for non- or moderately polar aprotic solvents in agreement with the original formulation of the Hildebrand solubility parameters. The observed deviations for more polar and protic solvents define robust limits from the original formulation which remain valid. Likewise, we show that the use of machine learning methods leads to only slightly better predictability.

Introduction

Solubilities of substances in different solvents are of crucial importance for many industrial applications and chemical syntheses.[1–3] Over the last decades, various numerical, theoretical and experimental approaches were introduced to estimate solubilities and to identify suitable solvent-substance pairs.[2–14] Most numerical methods are quite advanced and rely on thermodynamic or quantum chemical concepts. In contrast, there are also empirical approaches such as the "Like dissolves Like" rule which are based on simplified assumptions.[1,15,16] In general, the principles underlying these empirical considerations are based on the experience that molecular similarities between solutes and solvents favor high solubilities.

Following closely the LDL concept, Hildebrand solubility parameters (HSPs) were introduced as a useful extension for practical applications.[1,2,17–21] Based on the regular solution theory, the HSPs are represented by single numerical values to estimate the solubility of certain components in different solvents. More specifically, the HSPs are calculated from the square root of the cohesive energy densities and compared for the different components in a mixture. Good solubility is usually achieved when both HSPs are nearly comparable. Although this simple approach relies on many approximations such as the regular arrangement of the components on a lattice and the strict consideration of aprotic and apolar compounds,[20] it has already been shown to be effective in various contexts[17] and specifically for pharmaceutical applications.[22] This is all the more surprising since all electrostatic and dipolar interactions are ignored.[20] Accordingly, it has already been mentioned that the Hildebrand solubility parameters are applicable only to a very limited set of uncharged substances with low dipole moments and negligible hydrogen donor or acceptor properties. Despite the fact that simple approaches to estimating solubilities are attractive, it should be noted that real solutions have a large number of complex interactions leading to ideal and non-ideal effects. Therefore, it is highly questionable whether a single parameter will provide reliable estimates of good solubilities. Nevertheless, Hildebrand solubility parameters are still widely used today as a first approach to identify suitable solvents in various combinations.[21–23] In addition, many computer studies have already provided HSP values from DFT calculations, since experimental values for specific substances are usually difficult to determine.[24–28]

[a] Prof. Dr. R. A. Miranda-Quintana
Department of Chemistry and Quantum Theory Project, University of Florida, Gainesville, FL 32603, USA
E-mail: quintana@chem.ufl.edu
Homepage: https://quintana.chem.ufl.edu

[b] L. Chen
Department of Chemistry, University of Florida, Gainesville, FL 32603, USA

[c] Priv.-Doz. Dr. J. Smiatek
Institute for Computational Physics, University of Stuttgart, D-70569 Stuttgart, Germany
E-mail: smiatek@icp.uni-stuttgart.de
Homepage: https://www2.icp.uni-stuttgart.de/icp/Jens_Smiatek

[**] A previous version of this manuscript has been deposited on a preprint server (DOI: https://dx.doi.org/10.26434/chemrxiv-2023-33l70)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cphc.202300566

Special Issue An invited contribution to a Special Collection celebrating the 30th anniversary of the Physicochemistry of Interfacial Phenomena Section of the Polish Chemical Society

As another option for the study of solute properties, extensions of conceptual DFT for molecular solutions were recently introduced.[29–35] Previously, conceptual DFT was successfully applied for explaining various empirical observations in terms of chemical reactions and molecular binding behavior.[36–49] In particular, conceptual DFT introduces certain chemical reactivity descriptors that provide reliable estimates for the chemical hardnesses, electronegativities and reaction energies for the chemical species under consideration. Within this context, the formation of solvent-solute complexes is described by chemical reactions, such that charge transfer mechanisms in terms of directional solvation bonds between the species become important.[29] The corresponding approach was already successfully used for the study of donor numbers, specific ion effects and the identification of suitable solvent-solute combinations.[29–32,35]

In this article, we use some expressions from conceptual DFT for studying the contributions to HSPs. Our approach allows us to identify the cohesive energy densities with simple electrophilic densities computed from DFT calculations. The calculated values of Hildebrand parameters are compared with experimental values for more than 45 solvents and show good agreement for non- or moderately polar aprotic solvents. The observed deviations for more polar and protic solvents define robust limits where the Hildebrand approach remains valid. The use of machine learning methods for more accurate predictions also for protic and polar solvents shows only minor improvements in this respect. Accordingly, our assumptions are not suitable for extending the narrow limits of Hildebrand solubility theory. However, in particular the importance of electronic charge transfer mechanisms for solubility phenomena and fundamental parameters becomes more apparent. Hence, empirical concepts such as cohesive energy densities can be replaced by robust molecular parameters in terms of well-defined electrophilicities.

The article is organized as follows. The next section discusses the theoretical background of the HSPs and the conceptual DFT framework. All numerical details will be presented in section 3. The revised expressions for the HSPs and mixing enthalpies in combination with the computational results will be shown in section 4. We conclude and summarize in the last section.

Theoretical Background

In this section, we will introduce the main idea behind the HSPs and key concepts from conceptual DFT.

Hildebrand Solubility Parameters

The HSPs for a specific molecular substance are defined by[20]

$$\delta_H = \sqrt{\frac{\Delta H_V - RT}{V_m}} \tag{1}$$

where ΔH_V denotes the heat of vaporization, V_m the molecular volume of the molecule and R and T the universal gas constant and the actual temperature, respectively.[1,2,17–20] The expression under the root can be understood as cohesive energy density

$$\rho_H = \frac{\Delta H_V - RT}{V_m} \tag{2}$$

in terms of the energy of vaporization for a species that estimates the degree of van der Waals forces between the molecules in pure solid form. In consequence, the cohesive energy density is the amount of energy needed to completely remove one unit volume of molecules from their neighbours in a solid conformation. As can be seen from the original interpretation, any electrostatic or dipolar interactions are ignored such that the approach is solely applicable for apolar and aprotic species.

Noteworthy, in its original formulation, it was assumed that there is a correlation between vaporization and solubility behavior. This assumption is based on the idea that the same intermolecular attractive forces have to be overcome to vaporize a number of molecules as to dissolve it.[20] In consequence, it is proposed that the molecules from one substance in the mixture are physically separated from the molecules of the other component. Such questionable assumptions stimulated a lot of discussion and finally led to the development of more advanced parametric descriptions.

As an extension of HSPs, there also exists a direct connection to the thermodynamics of mixing in combination with regular solution and Flory-Huggins mixing theory.[50–52] As can be shown,[50] the entropy of mixing for two solvents is defined by

$$\Delta S_m = -k_B T(N_1 \ln \phi_1 + N_2 \ln \phi_2) \tag{3}$$

with Boltzmann constant k_B and the number of solvent molecules N_1 and N_2 and their respective mole fraction $\phi_1 = N_1/N$ and $\phi_2 = N_2/N$ with $N_1 + N_2 = N$ where N denotes the total number of solvent molecules. The resulting enthalpy changes can be written as

$$\Delta w = w_{12} - \frac{1}{2}(w_{11} + w_{22}) \tag{4}$$

where w_{11} and w_{22} denote the interaction energy between neighboring species of likewise molecules and w_{12} between different molecules. The total number of such contacts can be written as $zN_2\phi_1 = zN_1\phi_2$ where z is the assumed coordination number of surrounding molecules around a molecule from regular solution theory.[19] The resulting enthalpy change for each of these contacts can be written as $\Delta H_m = zN_1\phi_2\Delta w$ with the Flory-Huggins interaction parameter[50]

$$\xi_{12} = \frac{z\Delta w}{k_B T} \tag{5}$$

which leads to

$$\Delta H_m = k_B T N_1 \phi_2 \xi_{12}. \tag{6}$$

Noteworthy, the Flory-Huggins interaction parameter can also be written as

$$\xi_{12} \sim (\delta_{H,1} - \delta_{H,2})^2 \tag{7}$$

which introduces Hildebrand solubility parameters for two species.[53] The combination of Eqn. (6) and Eqn. (7) leads to the resulting enthalpy change upon mixing the two species[18,54]

$$\Delta H_m \sim RT n_1 \phi_2 (\delta_{H,1} - \delta_{H,2})^2 \tag{8}$$

with the molar concentration of n_1. Further evaluation of the free energy change of mixing in terms of $\Delta G_m = \Delta H_m - T\Delta S_m$ provides the expression

$$\Delta G_m \sim RT\left(n_1 \phi_2 (\delta_{H,1} - \delta_{H,2})^2 + n_1 \ln \phi_1 + n_2 \ln \phi_2\right) \tag{9}$$

which also relies on a lot of crucial approximations. However, as can be seen from Eqn. (9), the free energy decreases with smaller differences between $\delta_{H,1}$ and $\delta_{H,2}$ which indicates a more favorable mixing process. Such assumptions reveal surprisingly good estimates as long as the components of the mixture are not too polar with low or moderate dipole moments or are affected by hydrogen bonding.[2]

Conceptual DFT

In general, the conceptual DFT approach for the study of solution effects is based on the interpretation of solvation phenomena as chemical reactions. In principle, every chemical process can be interpreted as a reaction of the system to changes in the number of electrons ΔN in combination with an external potential $\Delta v(\mathbf{r})$.[36,48,49] The resulting energy change for a generic X + Y→XY reaction between the reactants X and Y reads[36,46,55,56]

$$\Delta E_{XY} = -\frac{1}{2}\frac{(\chi_X - \chi_Y)^2}{\eta_X + \eta_Y} \tag{10}$$

with the electronegativity χ as the first derivative of the ground state energy E_Z in the number of electrons N_Z as defined by

$$\chi_Z = \left(\frac{\partial E_Z}{\partial N_Z}\right)_{v(\mathbf{r})} \approx \frac{I_Z + A_Z}{2} \approx -\frac{\varepsilon_Z^{LUMO} + \varepsilon_Z^{HOMO}}{2} \tag{11}$$

for species $Z = X, Y$. In addition, the second derivative

$$\eta_Z = \left(\frac{\partial^2 E_Z}{\partial N_Z^2}\right)_{v(\mathbf{r})} \approx I_Z - A_Z = \varepsilon_Z^{LUMO} - \varepsilon_Z^{HOMO} \tag{12}$$

can be interpreted as the chemical hardness of the species.[36,57] In its original formulation, it also includes a prefactor of 1/2 which is often ignored for consistency with the HOMO-LUMO gap.[36] The parameters I_Z and A_Z are the first vertical ionization potential and electron affinity, respectively, and ε_Z^{HOMO} and ε_Z^{LUMO} the observed energies of the highest occupied (HOMO) and the lowest unoccupied (LUMO) molecular orbitals. In terms of the Koopmans and the Janak theorem,[58,59] one can also define $I_Z = -\varepsilon_Z^{HOMO}$ and $A_Z = -\varepsilon_Z^{LUMO}$. Furthermore, one can define the individual electrophilicities in terms of[38,60]

$$\omega_Z = \frac{\chi_Z^2}{2\eta_Z} \tag{13}$$

which leads to the interpretation of an electrophilic ligand immersed in an idealized zero-temperature free electron sea of zero electronegativity $\chi_S = 0$ and vanishing chemical hardness $\eta_S = 0$.[36] In presence of enough electrons to saturate the considered molecule in terms of full charge transfer $\Delta N_{max} = \chi_Z/\eta_Z$,[36] the reaction energy according to Eqn. (10) can be written as

$$\Delta E_{ZS} = -\frac{1}{2}\frac{(\chi_Z - \chi_S)^2}{\eta_Z + \eta_S} = -\frac{1}{2}\frac{\chi_Z^2}{\eta_Z} \tag{14}$$

which exactly corresponds to Eqn. (13) in terms of the negative electrophilicity $\Delta E_{ZS} = -\omega_Z$.

Numerical Details

The values for the electronegativity, the chemical hardness, the frontier molecular orbital energies and the reaction energies were computed by standard DFT calculations for isolated and geometry-optimized species with the software package Orca 4.2.1.[61,62] The individual molecules were optimized at the DFT level of theory using the PBE functional[63] in combination with the def2-TZVPP[64] basis set and the auxiliary Coulomb fitting basis set def2/J.[65] All calculations utilized an atom-pairwise dispersion correction with the Becke-Johnson damping scheme.[66,67] In addition, all calculations were performed for isolated species which were geometry-optimized until an energy convergence criterion of $5 \cdot 10^{-6}$ eH was reached. The presence of a continuous dielectric background ($\varepsilon_r = 80.4$) with the CPCM method[68] was further considered. All corresponding molecular volumes for the geometry-optimized species were estimated by the GEPOL approach.[69,70] The lists of considered solvents can be found in Tables 1 and 2. We assumed a temperature of $T = 300$ K for all calculations. The choice of the continuum solvent is mainly due to the fact that the proposed method should also be applicable to molecular ions such as carboxylic acids. In the presence of solvents with low dielectric constant or in combination with gas phase calculations, the degree of dissociation and the corresponding structure cannot be correctly represented. Accordingly, we decided to use a continuum solvent which allows to realize the different conformations of molecules and thus transfer the corresponding linear fit model after validation for further more complex and charged solutes. Moreover, the consideration of small molecular size solvents such as water also allows proper

Table 1. Experimental Hildebrand solubility parameters δ_H from References [20,83] in combination with HOMO (E^{HOMO}) and LUMO energies (E^{LUMO}) as well as molecular volumes V_m from the GEPOL method[70] for different solvents with $D<5$ Debye reflecting poor protic behavior (0) in accordance with Reference [83].

Solvent	δ_H [cal$^{1/2}$/cm$^{3/2}$]	ε^{HOMO} [eV]	ε^{LUMO} [eV]	V_m [10^{-3} nm^3]	D [Debye]	Protic
n-hexane	7.24	−7.25	0.57	968.74	0.00	0
n-heptane	7.40	−7.14	0.59	1108.16	0.11	0
n-octane	7.60	−6.88	0.55	1249.95	0.06	0
methylcyclohexane	7.80	−6.76	0.59	1000.42	0.61	0
dodecane	7.90	−6.93	0.56	1819.93	0.06	0
cyclohexane	8.18	−6.86	0.53	856.33	0.00	0
mesitylene	8.80	−5.76	−1.12	1167.05	0.13	0
ethylbenzene	8.80	−6.02	−1.16	1028.77	0.65	0
o-xylene	8.80	−5.85	−1.20	1020.64	1.00	0
decaline	8.80	−6.42	0.60	1319.28	0.00	0
toluene	8.91	−6.01	−1.22	884.97	0.65	0
1,1-dichloroethylene	9.10	−6.35	−1.52	650.32	1.66	0
benzene	9.15	−6.36	−1.24	745.46	0.00	0
trichloroethylene	9.20	−6.09	−1.69	793.01	1.17	0
chloroform	9.21	−7.40	−1.95	700.70	1.51	0
tetrachloroethylene	9.30	−6.05	−1.91	931.96	0.00	0
styrene	9.30	−5.70	−2.05	991.23	0.28	0
bromobenzene	9.50	−6.11	−1.50	934.50	2.41	0
chlorobenzene	9.50	−6.15	−1.48	891.63	2.23	0
tetrahydronaphtalene	9.50	−5.80	−1.18	1197.23	1.20	0
1,1,2,2-tetrachloroethane	9.70	−7.13	−1.61	980.39	0.00	0
ethylene dibromide	9.70	−6.82	−1.77	777.02	0.01	0
1,2-dichloroethane	9.80	−7.21	−1.00	691.46	0.00	0
naphtalene	9.90	−5.53	−2.13	1126.65	0.00	0
2-nitropropane	9.90	−6.94	−3.14	752.14	4.80	0
carbon disulfide	10.00	−6.71	−2.75	563.24	0.00	0
o-dichlorobenzene	10.00	−6.17	−1.66	1033.43	3.25	0
1-bromonaphtalene	10.60	−5.58	−2.34	1309.90	2.42	0
nitroethane	11.10	−6.93	−3.11	602.69	4.64	0
nitromethane	12.70	−6.96	−3.21	464.46	4.39	0

estimates of the molecular surface for the solutes when compared to more bulky solvent molecules.

Overall, it has to be noted that different methods, basis sets, and functionals lead to slightly different results. However, we can ignore the consideration of these factors, since we are only interested in studying the electronic contributions to HSPs. Accordingly, we can assume that the generically selected models and values provide a sufficiently accurate qualitative description of the contributions. Since the applicability of the HSPs is appropriately limited to a few aprotic and apolar compounds, we cannot assume that the introduction of highly accurate predictive models offers a decisive advantage. Moreover, it has to be noted that one can also use more advanced definitions for the electron affinities, ionization potentials and thus also the corresponding paramters from conceptual DFT.[36,71] However, previous comparisons already revealed a good agreement with more advanced calculations approaches,[72,73] such that we decided to use the simple approach which gives us the chance for fast high throughput calculations in agreement with the vast majority of conceptual DFT calculations.[36]

Nevertheless, we also applied machine learning approaches for a more accurate mapping of individual features to the experimental values of the HSPs. Hence, we used the full list of solvents with the corresponding features for the HOMO and LUMO energies, the dipole moment, the molecular volume, the electronegativity, the chemical hardness and the corresponding self energy (Eqn. (14)). As most accurate method for predictions, we identified the Extra Trees Method[74] with a number of estimators of 100 and a minimum sample split of 2. In previous publications, it was shown that decision-tree based methods usually provide the most accurate results for small datasets.[75] The source code was written in Python 3.9.1[76] in combination

Table 2. Experimental Hildebrand solubility parameters δ_H from References [20,83] in combination with HOMO (E^{HOMO}) and LUMO energies (E^{LUMO}) as well as molecular volumes V_m from the GEPOL method[70] for different poor (0), moderate (1) or strong (2) protic solvents with $D > 5$ Debye in accordance with Reference [83].

Solvent	δ_H [cal$^{1/2}$/cm$^{3/2}$]	ε^{HOMO} [eV]	ε^{LUMO} [eV]	V_m [10^{-3} nm^3]	D [Debye]	Protic
diethylether	7.62	−5.80	0.62	753.46	1.61	1
benzonitrile	8.40	−6.70	−2.43	914.50	6.28	0
butanone	9.27	−5.73	−1.68	710.98	4.21	1
tetrahydrofurane	9.52	−5.56	0.62	663.70	2.43	1
acetone	9.77	−5.82	−1.70	570.23	4.05	1
methylene chloride	9.93	−7.21	−1.17	554.08	2.20	1
nitrobenzene	10.00	−6.86	−3.63	947.41	6.21	0
acrylonitrile	10.50	−7.19	−2.47	538.62	5.29	0
morpholine	10.52	−5.25	0.61	758.45	2.07	1
pyridine	10.61	−6.12	−1.87	711.40	3.15	2
proprionitrile	10.80	−7.97	−0.47	570.44	5.34	0
2-propanol	11.60	−6.24	0.49	611.22	2.18	2
dimethyl formamide	12.14	−5.66	−0.77	677.80	4.84	1
ethanol	12.92	−6.24	0.55	480.17	2.14	2
dimethyl sulfoxide	12.93	−5.50	−0.10	644.48	5.56	1
ethylene glycol	16.30	−6.15	0.39	537.70	2.67	1

with the modules NumPy 1.19.5,[77] Scikit-earn 1.0.1[78] and Pandas 1.2.1.[79]

Results

Combination of Hildebrand Solubility Parameters with Conceptual DFT

In the following, it is assumed that the species Z in liquid state (l) is vaporized to the gas phase state (g) in terms of

$$Z_l \rightarrow Z_g \tag{15}$$

where the corresponding reaction energy can be estimated from conceptual DFT calculations. In more detail, we assume that the reaction energy of species Z_g with the reference state (R) in accordance with Eqn. (10) can be written as

$$\Delta E_{ZR} = -\frac{1}{2}\frac{(\chi_Z - \chi_R)^2}{\eta_Z + \eta_R} = -\frac{1}{2}\frac{\chi_Z^2}{\eta_Z} \tag{16}$$

where it is assumed that the electronegativity and chemical hardness of the reference state are $\chi_R = \eta_R = 0$. The previous relation exactly corresponds to Eqn. (13) and Eqn. (14) such that $\Delta E_{ZR} = \Delta E_{ZL} = -\omega_Z$. As a side remark, one has to notice that the back-reaction from the gas phase to the liquid state results in $E_{ZZ} = 0$. If we assume that $\Delta E_{ZR} \propto -\Delta H_V$ in accordance with previous considerations of enthalpic contributions,[29,30] one can immediately see from Eqn. (2) that

$$\rho_H \propto \frac{\omega_Z - RT}{V_m} \tag{17}$$

such that the cohesive energy density can also be interpreted as molecular electrophilicity density which yields

$$\delta_H \propto \sqrt{\frac{\omega_Z - RT}{V_m}} \tag{18}$$

as a reasonable estimate for the vaporization energy and the HSPs. Due to the fact that certain factors such as the latent heat of vaporization in combination with reasonable finite size scaling corrections are missing, one can further redefine

$$\delta_H = p\sqrt{\frac{\omega_Z - RT}{V_m}} + \delta_H^0 \tag{19}$$

with the fitted scaling factor p and the missing contributions δ_H^0. As already mentioned, the HSPs rely on a series of approximations, such that the further introduction of electrophilicities adds another level of uncertainty. Due to these reasons, one can assume that the experimental HSP values as obtained from Eqn. (1) might deviate from Eqn. (18). However, the corresponding values of p and δ_H^0 do not significantly alter the results, as the main correlation comes from the electrophilicity densities. As mentioned earlier, our goal is also more in the study of electronic contributions to solution parameters as opposed to accurate predictions for HSPs. After combination of Eqn. (18) with Eqn. (8), it also follows

$$\Delta H_m \propto RTn_1\phi_2\left(\sqrt{\frac{\omega_{Z,1}-RT}{V_{m,1}}}-\sqrt{\frac{\omega_{Z,2}-RT}{V_{m,2}}}\right)^2 \quad (20)$$

which shows that comparable electrophilicity densities lead to vanishing and thus more favorable mixing enthalpies.

Higher Order Contributions and Perturbation Effects

In the previous subsection, we assumed the presence of a parabolic charge transfer reaction which results in the corresponding reaction energies (Eqn. (16)).[80,81] In principle, one can introduce higher order contributions in terms of cubic charge transfer models which often provide a higher accuracy.[80,81] A generic expression for the cubic reaction energy between two components A and B reads[81]

$$\Delta E_{AB} = (\chi_B-\chi_A)\Delta N_{AB}+\frac{1}{2}(\eta_A+\eta_B)\Delta N_{AB}^2+\frac{1}{6}(\Gamma_A+\Gamma_B)\Delta N_{AB}^3 \quad (21)$$

with the number of transferred electrons ΔN_{AB} and the hyperhardness parameter Γ. The last term in the previous equation can be regarded as the cubic contribution while the first two terms are key components of parabolic charge transfer models. In more detail, the hyperhardness reads

$$\Gamma_Z = \left(\frac{\partial^3 E_Z}{\partial N_Z^3}\right)_{\nu(\mathbf{r})} \quad (22)$$

which thus points to a modified expression for the reaction energy after detailed evaluation of all contributions.[81] After applying the corresponding expressions in combination with charge conservation,[81] one derives a modified expression for the HSP as denoted by 'SI Model 4' in the supporting information (Equation (4) in the SI).

In addition, one can also introduce certain perturbation parameters for the electronegativity and the chemical hardness.[35] Such perturbations can be rationalized by the presence of a continuum solvent and its pronounced influence on shifted electronic energies of the frontier molecular orbitals.[82] The resulting expressions for the perturbed electronegativity and the perturbed chemical hardness read

$$\chi_Z = \frac{\gamma_Z I_Z + A_Z}{1+\gamma_Z} \quad (23)$$

and

$$\eta_Z = \xi_Z(I_Z - A_Z) \quad (24)$$

with the perturbation parameters γ_Z and ξ_Z. It has to be noted that these parameters are usually regarded as fitting constants for numerical data with regard to the fact that detailed analytic expressions are hard to derive.[35,82] The individual and combined consideration of such perturbations can be applied for parabolic and cubic charge transfer approaches and thus results in modified expressions for the HSPs as presented in the Supporting Information (SI Models 1–3 and SI Models 6–11 (Eqns. (1)–(3) and (5)–(11) in the SI). In combination with cubic and parabolic charge transfer descriptions, the individual consideration of such extensions results in the fitting expressions SI Models 1–11 for the HSPs in the supporting information. We will fit the corresponding models to experimental and computed values of the HSPs and compare their accuracy to our previous simple expression (Eqn. (19)) in order to evaluate the influence of more advanced higher order contributions and perturbation approaches. A reasonable hypothesis would be that these more sophisticated approaches provide a more accurate description of dipolar and hydrogen bonding effects which thus enables a higher predictive accuracy for the HSPs and *vice versa* a broader application. Hence, the corresponding comparison with the simple approach allows us to understand if the validity of the HSPs is limited to aprotic and apolar solvents or if certain ignored higher order and perturbation approaches even improve the accuracy.

Numerical Results

Simple Conceptual DFT Contributions

The corresponding results for the experimental Hildebrand solubility parameters from References [20,83] in combination with HOMO and LUMO energies, molecular volumes V_m, computed dipole moments D as well as their hydrogen bonding properties for all considered solvents are shown in Table 1 and Table 2. Here we distinguish between solvents with a dipolar moment $D < 5.0$ Debye in combination with poor hydrogen bonding properties (aprotic solvents) in Table 1 and solvents with poor, moderate or strong hydrogen bonding behavior in combination with high dipole moments ($D > 5.0$ Debye) as shown in Table 2.

As can be seen, the considered solvents span a relatively broad range of HSPs from 7 $cal^{1/2}/cm^{3/2}$ for n-pentane to 16.30 $cal^{1/2}/cm^{3/2}$ for ethylene glycol. The corresponding dipole moments of all solvents in combination with their experimental HSPs are shown in Figure 1. One observes a weak positive correlation between high HSP values and high dipole moments. Such findings reveal that most aprotic and apolar solvents with dipole moments $D < 2$ Debye reveal low or moderate HSP values between δ_H =7 $cal^{1/2}\cdot cm^{3/2}$–10 $cal^{1/2}\cdot cm^{3/2}$. In contrast, more polar solvents ($D > 2$ Debye) with moderate or strong hydrogen bonding properties reveal HSPs between $\delta_H = 8 cal^{1/2}\cdot cm^{3/2}$–16 $cal^{1/2}\cdot cm^{3/2}$. Even more important, one can also notice in Table 2 for the protic solvents a significant shift of HSP values with increasing protic properties. Such findings highlight that reasonable threshold values for dipole moments as well as protic properties are needed where the HSP approach remains valid. Due to these reasons, we assume the validity of the HSP concepts for all solvents with $D < 5$ Debye in combination with poor protic properties. For solvents with moderate or strong hydrogen bond properties in

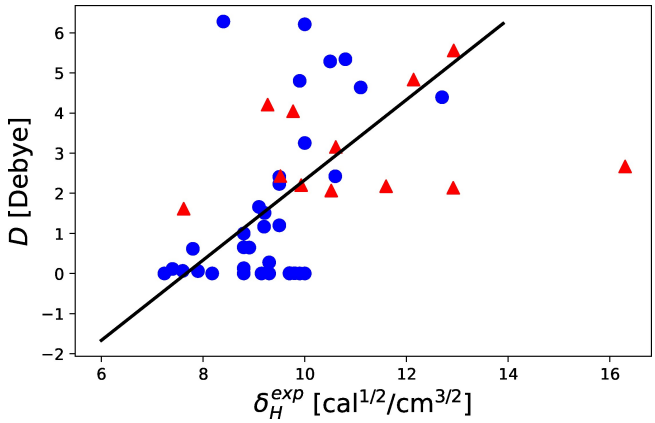

Figure 1. Experimental Hildebrand solubility parameters δ_H^{exp} in combination with dipole moments D from DFT calculations for the different solvents from Tables 1 and 2. Blue circles denote aprotic solvents while the red triangles denote solvents with moderate or strong hydrogen bonding. The black line denotes a least square linear regression fit for all aprotic solvents (blue circles).

combination with high dipole moments ($D > 5$ Debye), we expect certain deviations between our expression (Eqn. (18)) and the corresponding experimental values.

Further consideration of the chemical hardnesses and electronegativities of species also reveals a reasonable distribution of Lewis acidic and Lewis basic solvents[29,30] with hard and moderate chemical hardnesses. The corresponding results are depicted in Figure 2. Noteworthy, for aprotic and apolar solvents one can observe a gap between electronegativities $\chi = 2.5 - 3.5$ eV and chemical hardnesses $\eta = 3 - 6$ eV. However, for electronegativities $\chi > 3.5$ eV, a reasonable distribution of aprotic and apolar solvents can be observed. In addition, a significant linear ordering of chemical hardness values for $\eta > 6.0$ eV with increasing electronegativities has to be noticed. The corresponding protic and polar solvents include diethylether, 2-propanol, ethanol, ethylene glycol, tetrahydrofuran and morpholine in combination with the aprotic and apolar solvents cyclohexane, heptane, octane, methylcyclohexan, dodecane, decalin and hexane. Despite such ordered findings, one can conclude that the considered solvents reflect a broad range of hardly and softly polarizable as well as Lewis acidic or Lewis basic properties such that our calculations for the HSPs do not include a significant bias.

The corresponding calculated HSP values from Eqn. (18) in combination with the experimental values are shown in Figure 3. One observes a reasonable agreement for the aprotic and apolar solvents in terms of the calculated and the experimental values with a Pearson correlation coefficient of $R^2 = 0.72$. Specifically the aprotic and apolar solvents reveal a linear correlation. The resulting linear fit function in accordance with Eqn. (19) reads

$$\delta_H^{\text{exp}} = 0.969 \cdot \delta_H^{DFT} + 6.724\,\text{cal}^{1/2}/\text{cm}^{3/2} \tag{25}$$

with an average root mean squared error (RMSE) of $E(\delta_H^{exp}) = 0.585\,\text{cal}^{1/2}/\text{cm}^{3/2}$. The corresponding quality of the RMSE can be assessed in terms of the normalized RMSE according to $\text{nRMSE}_\sigma = E(\delta_H^{exp})/\sigma(\delta_H)$ and $\text{nRMSE}_{\Delta_H} = E(\delta_H^{exp})/\Delta_H$ where $\sigma(\delta_H)$ denotes the standard deviation of δ_H^{exp} and Δ_H the corresponding difference between the maximum and minimum values of δ_H^{exp}. The corresponding values read $\text{nRMSE}_\sigma = 0.532$ and $\text{nRMSE}_{\Delta_H} = 0.107$ which shows an acceptable accuracy of the fit function. In consequence, our calculated DFT values can be brought into reasonable agreement with the experimental values. Moreover, one can observe that even certain highly polar solvent molecules with $D > 5.0$ Debye show a reasonable agreement with the experimental HSP values. As expected from Figure 3, specifically the moderate and highly protic solvents reveal certain deviations from the experimental outcomes. Hence, it can be concluded that the presence of hydrogen bonds strongly limits the applicability of the HSP approach. In consequence, one can also conclude that Eqn. (18) remains

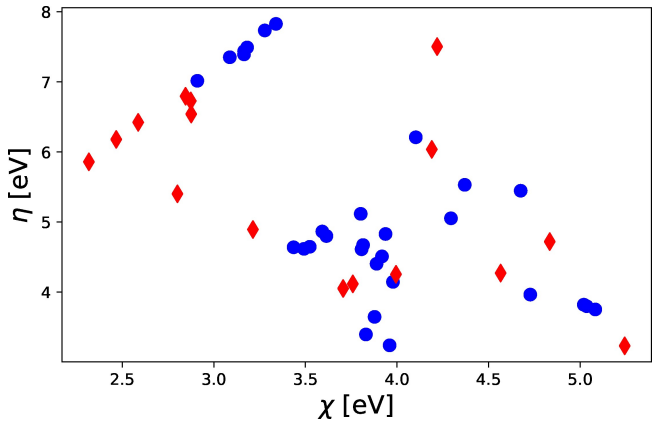

Figure 2. Electronegativities χ in combination with chemical hardnesses η from DFT calculations for the different solvents from Tables 1 and 2. Blue circles denote aprotic solvents with $D \leq 5$ Debye while the red diamonds denote solvents with moderate or strong hydrogen bonding and $D > 5$ Debye.

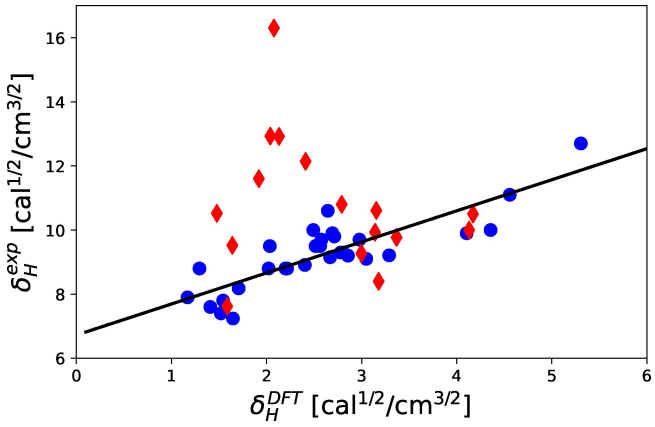

Figure 3. Calculated δ_H^{DFT} (Eqn. (18)) and experimental δ_H^{exp} Hildebrand solubility parameters. Blue circles denote aprotic solvents with dipole moments $D \leq 5$ Debye and red diamonds denote protic solvents with $D > 5$ Debye. The black line denotes a least square linear regression fit for all aprotic solvents (blue circles) with a correlation coefficient of $R^2 = 0.72$.

valid for aprotic and apolar compounds and provides reasonable estimates for HSPs values in terms of DFT calculations. Moreover, we have shown that the consideration of electrophilicity densities in Eqn. (18) results in a reasonable agreement with the experimental values. Hence, at least for the apolar and aprotic solvents, charge transfer effects in terms of binding energy effects contribute significantly to the vaporization energies. The corresponding results for the apolar and aprotic species in terms of gas phase calculations are shown in the supplementary material. As can be seen, there are only slight changes in the parameters of the fit function and the general accuracy, such that one can conclude, that our approach is valid for gas-phase and continuum solvent calculations.

Machine Learning

Furthermore, to assess the accuracy of the linear correlation (Eqn. (25)), we also developed an Extra Trees model to predict experimental HSP values. Using the Extra Trees model, we can now check whether we can achieve a higher overall accuracy by means of further features. Since we only considered a small data set, a k-fold cross validation approach is used, where the training data consists of each N-1 data samples with one test data point from the total data set including N samples.[84,85] In more detail, each Extra Trees model M_j is trained with the feature data including the samples $X = [x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_N]$ and the associated target data $Y = [y_1, y_2, \ldots, y_{j-1}, y_{j+1}, \ldots, y_N]$ for the experimental HSP values and predicts the corresponding test target sample Y_j from X_j. As can be seen, X_j and Y_j are not part of the corresponding training data for model M_j. This procedure, which is also called leave-one-out cross-validation (LOOCV) is repeated for all N models $M_1, M_2, \ldots, M_j$ and the corresponding predictions. The corresponding results are shown in Figure 4. As can be seen, one still observes certain deviations between the predicted and the experimental values. Such findings reveal the crucial limits of the HSP approach. In consequence, it can be concluded that the HSP approach is only valid for the limited class of apolar and aprotic solvents. Here, one can assume that the electrophilicity densities provide a reliable interpretation of the binding properties instead of the vaporization energies. A further evaluation of the HSP approach for protic polar solvents is not possible. The corresponding considered features are not sufficient and informative in order to reflect the corresponding hydorgen bonding and electrostatic contributions.

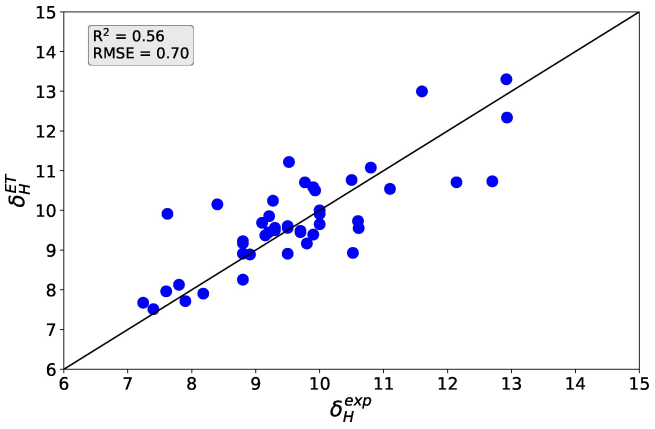

Figure 4. Experimental δ_H^{exp} and predicted δ_H^{ET} Hildebrand solubility parameters with Extra Trees Models in combination with a LOOCV approach. The black line denotes a least square linear regression fit with a correlation coefficient of $R^2 = 0.57$.

Perturbed and Cubic Conceptual DFT Contributions

In order to study the influence of higher order contributions such as cubic charge transfer models and the influence of perturbations for the electronegativities and chemical hardnesses, we fitted the data for the experimental and the computed HSPs with the corresponding models as listed in Eqns. (1)–(11) in the supporting information. The results are compared to the simple HSP fitting approach (Eqn. (19)) in order to evaluate the influence of further and higher order contributions. More details on the individual fitting equations can be found in the supporting information. In the following, we list the corresponding values as well as error estimates of the models for all solvents and for solvents with $D < 5$ Debye separately. Here, we do not differentiate between protic and aprotic solvents. The corresponding fitting results for all solvents are listed in Table 3.

The related plots are shown in the supporting information. As can be seen, the consideration of all solvents increases the error value for our simple fitting approach (Eqn. (19)) with $E(\delta_H) = 1.575\ \mathrm{cal}^{1/2}/\mathrm{cm}^{3/2}$ when compared to the results shown in Figure 3. Hence, it becomes clear that a separation into protic and low dipolar solvents is definitely needed for high fitting accuracies. Even the consideration of more advanced expressions (SI Models 1–11 in the supporting information) does not significantly reduce the error values from a general point of view. In comparison to the simple approach, it can be seen that the higher order and perturbation contributions slightly improve the error values but it becomes clear that an evaluation of HSPs for all solvents is not feasible. One can also assume, that the slight improvement over the simple approach can be related to the larger number of adaptable fitting parameters such that an improved matching to the data distribution can be achieved. However, the computed results for the Akaike information criterion (dAICc) show only minor differences between the various models with dAICc = 0.00–5.26 which shows that such effects can be ignored.

The corresponding values for all solvents with dipole moments $D < 5$ Debye are shown in Table 4. The related plots are shown in the supporting information. In comparison to the results shown in Table 3, one can observe a slight improvement in terms of lower error values. Even the simple approach shows a higher accuracy. However, the overall high error values indicate that mainly a distinction between protic and aprotic properties is necessary for accurate fitting procedures and robust HSP expressions. Accordingly, the prediction accuracy is not significantly increased when distinguishing between low

Table 3. Mean errors $E(\delta_H)$ between computed and experimental values of HSPs for all solvents in terms of the fitting function Eqn. (19) (Simple Approach) and the more advanced fitting functions with higher order and perturbation contributions as listed in the supplementary information (Eqns. (1)–(11) with the same nomenclature of SI – Model 1 – SI – Model 12). The corresponding values for the well-normalized hyperhardness parameters a, free hyperhardness parameters Γ, the slope p and δ_H^0 are also shown.

Model	$E(\delta_H)$ [cal$^{1/2}$/cm$^{3/2}$]	Γ [cal^3]	a [cal^2]	p	δ_H^0 [cal$^{1/2}$/cm$^{3/2}$]
Simple Approach (Eqn. (19))	1.575	–	–	0.288	8.353
SI – Model 1	1.054	–	–	0.158	8.510
SI – Model 2	1.078	–	–	0.154	8.902
SI – Model 3	1.050	–	–	0.110	7.787
SI – Model 4	1.055	1.000	–	0.109	8.353
SI – Model 5	1.074	–	−407400	0.183	7.750
SI – Model 6	1.054	1.000	–	0.158	8.510
SI – Model 7	1.061	–	−3109000	0.218	8.191
SI – Model 8	1.078	1.000	–	0.154	8.902
SI – Model 9	1.075	–	−2342000	0.043	8.949
SI – Model 10	1.050	1.000	–	0.110	7.787
SI – Model 11	1.067	–	−296100	0.172	7.117

Table 4. Mean errors $E(\delta_H)$ between computed and experimental values of HSPs for all solvents with $D<5$ Debye in terms of the fitting function Eqn. (19) (Simple Approach) and the more advanced fitting functions with higher order and perturbation contributions as listed in the supplementary information (Eqns. (1)–(11) with the same nomenclature of SI – Model 1 – SI – Model 12). The corresponding values for the well-normalized hyperhardness parameters a, free hyperhardness parameters Γ, the slope p and δ_H^0 are also shown.

Model	$E(\delta_H)$ [cal$^{1/2}$/cm$^{3/2}$]	Γ [cal^3]	a [cal^2]	p	δ_H^0 [cal$^{1/2}$/cm$^{3/2}$]
Simple Approach (Eqn. (19))	1.541	–	–	0.349	8.015
SI – Model 1	0.988	–	–	0.194	8.175
SI – Model 2	1.006	–	–	0.197	8.602
SI – Model 3	0.998	–	–	0.129	7.419
SI – Model 4	0.991	1.000	–	0.132	8.015
SI – Model 5	1.027	–	−381300	0.202	7.491
SI – Model 6	0.988	1.000	–	0.194	8.175
SI – Model 7	0.999	–	−2370000	0.236	7.967
SI – Model 8	1.006	1.000	–	0.197	8.602
SI – Model 9	1.000	–	−4414000	0.042	8.811
SI – Model 10	0.998	1.000	–	0.129	7.419
SI – Model 11	1.036	–	−277100	0.190	6.774

and high dipolar solvents. This finding applies to all approaches.

From this it can be concluded that the key contributions for aprotic and apolar solvents are already contained in Eqn (19). The introduction of more advanced approaches comes at the price of a higher number of free fitting parameters combined with a more complex expression. We conclude that using the simple approach is sufficient for properly chosen solvents, as reflected in aprotic and apolar properties. A further application of more advanced expressions for aprotic and apolar solvents does not increase the accuracy. Thus, one can conclude that the HSP approach is not applicable for protic and polar solvents in combination with the corresponding chosen molecular parameters.

Summary and Conclusions

We introduced a new expression for the Hildebrand solubility parameters by means of conceptual DFT calculations. Our results reveal that the cohesive energy density of the HSPs can be replaced by an electrophilicity density. The corresponding expressions are compared to experimental HSP values for a broad range of solvents. Our results reveal a good agreement between the calculated and the experimental HSP values. Slighter deviations can only be observed for solvents with moderate or strong hydrogen bonds and high dipole moments $D > 5$ Debye. Noteworthy, these deviations were expected as the original HSP approach is restricted to aprotic and apolar species. In addition, our results reveal that higher order and perturbation contributions only slightly improve the accuracy of predictions. Comparable conclusions can also be drawn for the

consideration of machine learning approaches. As already discussed, such approaches come with the cost of more free fitting parameters in combination with complex expressions. Hence, even more sophisticated approaches are not able to accurately describe hydrogen bonding or electrostatic effects.

Although we introduced a useful linear mapping relation between our computed HSP and the experimental values for practical purposes, the implications of this approach are more far reaching. In fact, the experimentally determined heat of vaporizations as cohesive energy densities can be replaced by electrophilicity densities or stabilization energy densities. Such expressions, which are more easily accessible than experimental values clearly reveal the presence of charge transfer effects as key contributions. In recent works, we have already shown that charge transfer effects are omnipresent in solution and provide the vast majority of contributions to solvation effects. The corresponding expression of the mixing enthalpy (Eqn. 20) points into the same direction. As was shown, solvents are hardly miscible if they have a large difference in the electrophilicity densities. Although the limits of Eqn. (20) in combination with regular solution theory were already discussed, it has to be noted that the underlying physical picture provides some useful insights into the fundamental mechanisms. Hence, one can conclude that charge transfer effects in terms of electrophilicity densities also dominate the dissolution processes for aprotic and apolar molecular species. Closely related, the electrophilicity accounts for potential electron transfer processes, such that it can be expected that less miscible solvents show certain deficits in full compensation of electron transfer values. In summary, our findings show that even apolar solvents and the corresponding solubilities are crucially affected by charge transfer effects and the corresponding molecular properties.

Supporting Information

The results for the different models and the outcomes of the gas phase calculations are presented in the Supporting Infomation. Additional references listed only in the Supporting Information include Refs. [80–82].

Acknowledgements

The authors thank Moritz Schulze, Niklas Adebar, Robert Möckel and Julian Keupp for valuable discussions. R.A.M.Q acknowledges funding from the Univesity of Florida in terms of a start-up grant. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interests

J. S. is also an employee of Boehringer Ingelheim Pharma GmbH & Co. KG.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: Conceptual DFT · Electrophilicity · Hildebrand Solubility Parameters · Mixing Enthalpy · Solvent Mixing

[1] C. Reichardt, T. Welton, *Solvents and Solvent Effects in Organic Chemistry*, John Wiley & Sons **2011**.
[2] E. Buncel, R. A. Stairs, *Solvent Effects in Chemistry*, John Wiley & Sons **2015**.
[3] A. Veseli, S. Žakelj, A. Kristl, *Drug Dev. Ind. Pharm.* **2019**, *45*, 1717.
[4] C. L. Silveira, A. C. Galvao, W. S. Robazza, J. V. T. Feyh, *Fluid Phase Equilib.* **2021**, *535*, 112970.
[5] C. H. Mehta, R. Narayan, U. Y. Nayak, *Drug Discovery Today* **2019**, *24*, 781.
[6] S. Hossain, A. Kabedev, A. Parrow, C. A. Bergström, P. Larsson, *Europ. J. Pharm. Biopharm.* **2019**, *137*, 46.
[7] S. Boobier, D. R. Hose, A. J. Blacker, B. N. Nguyen, *Nat. Commun.* **2020**, *11*, 5753.
[8] M. A. Lovette, J. Albrecht, R. S. Ananthula, F. Ricci, R. Sangodkar, M. S. Shah, S. Tomasi, *Cryst. Growth Des.* **2022**, *22*, 5239.
[9] R. Neff, D. McQuarrie, *J. Phys. Chem.* **1973**, *77*, 413.
[10] P. E. Smith, R. M. Mazo, *J. Phys. Chem. B* **2008**, *112*, 7875.
[11] J. Smiatek, A. Heuer, M. Winter, *Batteries* **2018**, *4*, 62.
[12] F. Tumakaka, J. Gross, G. Sadowski, *Fluid Phase Equilib.* **2005**, *228*, 89.
[13] S. Shimizu, N. Matubayasi, *J. Mol. Liq.* **2019**, *273*, 626.
[14] S. Gracin, T. Brinck, Å. C. Rasmuson, *Ind. Eng. Chem. J.* **2002**, *41*, 5114.
[15] R. Schmid, *Monatsh. Chem.* **2001**, *132*, 1295.
[16] B. Zhuang, G. Ramanauskaite, Z. Y. Koa, Z.-G. Wang, *Sci. Adv.* **2021**, *7*, eabe7275.
[17] J. Hildebrand, R. Scott, *Annu. Rev. Phys. Chem.* **1950**, *1*, 75.
[18] J. H. Hildebrand, R. L. Scott, *Regular solutions*, Prentice-Hall **1962**.
[19] J. H. Hildebrand, *Solubility of Non-electrolytes*, Reinhold Pub. **1936**.
[20] J. Burke, *Solubility parameters: theory and application*, The Book and Paper Group of the American Institute for Conservation **1984**.
[21] C. M. Hansen, *Hansen solubility parameters: a user's handbook*, CRC press **2007**.
[22] B. C. Hancock, P. York, R. C. Rowe, *Int. J. Pharm.* **1997**, *148*, 1.
[23] S. Venkatram, C. Kim, A. Chandrasekaran, R. Ramprasad, *J. Chem. Inf. Model.* **2019**, *59*, 4188.
[24] N. E. Jackson, L. X. Chen, M. A. Ratner, *J. Phys. Chem. B* **2014**, *118*, 5194.
[25] H. S. Salehi, M. Ramdin, O. A. Moultos, T. J. Vlugt, *Fluid Phase Equilib.* **2019**, *497*, 10.
[26] M. Belmares, M. Blanco, W. Goddard Iii, R. Ross, G. Caldwell, S.-H. Chou, J. Pham, P. Olofson, C. Thomas, *J. Comput. Chem.* **2004**, *25*, 1814.
[27] L. Utracki, R. Simha, *Polym. Int.* **2004**, *53*, 279.
[28] N. Rai, A. J. Wagner, R. B. Ross, J. I. Siepmann, *J. Chem. Theory Comput.* **2008**, *4*, 136.
[29] J. Smiatek, *J. Chem. Phys.* **2019**, *150*, 174112.
[30] J. Smiatek, *J. Phys. Chem. B* **2020**, *124*, 2191.
[31] R. A. Miranda-Quintana, J. Smiatek, *ChemPhysChem* **2020**, *21*, 2605.
[32] R. A. Miranda-Quintana, J. Smiatek, *Electrochim. Acta* **2021**, *384*, 138418.
[33] R. A. Miranda-Quintana, J. Smiatek, *J. Mol. Liq.* **2021**, *322*, 114506.
[34] R. A. Miranda-Quintana, J. Smiatek, *Phys. Chem. Chem. Phys.* **2022**, *24*, 22477.
[35] R. A. Miranda-Quintana, L. Chen, V. S. J. Craig, J. Smiatek, *J. Phys. Chem. B* **2023**, *127*, 2546.
[36] P. Geerlings, F. De Proft, W. Langenaeker, *Chem. Rev.* **2003**, *103*, 1793.
[37] H. Chermette, *J. Comput. Chem.* **1999**, *20*, 129.
[38] P. K. Chattaraj, S. Utpal, D. R. Roy, *Chem. Rev.* **2006**, *106*, 2065.
[39] J. L. Gazquez, A. Cedillo, A. Vela, *J. Phys. Chem. A* **2007**, *111*, 1966.
[40] J. L. Gázquez, *J. Mex. Chem. Soc.* **2008**, *52*, 3.
[41] P. Geerlings, E. Chamorro, P. K. Chattaraj, F. De Proft, J. L. Gázquez, S. Liu, C. Morell, A. Toro-Labbé, A. Vela, P. Ayers, *Theor. Chem. Acc.* **2020**, *139*, 1.
[42] P. K. Chattaraj, S. Giri, *Ann. Rep. Phys. Chem. C* **2009**, *105*, 13.
[43] S.-B. Liu, *Acta Phys. Chim. Sin.* **2009**, *25*, 590.
[44] R. A. Miranda-Quintana, M. Martinez Gonzalez, P. W. Ayers, *Phys. Chem. Chem. Phys.* **2016**, *18*, 22235.

[45] R. A. Miranda-Quintana, P. W. Ayers, *Phys. Chem. Chem. Phys.* **2016**, *18*, 15070.
[46] R. A. Miranda-Quintana, *J. Chem. Phys.* **2017**, *146*, 046101.
[47] R. A. Miranda-Quintana, *J. Chem. Phys.* **2017**, *146*, 214113.
[48] N. Islam, S. Kaya, *Conceptual density functional theory and its application in the chemical domain*, CRC press **2018**.
[49] S. Liu, *Conceptual density functional theory: Towards a new chemical reactivity theory*, John Wiley & Sons **2022**.
[50] P.-G. De Gennes, P.-G. Gennes, *Scaling concepts in polymer physics*, Cornell University Press **1979**.
[51] M. Doi, S. F. Edwards, S. F. Edwards, *The theory of polymer dynamics*, Oxford University Press **1988**.
[52] J. H. Hildebrand, *Chem. Rev.* **1936**, *18*, 315.
[53] T. Lindvig, M. L. Michelsen, G. M. Kontogeorgis, *Fluid Phase Equilib.* **2002**, *203*, 247.
[54] G. Scatchard, J. Hildebrand, *J. Am. Chem. Soc.* **1934**, *56*, 995.
[55] P. W. Ayers, *J. Chem. Phys.* **2005**, *122*, 141102.
[56] P. W. Ayers, *Faraday Discuss.* **2007**, *135*, 161.
[57] R. G. Parr, R. G. Pearson, *J. Am. Chem. Soc.* **1983**, *105*, 7512.
[58] T. Koopmans, *Physica* **1934**, *1*, 104.
[59] J. F. Janak, *Phys. Rev. B* **1978**, *18*, 7165.
[60] R. G. Parr, L. v Szentpály, S. Liu, *J. Am. Chem. Soc.* **1999**, *121*, 1922.
[61] F. Neese, *WIREs Comput. Mol. Sci.* **2012**, *2*, 73.
[62] F. Neese, *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.
[63] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
[64] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
[65] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
[66] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[67] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456–1465.
[68] V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995.
[69] E. Silla, I. Tunon, J. L. Pascual-Ahuir, *J. Comput. Chem.* **1991**, *12*, 1077.
[70] J.-L. Pascual-ahuir, E. Silla, I. Tunon, *J. Comput. Chem.* **1994**, *15*, 1127.
[71] L. Kronik, T. Stein, S. Refaely-Abramson, R. Baer, *J. Chem. Theory Comput.* **2012**, *8*, 1515.
[72] R. A. Miranda-Quintana, P. W. Ayers, *Theor. Chem. Acc.* **2016**, *135*, 172.
[73] R. A. Miranda-Quintana, P. W. Ayers, *Theor. Chem. Acc.* **2016**, *135*, 1.
[74] P. Geurts, D. Ernst, L. Wehenkel, *Mach. Learn.* **2006**, *63*, 3.
[75] I.-T. Ho, M. Matysik, L. M. Herrera, J. Yang, R. J. Guderlei, M. Laussegger, B. Schrantz, R. Hammer, R. A. Miranda-Quintana, J. Smiatek, *Phys. Chem. Chem. Phys.* **2022**, *24*, 28314.
[76] G. Van Rossum, F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA **2009**.
[77] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, *Nature* **2020**, *585*, 357.
[78] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
[79] Wes McKinney, Data Structures for Statistical Computing in Python, in Stéfan van der Walt, Jarrod Millman (Editors), *Proceedings of the 9th Python in Science Conference* **2010** pages 56–61.
[80] R. A. Miranda-Quintana, P. W. Ayers, F. Heidar-Zadeh, *ChemistrySelect* **2021**, *6*, 96.
[81] R. A. Miranda-Quintana, P. W. Ayers, F. Heidar-Zadeh, *Theor. Chem. Acc.* **2021**, *140*, 1.
[82] R. A. Miranda-Quintana, *Theor. Chem. Acc.* **2017**, *136*, 76.
[83] A. F. Barton, *Chem. Rev.* **1975**, *75*, 731.
[84] J. Gareth, W. Daniela, H. Trevor, T. Robert, *An introduction to statistical learning: with applications in R*, Spinger **2013**.
[85] T.-T. Wong, *Pattern Recogn.* **2015**, *48*, 2839.

Manuscript received: August 10, 2023
Revised manuscript received: October 23, 2023
Accepted manuscript online: October 26, 2023
Version of record online: November 14, 2023