On a relationship between molecular polarizability and partial molar volume in water

ARTICLE *in* THE JOURNAL OF CHEMICAL PHYSICS · DECEMBER 2011

Impact Factor: 2.95 · DOI: 10.1063/1.3672094 · Source: PubMed

2 AUTHORS:

Ekaterina Ratkova
Helmholtz Zentrum München
19 PUBLICATIONS **193** CITATIONS
SEE PROFILE

Maxim V Fedorov
University of Strathclyde
98 PUBLICATIONS **2,073** CITATIONS
SEE PROFILE

On a relationship between molecular polarizability and partial molar volume in water

Ekaterina L. Ratkova and Maxim V. Fedorov

Citation: J. Chem. Phys. **135**, 244109 (2011); doi: 10.1063/1.3672094
View online: http://dx.doi.org/10.1063/1.3672094
View Table of Contents: http://jcp.aip.org/resource/1/JCPSA6/v135/i24
Published by the American Institute of Physics.

On a relationship between molecular polarizability and partial molar volume in water

Ekaterina L. Ratkova[1,a)] and Maxim V. Fedorov[1,2,b)]
[1]*The Max Planck Institute for Mathematics in Sciences, Inselstrasse 22, Leipzig 04103, Germany*
[2]*Department of Physics, Scottish Universities Physics Alliance (SUPA), Strathclyde University, John Anderson Building, 107 Rottenrow East, Glasgow G4 0NG, United Kingdom*

(Received 29 July 2011; accepted 5 December 2011; published online 30 December 2011)

We reveal a universal relationship between molecular polarizability (a *single-molecule* property) and partial molar volume in water that is an *ensemble* property characterizing solute-solvent systems. Since both of these quantities are of the key importance to describe solvation behavior of dissolved molecular species in aqueous solutions, the obtained relationship should have a high impact in chemistry, pharmaceutical, and life sciences as well as in environments. We demonstrated that the obtained relationship between the partial molar volume in water and the molecular polarizability has in general a non-homogeneous character. We performed a detailed analysis of this relationship on a set of ~200 organic molecules from various chemical classes and revealed its fine well-organized structure. We found that this structure strongly depends on the chemical nature of the solutes and can be rationalized in terms of specific solute-solvent interactions. Efficiency and universality of the proposed approach was demonstrated on an external test set containing several dozens of polyfunctional and druglike molecules. © *2011 American Institute of Physics*. [doi:10.1063/1.3672094]

Determination of physical/chemical properties of a solute-solvent system using only molecular parameters of its components derived from the chemical structure (quantitative structure-property relationship, QSPR) (Ref. 1) is a common approach in chemistry, pharmaceutical sciences, environmental protection policy, and health research. Partial molar volume (PMV) is of the key importance to describe solvation behavior of molecular compounds in aqueous solutions since it contains information both about the immersed solute structure as well as on specific solute-solvent interactions.[2] From another side, redistribution of charges in a solute molecule upon solvation can be characterized by the molecular polarizability (α). Therefore, polarizability effects should play an essential role in molecular solvation, particularly in polar solvents such as water. In this article we report a novel way how to relate these two key properties (PMV and α) and how to determine PMV of an organic molecule in water in terms of the static electric polarizability of the molecule.

It has long been known that the molecular polarizability correlates with *molecular* volume (V_m).[3] For a conductive sphere, for example, the polarizability is given analytically as $\alpha = 3V_m/4\pi$. The linear relation $\alpha = b_1 V_m + b_0$ (where b_1 is the scaling coefficient, b_0 is the intercept) gives a high correlation coefficient between experimental and predicted data for both spherical particles ($R = 0.988$) (Ref. 4) and simple molecular compounds ($R = 0.960$).[3,5] In turn, molecular volume also correlates with the PMV. Lepori and Gianni demonstrated PMV – V_m correlations which have the same scaling coefficient and different intercept according to the different functional groups.[6] It was shown that PMVs of non-polar solutes (e.g., hydrocarbons, noble gases, etc.) can be obtained by the linear relation: $\bar{V}^{np} = c_1 V_m + c_0^{np}$ (where c_1 is a universal scaling coefficient, c_0^{np} is the intercept for non-polar compounds). However, accurate prediction of the class-specific changes of the intercept for polar molecular compounds still remains an open question. The goals of the current study were (i) to reveal the basic relationships between PMV and the polarizability of the solute and (ii) to develop *a universal* model describing their relationship.

We performed our investigation on a set of 210 neutral organic compounds taken from Refs. 7 and 8. This set comprises molecules that belong to multiple different chemical classes, from simple hydrocarbons to druglike compounds. We note that the dataset contains homologous series of monofunctional organic molecules as well as congeners of polyfunctional molecules. The congeners are of two types: (i) with the same hydrocarbon chain but different number of identical functional groups or (ii) with different hydrocarbon chains and different number of identical functional groups. In the following analysis, we used a training set of 74 molecules and a test set of 136 molecules. We divided the test set into an internal test set of 93 solutes (taken from the same chemical classes as the molecules in the training set), and an external test set containing 19 druglike compounds, and 24 other polyfunctional molecules.

Analysis of the available literature data reveals that, in general, there is a lack of experimental data on polarizability and PMV for organic compounds. Due to that, we used robust and accurate computational methods to obtain PMV and static electric polarizability for those compounds for which we were not able to find the experimental data (more details are provided below). To check the accuracy of the employed

a)Electronic mail: ratkova@mis.mpg.de.
b)Electronic mail: maxim.fedorov@strath.ac.uk. Tel: +44 141 5484 012. Fax: +44 141 5522 891.

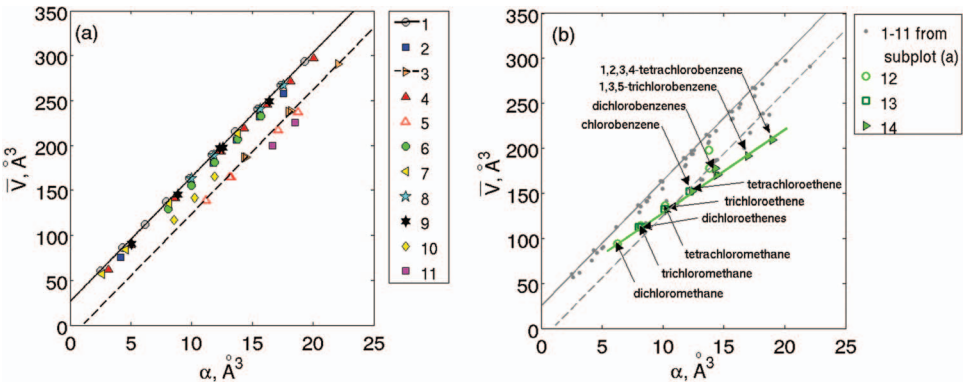

FIG. 1. Correlations between the benchmark partial molar volume $(\bar{V})$ and the benchmark static electric polarizability (α) for the training set of solutes: (a) homologous series of monofunctional solutes: 1 – alkanes, 2 – alkenes, 3 – alkylbenzenes, 4 – monohydric alcohols, 5 – phenols, 6 – 1-chloroalkanes, 7 – aldehydes, 8 – ketones, 9 – ethers, 10 – 1,3-dienes, and 11 – styrenes; the solid black line gives the correlation for the homologous series of alkanes (No. 1), while the dashed black line illustrates the correlation for the homologous series of alkylbenzenes (No. 3); (b) congeners of polyfunctional compounds: 12 – polychlorinated alkanes, 13 – polychlorinated alkenes, 14 – polychlorinated benzenes; grey dots represent the solutes shown on the subplot (a); grey solid and dashed lines reproduce the corresponding lines from subplot (a); the green solid line illustrates the correlation for polychlorinated compounds with the same hydrocarbon chain but different numbers of chlorine atoms (see the notation of the subplot).

methods we selected a "verification" set from molecules for which experimental values of the PMV and/or polarizability are available.[9]

Previously, it was shown that a combination of the Kirkwood-Buff (KB) solution theory with the reference interaction site model (RISM) of the integral equation theory of molecular liquids has succeeded in PMV estimations for various molecular systems: ions,[10] hydrocarbons,[11–13] and biomolecules.[14,15] In this work, PMV values were calculated as follows. Direct correlation functions of the solute–solvent systems were obtained by 3D RISM method combined with the Kovalenko-Hirata closure (3D RISM-KH).[16] We refer to Ref. 9 for the computational details. The resulted correlation functions were used to calculate the PMV values by the following KB equation:

$$\bar{V} = k_B T \chi_T \left(1 - \rho \sum_{s=1}^{N} \int_{R^3} c_s(\mathbf{r}) d\mathbf{r}\right), \quad (1)$$

where $\bar{V}$ is the partial molar volume, k_B is the Boltzmann constant, T is the temperature, χ_T is the pure solvent isothermal compressibility, ρ is the bulk density of the solvent, $c_s(\mathbf{r})$ is the intermolecular solute-solvent site direct correlation function (s denote the index of the corresponding solvent molecule site), N is the overall number of the solvent sites in one solvent molecule, and $\mathbf{r}$ is the radius-vector of the solvent site. We note that previously this formula was referred as the 3D RISM-KH/KB theory.[2] For the "verification" set of molecules we found a strong linear correlation between the calculated and experimental[17–21] PMVs with the correlation coefficient of R = 0.987 and standard deviation of error of 6.61 Å^3 (≈4% of the corresponding average value of PMV).

To obtain the static electric polarizability values, we performed gas-phase calculations at the B3LYP/aug-cc-pVDZ level of theory with the GAUSSIAN 03 software.[22] According to the analysis carried out by the National Institute of Standards and Technology (NIST),[23] this level of theory provides reliable calculations of the static electric polarizability. Indeed, for the "verification" set of molecules we obtained a good agreement between the calculated static electric polarizabilities and the corresponding experimental data[23,24] (R = 0.997, standard deviation of error equals 0.33 Å^3 (≈3% of the corresponding average value of polarizability)).

The fact that for the selected set of molecules the calculated values of PMV and static electric polarizability are so well correlated with the corresponding experimental values allows us to use them in our study as accurate estimations of unavailable experimental data. For the rest of this article we refer to the values obtained by the accurate computational methods as the "benchmark" data.

Figure 1 shows that for the whole training set the PMV – α correlation has a non-homogeneous character. We found that separation of the training set to subsets of mono- and polyfunctional solutes simplifies the analysis of the correlations (Figure 1). Thus, for non-polar solutes (e.g., alkanes) and monofunctional homologues (i.e., compounds with the same functional group but different length of the hydrocarbon tail) there are strong linear correlations between the benchmark PMV and polarizability (Figure 1(a)). The linear dependencies have the same *universal* scaling coefficient A (which reflects the nature of the saturated hydrocarbon tail) but differ by the class–specific intercept B^{class} that reflects the nature of unsaturated bonds, polar, and charged groups (fine structure of the PMV – α relationship):

$$\bar{V}^{class} = A\alpha + B^{class}. \quad (2)$$

In the case of polyfunctional solutes, the PMV correlation with α has another scaling coefficient which, in this case, reflects the nature of the functional groups. Thus, polychlorinated compounds from several classes (within one class the compounds have the same hydrocarbon chain but different number of chlorine atoms) can be described by the same correlation (see Figure 1(b), green solid line). However, several solutes (1,4-dichloropentan, 2-chlorotoluene, and pentachloroethane) are placed above this line. These molecules have larger hydrocarbon chain than that of the molecules on the line (Figure 1(b)) and, as a result, larger value of PMV. This result allows us to assume that for polyfunctional solutes

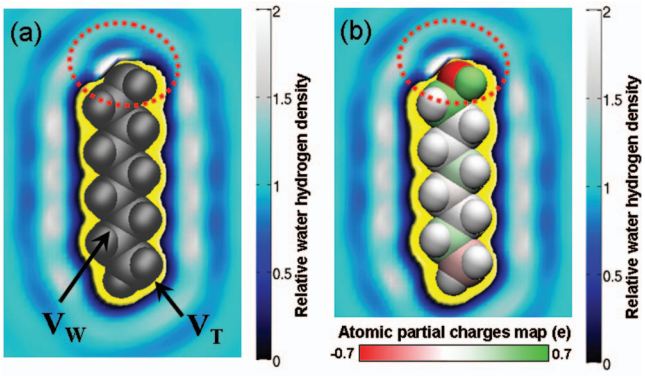

FIG. 2. Water hydrogen density distributions around the solute (1–octanol): (a) The solute molecule is represented by the van der Waals volume, V_W (grey balls), the yellow area corresponds to the zero water hydrogen density (so-called thermal volume, V_T), the dashed red circle shows the area with the minimal thermal volume; (b) The van der Waals volume of the solute is mapped with the atomic partial charges,[26,27] the dashed red curve shows the most negatively charged site of the solute (oxygen atom). Due to the negative partial charge, this site strongly attracts water hydrogens that resulted in minimal thermal volume around it.

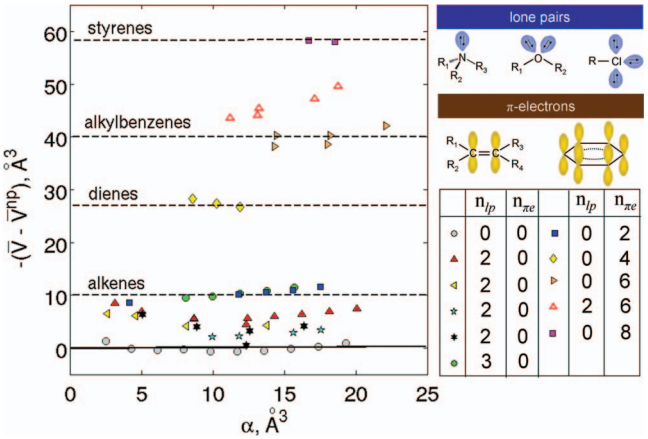

FIG. 3. Difference between benchmark PMV and hypothetical "non-polar" PMV ($\bar{V} - \bar{V}^{np}$) calculated with Eq. (4) versus the benchmark static electric polarizability (α) for homologous series of solutes from the training set (see the symbol notations on the caption of Figure 1). The solid black line gives the difference for alkanes. The dotted black lines give the difference for unsaturated compounds without lone pairs, $n_{lp} = 0$ (alkenes, dienes, alkylbenzenes, styrenes). As one can see, the difference ($\bar{V} - \bar{V}^{np}$) changes with the change in the number of π-electrons ($n_{\pi e}$) as well as with the change in the number of lone pairs (n_{lp}) (see the table in right part).

the PMV – α correlation also has a fine structure which, in this case, reflects the size of the hydrocarbon chain.

To understand and rationalize this intricate behavior of the PMV – α relationship we performed an additional analysis representing the PMV as a complex quantity that depends on polarizability. According to the previous works,[2,6,25] the PMV of a solute at infinite dilution can be decomposed into the following terms:

$$\bar{V}(\alpha) = \underbrace{V_W(\alpha) + V_T(\alpha)}_{V_C(\alpha)} + V_I(\alpha) + k_B T \chi_T, \qquad (3)$$

where V_C ("cavity volume") is the volume of the cavity created in the solvent to accommodate the solute molecule; it contains two contributions (Figure 2(a)): (i) the solute molecular volume (for low molecular weight solutes it can be approximated by the van der Waals volume, V_W) and (ii) the thermal (empty, void) volume, V_T, associated with thermally induced molecular vibrations of both the solute and solvent molecules which lead to the creation of an empty (void) space around the solute molecule;[2,25] V_I ("interaction volume") represents the change in the solvent volume associated with specific interactions of solvent molecules with charged and polar groups of the solute; the last term, $k_B T \chi_T$, describes the volume effect related to the kinetic contribution of a solute molecule to the pressure due to translation degrees of freedom, it is a property of the solvent only and does not depend on the solute properties (e.g., static electric polarizability).

In the case of non-polar saturated solutes, there are no specific solute-solvent interactions (interaction volume equals zero) and the PMV corresponds to the solute cavity volume (value of the ideal term is small and, therefore, usually can be ignored).[25] Thus, for alkanes the relationship between PMV and polarizability reduces to the V_C – α correlation which is similar to the correlation between molecular volume and polarizability observed previously in Refs. 3 and 5. This correlation has the scaling coefficient $A = 13.85$ and the intercept $B^{np} = 26.84$ Å^3 (Figure 1(a), black solid line). The high correlation coefficient ($R = 0.999$) between predicted and benchmark data for the test subset of alkanes and the small standard deviation of error ($std = 1.24$ Å^3 which is about 1% of the corresponding average value of PMV) allows us to conclude that this correlation can be applied to the whole chemical subspace of solutes consisting of open-loop alkyl chain with different degrees of branching.

In turn, solutes with unsaturated bonds, polar groups, and charges have specific interactions with solvent molecules which leads to a decrease of solute PMV. Thus, PMVs of alkylbenzenesare are smaller than that of alkanes with the same polarizability (see Figure 1(a), black solid and dashed lines). One may note that the decrease is caused not only by the change of the solvent volume (interaction volume) but also by the decrease of the thermal volume (see Figure 2(b)):

$$\bar{V}(\alpha) - \bar{V}^{np}(\alpha) = \left[V_T(\alpha) - V_T^{np}(\alpha)\right] + V_I(\alpha), \qquad (4)$$

where $\bar{V}^{np}(\alpha)$ is the hypothetical "non-polar" PMV, calculated with Eq. (2) using coefficients A and B^{np} (we note that van der Waals volume of a solute does not depend on the solute polarizability: $V_W(\alpha) = \text{const}$). For homologous series of monofunctional solutes the PMVs differences (Eq. (4)) are shown on Figure 3. As one can see, they strongly correlate with the shifts of PMV – α correlations depicted on Figure 1(a). This result shows that for monofunctional homologous the fine structure of the PMV – α correlation is determined by the specific solute–solvent interactions. We assume that analysis of the PMVs difference may be of key importance for understanding the PMV – α correlations for polyfunctional solutes (see Figure 1(b)).

For further analysis we apply the fact that polarizability is not a uniform quantity throughout a solute molecule. Thus, groups with π-electrons and lone pairs are expected to be the most polarizable.[4] Therefore, the difference of PMVs (as a function of polarizability) should depend on the number

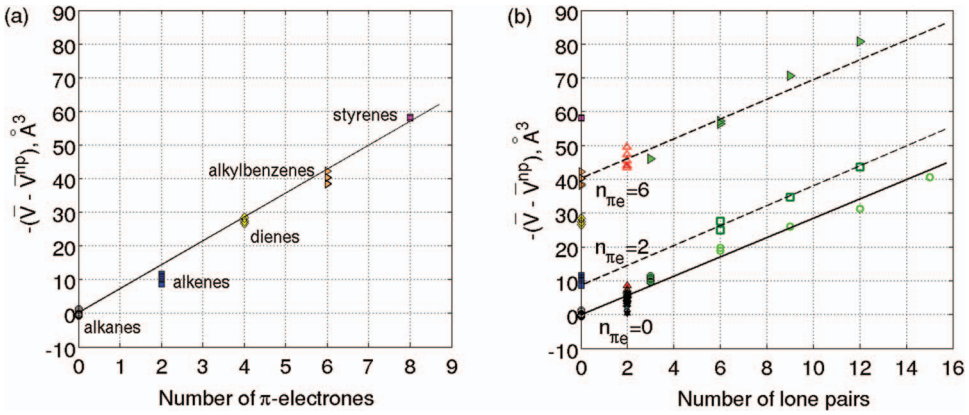

FIG. 4. (a) The difference between the benchmark PMV and hypothetical "non-polar" PMV ($\bar{V} - \bar{V}^{np}$) calculated with Eq. (4) versus the number of π-electrons for the training set of alkanes and unsaturated hydrocarbons without lone pairs. Black solid lines give the line-of-best-fit. (b) ($\bar{V} - \bar{V}^{np}$) versus the number of lone pairs for the training set of solutes (see the symbol notations on the caption of Figure 1). The black solid line gives the line-of-best-fit for saturated compounds, while the black dashed lines illustrate the corresponding correlations for unsaturated molecules.

of lone pairs (n_{lp}) and the number of π-electrons ($n_{\pi e}$) (see Figure 3, inset table). Indeed, we found strong linear correlations between the difference of PMVs and the number of π-electrons (Figure 4(a)) as well as the number of lone pairs (Figure 4(b)). That is resulted in the following equation:

$$-(\bar{V} - \bar{V}^{np}) = C_1 n_{lp} + C_2 n_{\pi e}, \quad (5)$$

where $C_1 = 2.86$ Å^3 and $C_2 = 7.12$ Å^3 are the corresponding scaling coefficients. We found that this relation has a universal character and can be reasonably well applied to mono- and for polyfunctional solutes. Substituting Eq. (2) for non-polar solutes into Eq. (5) yields

$$\bar{V} = \underbrace{A\alpha + B^{np}}_{\bar{V}^{np}} - (C_1 n_{lp} + C_2 n_{\pi e}). \quad (6)$$

The predictive ability of the Eq. (6) was analyzed using the internal test set of solutes and the same coefficients A, B^{np}, C_1, and C_2 as for the training set. Comparison of the predicted and benchmark PMVs is shown in Figure 5(a). The figure shows a strong linear correlation between these two sets of data with the correlation coefficient of R = 0.998 and the standard deviation of error of 2.58 Å^3 (≈2% of the corresponding average value of PMV).

To demonstrate the transferability of the coefficients for the proposed model we apply it to the external test set containing 19 neutral druglike molecules and 24 other polyfunctional solutes. Majority of the pharmaceutical molecules are non-steroidal anti-inflammatory drug with analgesics properties: aspirin, paracetamol (4–acetaminophen), 3–acetaminophen, 2–acetaminophen, ibuprofen, phenacetin, acetanilide, salicylic acid (2–hydroxybenzoic acid), 3–hydroxybenzoic acid, 4–hydroxybenzoic acid, 2–methoxybenzoic acid, 3–methoxybenzoic acid, and 4–methoxybenzoic acid. The rest of druglike molecules, studied in the work, is antimicrobial preservatives: triclocarban, methylparaben, ethylparaben, propylparaben, butylparaben, and benzoic acid. We note that the pharmaceutical molecules contain carboxylic-, amino-, and ester-groups which are not present in the solutes from the training set. Nevertheless, the predicted PMVs for druglike molecules (Figure 5(b), red circles No. 15), are still in good agreement with the corresponding benchmark data (R = 0.986 and root mean square of error is 9.22 Å^3 that

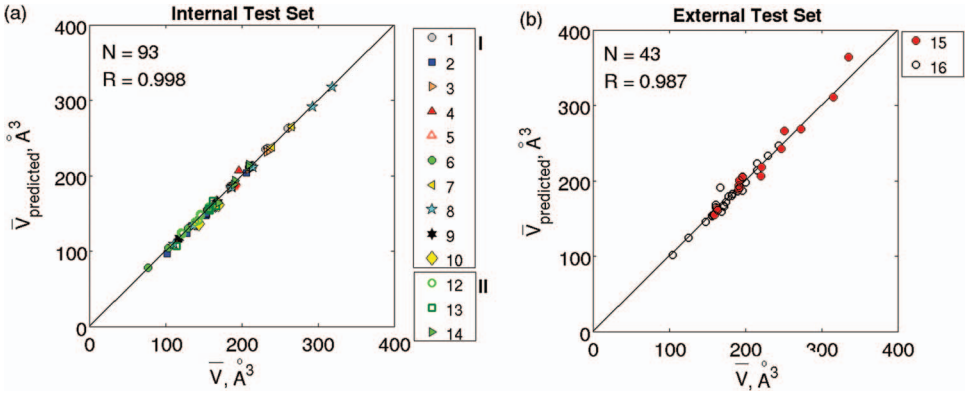

FIG. 5. Correlation between PMVs predicted with Eq. (6) and the corresponding benchmark values for the test set of solutes. (a) *Internal test set*: (**I**) homologous series of monofunctional compounds (the symbol notations equivalent to that on Figure 1(a)): 1 – alkanes, 2 – alkenes, 3 – alkylbenzenes, 4 – alcohols, 5 – phenols, 6 – 1-chloroalkanes, 7 – aldehydes, 8 – ketones, 9 – ethers, and 10 – dienes; (**II**) congeners of polyfunctional compounds (the symbol notations equivalent to that on Figure 1(b)): 12 – polychlorinated alkanes, 13 – polychlorinated alkenes, 14 – polychlorinated benzenes. (b) *External test set*: 15 – druglike molecules, 16 – other polyfunctional molecules.[9] The black solid lines give the line-of-best-fit. N is the number of solutes in the set. R is the correlation coefficient between predicted and benchmark PMVs.

is less than 5% of the corresponding average value of PMV). The results show that the developed model (Eq. (6)) accurately describes the PMV – α relationship for multiple classes of molecules ranging from simple organic solutes to different classes of polyfunctional molecules including druglike compounds.

To conclude, we demonstrated that the relationship between the partial molar volume in water and the static electric polarizability has in general a non-homogeneous character. However, a detailed analysis of this relationship reveals its fine well-organized structure that strongly depends on the chemical nature of the solutes. Thus, for compounds within a specific chemical class the PMV – α correlation has a strong linear behavior. For different chemical classes these correlations are parallel-shifted with respect to each other. We found that these shifts are caused by the specific solute-solvent interactions and can be presented as a function of the solute lone pairs and π-electrons. This concept resulted in a universal model that accurately describes the PMV – α relationship for various classes of polyfunctional solutes.

The authors thank David S. Palmer and Andrey I. Frolov for critical reading of the manuscript. Some part of the work was done when M.V.F. was visiting University of Lille 1 Science & Technology (Université Lille 1 Sciences et Technologies) as a Visiting Professor in the LASIR laboratory. M.V.F. appreciates the financial support from the University of Lille 1 for his visit. M.V.F. is thankful to Abdenacer Idrisi for organizing this visit and useful discussions. We acknowledge the supercomputing support from the John von Neumann-Institut für Computing (NIC), Juelich Supercomputing Centre (JSC), Forschungszentrum Juelich GmbH, Germany. Project ID: HLZ18. This work was partially supported by Deutsche Forschungsgemeinschaft (DFG) - German Research Foundation, Research Grant No. FE 1156/2-1 and Research Executive Agency (REA) (Grant No. 247500) "BioSol," Programme: FP7-PEOPLE-2009-IRSES.

[1] S. M. Free and J. W. Wilson, J. Med. Chem. **7**, 395 (1964).
[2] T. Imai, S. Ohyama, A. Kovalenko, and F. Hirata, Protein Sci. **16**, 1927 (2007).
[3] T. Brinck, J. S. Murray, and P. Politzer, J. Chem. Phys. **98**, 4305 (1993).
[4] P. Politzer, J. S. Murray, and F. A. Bulat, J. Mol. Model. **16**, 1731 (2010).
[5] P. Jin, J. S. Murray, and P. Politzer, Int. J. Quantum Chem. **96**, 394 (2004).
[6] L. Lepori and P. Gianni, J. Solution Chem. **29**, 405 (2000).
[7] E. L. Ratkova, G. N. Chuev, V. P. Sergiievskyi, and M. V. Fedorov, J. Phys. Chem. B **114**, 12068 (2010).
[8] D. S. Palmer, A. I. Frolov, E. L. Ratkova, and M. V. Fedorov, Mol. Pharmaceutics **8**, 1423 (2011).
[9] See supplementary material at http://dx.doi.org/10.1063/1.3672094 for description of training and test sets; experimental, benchmark, and predicted data for PMV and electric polarizability, their comparison; details of 3D RISM calculations.
[10] T. Imai, H. Nomura, M. Kinoshita, and F. Hirata, J. Phys. Chem. B **106**, 7308 (2002).
[11] L. Leu and D. Blankschtein, J. Phys. Chem. **96**(21), 8582 (1992).
[12] T. Imai and F. Hirata, J. Chem. Phys. **119**, 5623 (2003).
[13] T. Imai and F. Hirata. J. Chem. Phys. **122**, 094509 (2005).
[14] T. Imai, M. Kinoshita, and F. Hirata, J. Chem. Phys. **112**, 9469 (2000).
[15] M. Kinoshita, T. Imai, A. Kovalenko, and F. Hirata, Chem. Phys. Lett. **348**, 337 (2001).
[16] A. Kovalenko and F. Hirata, J. Phys. Chem. B **103**, 7942 (1999).
[17] S. Cabani, P. Gianni, V. Mollica, and L. Lepori, J. Solution Chem. **10**, 563 (1981).
[18] H. Durchschlag, T. Hefferle, and P. Zipper, Radiat. Phys. Chem. **67**, 479 (2003).
[19] A. V. Plyasunov and E. L. Shock, Geochim. Cosmochim. Acta **64**, 439 (2000).
[20] J. T. Edward, P. G. Farrell, and F. Shahidi, J. Chem. Soc., Faraday Trans. 1 **73**, 705 (1977).
[21] S. Sawamura, K. Nagaoka, and T. Machikawa, J. Phys. Chem. B **105**, 2429 (2001).
[22] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Gaussian, Inc., 2004.
[23] See http://cccbdb.nist.gov. for NIST, Computational Chemistry Comparison and Benchmark Database (accessed April, 2010).
[24] K. Miller, J. Am. Chem. Soc. **112**, 8533 (1990).
[25] T. V. Chalikian and K. J. Breslauer, Biopolymers **39**, 619 (1996).
[26] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, J. Comput. Chem. **21**, 132 (2000).
[27] A. Jakalian, D. B. Jack, and C. I. Bayly. J. Comput. Chem. **23**, 1623 (2002).