Parametrization Strategy for the MolFESD Concept: Quantitative Surface Representation of Local Hydrophobicity

Robert Jäger,† Stefan M. Kast, and Jürgen Brickmann*

Institut für Physikalische Chemie, Technische Universität Darmstadt, Petersenstrasse 20, 64287 Darmstadt, Germany

Received August 9, 2002

We derive a new model for the established concept of the *molecular free energy surface density* (MolFESD) yielding a more rigorous representation of local surface contributions to the overall hydrophobicity of a molecule. The model parametrization makes efficient use of both local and global information about solvation thermodynamics, as formulated earlier for the problem of predicting free energies of hydration. The free energy of transfer is separated into an interaction contribution and a term related to the cavity formation. Interaction and cavity components are obtained from the statistical three-dimensional (3D) free energy density and a linear combination of surface and volume terms, respectively. An appropriate molecular interaction field generated by the program *Grid* is used as an approximate representation of the interaction part of the 3D free energy density. We further compress the 3D density by means of a linear combination of localized surface functions allowing for the derivation of local hydrophobic contributions in the form of a free energy surface density. For a set of 400 compounds our model yields significant correlation ($R^2 = 0.95$, $\sigma = 0.57$) between experimental and calculated log *P* values. The final model is applied to establish a correlation between partial free energies of transfer for a series of sucrose derivatives and their relative sweetness, as studied earlier in the group of the authors. We find considerable improvement regarding the *rms* error of the regression thus validating the presented approach.

INTRODUCTION

Comprehension of the hydrophobic effect[1] on both microscopic and macroscopic scales has drawn tremendous attention because of its implication to a variety of physiological processes, i.e., metabolism, bioavailability, drug action, or receptor binding. There is an ongoing discussion in the literature concerning the molecular origin of the hydrophobic effect in terms of hydrophobic hydration and hydrophobic interaction.[2−7] A general theory on the molecular level to explain the positive free energy of transferring a nonpolar solute from the gas phase into water is yet to be conceived.

A quantitative, macroscopic treatment of the hydrophobic effect is accomplished using liquid−liquid systems and measuring the distribution behavior of a particular compound between an aqueous and an organic, immiscible phase. 1-Octanol has become the standard solvent in such a partitioning experiment,[8] and the logarithm of the partition coefficient, log *P*, is a prominent descriptor in many quantitative structure−activity relationships (QSAR).[9] The availability of an extensive amount of experimental log *P* values[10] has triggered the development of numerous empirical models designed to predict the log *P* value of a given compound.[11,12] These models, based on a two-dimensional representation of the molecular structure, fail when incorporated into studies of ligand−receptor interactions. As a consequence, use of molecular fields with respect to electrostatic and steric properties derived from the three-dimensional (3D) structure of the molecule, opened the field for the more successful 3D-QSAR[13] and 4D-QSAR[14] approaches, which in addition consider conformational aspects. Besides steric and electrostatic fields further attention was directed toward the establishment of other fields.[15] In particular the generation of hydrophobic fields, for instance by the program HINT,[16] has drawn broad interest since hydrophobic interactions play an important role in the binding process of a ligand and a receptor. Multiplication of atomic constants, derived from 1-octanol/water partition coefficients, with a particular distance function yields such hydrophobic fields. This spatial representation of lipophilicty remains on a qualitative level and renders inadequate for the calculation of log *P* values. Carrupt and co-workers present an excellent review of current computational approaches to lipophilicity including methodical aspects, limitations, and areas of application.[17]

To capture the physics of local hydrophobic interaction, the concept of the molecular free energy surface density (MolFESD) has been developed in the group of the authors.[18,19] Based on previously determined atomic lipophilicity parameters,[20] it was shown that the transfer free energy from one solvent to another can be represented by a surface integral of a scalar quantity, the MolFESD, over the solvent accessible surface of that molecule.[19] This concept was successfully applied to a series of sucrose derivatives by correlating the surface integral of partial surfaces to the relative sweetness of this series of compounds.[19]

* Corresponding author phone: +49 6151 162198; e-mail: brick@pc.chemie.tu-darmstadt.de.

† Present address: Aventis Pharma Deutschland GmbH, Industriepark Höchst G838, 65926 Frankfurt, Germany.

10.1021/ci025576h CCC: $25.00
Published on Web 12/14/2002

The local hydrophobicity models developed so far make no explicit use of the microscopic physics of local solvation.[13,17] In the present work, a new parametrization strategy for the MolFESD is presented that accounts for local and global information simultaneously. *Local* contributions are considered by defining the statistical *three-dimensional free energy density*[21,22] (3D-FED) as the objective function for a surface based regression model. The rigorous derivation of the 3D-FED from an exact thermodynamic, statistical basis[23] allows the plausible interpretation of model parameters in terms of local contributions. *Global* information, e.g. the log *P* value or the transfer free energy for the 1-octanol/water system, is used with an optimal weight in the parametrization process, leaving the parameter basis unchanged. The optimal regression model leads to the treatment of the transfer free energy as a surface function. This allows the local representation of lipophilicity on the molecular surface in a quantitative manner so that the surface integral over this density yields the transfer free energy. Based on our results, in a final step we obtain partial free energies of transfer from incremental surface areas of a series of sucrose derivatives. These energies are correlated with the relative sweetness of this series of compounds, and the results are compared with a correlation established from an earlier model.[19]

The newly derived model is a generalization of earlier work on the prediction of free energies of hydration,[21] termed "ΔG_{hyd} model" in what follows. Therefore, the theoretical basis is only briefly given, and the reader is referred to ref 21 for conceptual and technical details. Ahead of a brief outline of our model, we give a short review of the MolFESD concept and its main aspects. For further details, the reader is referred to refs 18 and 19. Although this new model can be established in a way to adequately predict log *P* values, the intention of the present work is to validate the concept of deriving a surface-based representation of hydrophobicity derived from 3D thermodynamic information. Extension to a competitive program to predict log *P* values is certainly feasible but not pursued at this point.

THE MOLFESD CONCEPT

To determine the free energy of solvation of a given compound in water two main contributions have to be considered: (a) the interaction energy between the solute and the surrounding solvent and (b) entropic contributions which are mainly a function of shape and size of the cavity which accommodates the solute. Consequently, one can distinguish two classes of molecules that constitute limiting cases in terms of their solubility in water: First, the class of ionic or strong polar compounds, where the solution thermodynamics are dominated by the electrostatic energy and second, nonpolar compounds, where the positive free energy of solvation (unfavorable solubility) is primarily influenced by the work spent for cavity formation.

In our previous work we showed that the electrostatic interaction energy, for dilute solutions, can be represented strictly as a function of the molecular surface.[19] Moreover, following the argumentation of several other authors, we arrived at a surface-based description for the free energy of cavity formation.[19] From these results we postulated that for all intermediate cases the free energy of solvation can be adequately described by a suitably chosen continuous function, $\rho^{FESD}(\mathbf{r}_S)$, of surface coordinates $\mathbf{r}_S$ in such a way as the integral over the entire surface *S* recovers the solvation free energy

$$\Delta G_{solv} = \Delta G_e + \Delta G_n = \int_S \rho^{FESD}(\mathbf{r}_S) d\mathbf{r}_S \qquad (1)$$

where ΔG_e and ΔG_n are the electrostatic contributions to the free energy of solvation and the free energy of cavity formation, respectively. Having established this physical basis, we derived an empirical model for the transfer free energy. Here, the MolFESD, $\rho^{FESD}(\mathbf{r}_S)$, represents the transfer free energy per unit surface and has its physical meaning only on the molecular surface. Following a similar notion, Luque et al. obtained the solvation free energy from a summation of discrete surface elements.[24]

In the previous paper[19] it is demonstrated that the MolFESD can be represented as a superposition of local contributions

$$\rho^{FESD}(\mathbf{r}_s) = \sum_l \rho_l(\mathbf{r}_s) = \sum_l F_l \mu_l(\mathbf{r}_s) \qquad (2)$$

where ρ_l is the MolFESD of an atomic or molecular increment *l*, F_l is an increment constant, and μ_l is a membership function which governs the influence of an increment *l* on the surface. The set of surface-related parameters F_l was determined based on partial log *P* values from an atom classification scheme.[20] With this approach, we were able to assess the hydrophobic contribution of a given surface element in a quantitative manner. This was successfully demonstrated by correlating partial transfer free energies (obtained from incremental surface areas) to the relative sweetness of a series of sucrose derivatives.[19]

In this paper we apply a parametrization strategy to arrive at an improved model for the MolFESD. This strategy has proven useful for the prediction of hydration free energies,[21] and it is now extended to build a model for a surface-based description of the transfer free energy. Essential aspects of this strategy are briefly reviewed in what follows.

NEW PARAMETRIZATION OF THE *MOLECULAR FREE ENERGY SURFACE DENSITY*

Model I: The *Grid* Approximation to the 3D-FED. As in the ΔG_{hyd} model the solvation free energy ΔG_{solv} is separated into an interaction and a cavity term

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{int} \qquad (3)$$

where the subscripts *cav*/*int* denote cavity and interaction contributions, respectively. Applied to the transfer free energy ΔG_{trans} for the 1-octanol/water system (superscript *oct* and *water*, respectively), we get

$$\Delta G_{trans} = \Delta G_{solv}^{water} - \Delta G_{solv}^{oct} = (\Delta G_{int}^{water} - \Delta G_{int}^{oct}) + (\Delta G_{cav}^{water} - \Delta G_{cav}^{oct}) \qquad (4)$$

The transfer free energy or its proportional quantity, the log *P* value, is related to the 3D-FED

$$\Delta G_{trans} = -2.303RT \log P = \int_V (\rho_{solv}^{water}(\mathbf{r}) - \rho_{solv}^{oct}(\mathbf{r})) d\mathbf{r}$$

$$= \int_V \rho_{trans}(\mathbf{r})d\mathbf{r} \quad (5)$$

where ρ_{trans} is the 3D-FED, $\mathbf{r}$ is a spatial vector, V is the total volume of the system, R is the gas constant, and T is the temperature. In analogy to eq 4, ρ_{trans} can be separated in a similar manner by redefining integration boundaries. The solvent accessible surface[25] is a reasonable border dividing the total volume into cavity and interaction regions. In the following the term *molecular surface* is taken as the solvent accessible surface described by Conolly.[25]

In the ΔG_{hyd} model[21] the interaction part of the 3D-FED was approximated with the program *Grid*[26] using the water probe only. For modeling the interaction part of the *transfer* free energy this becomes the difference of the water and the 1-octanol interaction terms as in eq 4. Regarding the 1-octanol phase as a simple nonpolar solvent, neglecting its partially polar properties due to the hydroxyl group and the ability to dissolve a substantial amount of water,[27,28] the interaction part is best approximated by the *hydrophobic probe* supplied by *Grid*. Although characteristic features of the 1-octanol phase are ignored, the hydrophobic probe offers the best model within the *Grid* approximation. Thus, the interaction parts of the 3D-FED of water and 1-octanol are approximated by *Grid* using the water and the hydrophobic probe, respectively, as

$$\Delta G_{int} = \int_{V_{int}} (\rho_{int}^{water}(\mathbf{r}) - \rho_{int}^{oct}(\mathbf{r}))d\mathbf{r} \approx \alpha \sum_m (\epsilon_m^{water} - \epsilon_m^{phob}) \quad (6)$$

where ϵ_m is the energy value of grid point m of the water and the hydrophobic probe (superscript *phob*) and α is a model parameter.

The cavity contributions are assumed to be a linear function of a surface and a volume term.[22,29] While the issue of surface and/or volume proportionality of the transfer free energy remains a subject of active debate,[30,31] both terms were included to avoid a biased cavity model

$$\Delta G_{cav}^{water} - \Delta G_{cav}^{oct} = (\beta^{water} - \beta^{oct})S + (\gamma^{water} - \gamma^{oct})V = \beta S + \gamma V \quad (7)$$

with S being the molecular surface of the water phase, V is the volume enclosed by this surface, and β and γ are the parameters. In this equation we implicitly assume that the molecular surface is similar in both phases and the difference of surface and volume parameters is described by a single parameter only. A constant term in eq 7 is neglected for reasons discussed in ref 21. Equations 6 and 7 are combined to establish model I for partition coefficient log P

$$\log P = \alpha \sum_m (\epsilon_m^{water} - \epsilon_m^{hydro}) + \beta S + \gamma V = \alpha \sum_m \epsilon_m^{diff} + \beta S + \gamma V \quad (8)$$

where superscript diff denotes the difference in grid energy densities.

The clearance of the grid was set to 4 Å, and the spacing of grid points was chosen to be 0.5 Å. Such grid dimensions result in an average number of grid points of approximately 34 400 per molecule. For each grid point, the energy value of the hydrophobic probe was subtracted from the energy value of the water probe. This difference according to eq 6 divided by the volume element of the grid gives the approximation of the interaction part of the 3D-FED. The molecular surface and volume were computed for each molecule,[25] and for a set of 400 compounds with known experimental log P value the parameters α, β, and γ were determined by linear least squares regression. The majority of the 400 compounds represent a subset of molecules that were given to the authors to be evaluated with the MolFESD software for a comparative analysis of several methods for predicting log P values.[32] This subset represents those compounds for which we were able to obtain experimental log P values. For this study, no further attention was directed toward the design of an appropriate training and test set which would allow to establish a model with high predictive capability. Work in this direction is currently ongoing.

Despite the simplicity of the model concerning the approximation of the interaction part of the 3D-FED for the transfer free energy the fit results, $R^2 = 0.81$ with a *rms* error of $\sigma = 1.64$ kcal/mol, are good.

The same regression was then repeated omitting the volume term. Interestingly, the fit statistics remain almost identical ($R^2 = 0.80$, $\sigma = 1.67$ kcal/mol) to the former fit, where the volume term was part of the regression model. This leads to the assumption that volume-proportional terms play a negligible role for the transfer free energy. In contrast, leaving the surface-proportional term out of the regression model reduced the quality of the fit significantly ($R^2 = 0.73$, $\sigma = 1.94$ kcal/mol). From these results we conclude that the transfer free energy is dominated by contributions proportional to the molecular surface. We further discuss this issue below.

Next, following the outline of the ΔG_{hyd} model derivation,[21] the reference grid is approximated by a linear combination of surface-based functions. This is an intermediate step toward the final model where local and global thermodynamic information will simultaneously enter the regression model.

Model II: Expansion of the 3D-FED in Surface-Based Functions. The rational behind this second model is 2-fold: one is the improvement of computational speed, the other is the establishment of a set of structure related parameters that allow the fast prediction of log P values by means of a group contribution approach. Furthermore, since the model functions are located on the molecular surface, the surface representation of local lipophilicity in a quantitative manner becomes possible. This aspect will be further examined after the final model section.

The essential characteristic of this second model is the approximation of the reference FED from *Grid* by a linear expansion of surface-based functions $\varphi_j(\mathbf{r}\text{-}\mathbf{r}_i)$. These functions were chosen as the product of an exponential distance dependence and angular dependent terms, which are given by Legendre polynomials with j being the degree of the expansion. The vectors $\mathbf{r}$ and $\mathbf{r}_i$ point from the origin of the coordinate system to a particular grid point and to a given reference point of surface patch i, respectively. The angle is determined from the normal vector of the reference point of patch i and $\mathbf{r}\text{-}\mathbf{r}_i$. Also, a parameter k was introduced to control

the exponential behavior and was set to an optimized value of 0.8 $Å^{-1}$ which was derived from preliminary investigations. The linear combination to approximate the *Grid* FED then reads

$$\rho_{\text{int}}^{\text{diff}}(\mathbf{r}) = \alpha\frac{\epsilon^{\text{diff}}(\mathbf{r})}{\Delta V} \approx \alpha\sum_i\sum_j a_{ij}\phi_j(\mathbf{r} - \mathbf{r}_i) \qquad (9)$$

where ΔV is the volume element occupied by a single grid point and α is a model parameter.

To evaluate this equation several steps are necessary that are only briefly sketched here (for details see ref 21). For each molecule of the training set the molecular surface was computed.[25] The surface was then divided into a certain number of patches *i*, each patch representing the surface area of the underlying atom. An atom classification scheme[20] was employed so that each patch is not only associated with the closest atom but also with the appropriate type of this atom from the classification scheme. Since this classification scheme uses roughly 120 different atom types, the number of different surface patch types is also restricted. The energy density of a grid point at position $\mathbf{r}$ is given by

$$\frac{\epsilon(\mathbf{r})}{\Delta V} = \sum_i\sum_j a_{lj}^T(i)\phi_j(\mathbf{r} - \mathbf{r}_i) \qquad (10)$$

where *i* is the surface patch index and *j* the degree of the Lengendre expansion. The parameter $a_{lj}^T(i)$ assigns an identical value to all patches *i* that belong to the same atom type *l*.

Integrating eq 10 gives the interaction part of the transfer free energy. Analytical integration yields a "structural constant", *c*, for each expansion degree *j* of the Legendre polynomial. With the cavity contribution from eq 7 and the number of occurrences *h* of a surface patch associated with atom type *l*, model II for log *P* value therefore reads

$$\log P = \alpha\sum_{l=1}^{120} h_l\sum_j a_{lj}^T c_j + \beta S + \gamma V \qquad (11)$$

The parametrization is accomplished in two steps. First, we determine the parameters a_{ij} by linear least squares regression according to eq 10. The enormous amount of data points can be substantially reduced, since only a fraction of all grid points is relevant in order to capture the interaction of a chosen probe and the target structure. For that purpose, a cubic sub grid, or *patch grid*, was assigned to each surface patch enclosing only those grid points that are in close neighborhood to the surface patch and therefore being most influential. Again, details of such reduction strategies are described elsewhere.[21] For the given training set of 400 compounds the optimal side length of the patch grids was found to be 2.5 Å.

The parameters α, β, and γ were subsequently fitted by linear regression. The results for the first step are reported in Table 1 for different degrees of expansion of the Legendre polynomial. The excellent approximation of the grid energy density by the linear combination of surface-based functions is shown in Figure 1. The parameters a_{ij} are taken from the $j = 0$ and $j = 1$ expansion. The yellow solid surface represents a constant energy density calculated from *Grid*.

Table 1. Results of the Least Squares Fit of the Model Density with Respect to the Reference *Grid* Values According to Model II (Eq 10)[a]

order *j* of Legendre polynomial	0	1	2
c_j [$Å^3$]	32.9381	10.8363	−3.7062
R^2	0.882	0.890	0.893
σ [kcal/mol]	0.623	0.602	0.596
F	213201	116423	79724

[a] The correlation coefficient R^2, the mean error σ, and the *F* statistic are listed depending on the maximum order of the Legendre expansion.

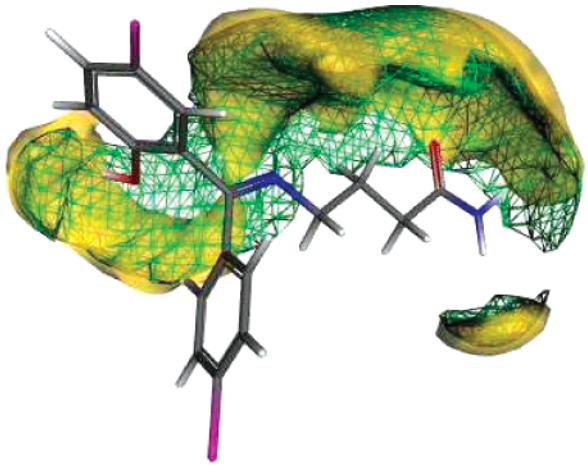

Figure 1. Overlap of two iso-surfaces for the molecule progabide. The yellow surface represents a constant density value from the *Grid* calculations. The green surface is the result for the same density modeled by the linear combination of surface values a_{lj}^T (see eq 10).

The green, mesh-like surface corresponds to the same energy density modeled from the surface functions.

Although the size of the training set is five times greater than the one used to predict free energies of hydration,[21] the correlation coefficients are very similar. Also, since the average size of the molecules is larger in the log *P* data set, the amount of data points of the regression model exceeds by a factor of 5. Taking this into consideration the correlation is excellent. As previously observed,[21] the exponential distance function alone ($j = 0$) captures a substantial part of the reference data. Adding angular terms yields an expected improvement, although it becomes insignificant beyond $j = 1$ with respect to the accuracy of the model. Consequently, the fit results for the free energy of transfer are reported for the $j = 0$ and $j = 1$ expansion only. According to eq 11 we obtain a correlation coefficient of $R^2 = 0.83$ with $\sigma = 1.54$ kcal/mol. The slight improvement when compared to the simple *Grid* model I might not be overrated. We conclude that the energy density determined by *Grid* can be modeled well by an expansion in terms of Legendre polynomials according to eq 10 yielding comparable fit statistics to model I.

The observations made in model I regarding the omission of the volume term are again confirmed. The fit statistics are only marginally affected when the volume term is left out of regression model II. Since both models show this invariance to a volume-proportional term, for the final model described in the next section, the cavity contribution is assumed to be a function of the surface only.

Final Model: Simultaneous Use of Local and Global Information. While in model II the parametrization was performed in two distinct steps, rearrangement of eq 11 allows the global information to enter the regression model together with the local grid data to find an optimal set of parameters a_{ij}. Omitting the volume term eq 11 is rewritten as

$$\frac{\log P_{\text{exp}} - \beta S}{\alpha} = \sum_{l=1}^{120} h_l \sum_{j=0}^{1} a_{lj}^T c_j \qquad (12)$$

where $\log P_{\text{exp}}$ is the experimental log P value of a given substance of the training set. All other variables have been defined above. The initial design matrix therefore consists of two blocks, termed block A and block B. The parameter basis is equal for both blocks as one can easily see by comparing eqs 10 and 12. The dependent variables for block A are the energy values computed from *Grid* divided by the volume element, whereas block B uses the LHS of eq 12 as reference data. With such a matrix scheme the function χ^2 to be minimized is given by eq 13

$$\chi^2 = \sum_m (obs_{m,\text{A,ref}} - obs_{m,\text{A}})^2 + w \sum_p (obs_{p,\text{B,ref}} - obs_{p,\text{B}})^2 \qquad (13)$$

The summation of block A is carried out over all relevant grid points m. The reference data, $obs_{m,\text{A,ref}}$, and the modeled data, $obs_{m,\text{A}}$, are given by the LHS and RHS of eq 10, respectively. Based on eq 12 the B block is built in analogous manner with p being the number of molecules. Since the number of grid points m is much larger than the number of compounds p, a weight factor w is introduced to find the optimal balance between the two blocks.

Iteration Cycle. To initially compute the LHS of eq 12 the parameters α and β from model II are used as starting values. The function χ^2 is minimized (see eq 13) yielding a set of parameters a_{lj}^T. From these parameters new values of α and β are determined by least squares regression according to model II (see eq 11) omitting the volume term. This allows to recalculate the LHS of eq 12 and to repeat the parametrization procedure in an iterative manner until convergence is achieved.

RESULTS

As before, the results for the iterative procedure are reported for the $j = 0$ and $j = 1$ expansion only. The convergence of the *rms* error for the prediction of the experimental log P values was taken as a stopping criterion for the iteration cycle. We obtain a minimum *rms* error for the log P prediction of $\sigma = 0.57$ log units (or 0.78 kcal/mol), while the corresponding correlation coefficient is $R^2 = 0.95$. The training set of 400 molecules comprised 89 atom types from the atom classification scheme.[20] For these atom types, the resulting parameters a_{lj}^T are listed in Table 2. We obtain for the final model

$$\log P = 1.754 \cdot 10^{-4} \text{mol kcal}^{-1} \sum_{l=1}^{120} h_l \sum_{j=0}^{1} a_{lj}^T c_j + 0.019598 \text{ Å}^{-2} S \qquad (14)$$

Table 2. Atom Types According to the Classification Scheme by Ghose et al.[20] and the Parameters of the Final Model for $j = 0$ and $j = 1$ Terms in the Legendre Expansion

atom type (l)	a_{l0}^T [kcal mol^{-1} Å^{-3}]	a_{l1}^T [kcal mol^{-1} Å^{-3}]	atom type (l)	a_{l0}^T [kcal mol^{-1} Å^{-3}]	a_{l1}^T [kcal mol^{-1} Å^{-3}]
1	53.409	−49.752	51	−66.325	41.462
2	14.208	−12.501	52	−14.056	−1.525
3	26.689	−26.843	53	−17.787	3.503
5	19.038	−1.378	54	−4.510	−14.266
6	−9.692	17.823	56	−98.561	66.538
7	−0.014	12.107	57	−13.460	−67.202
8	−15.172	3.902	58	−104.299	27.255
9	−34.025	7.138	59	−19.112	−39.154
11	−337.530	266.667	60	37.136	−91.475
12	725.400	−912.779	61	−89.267	17.870
13	−43.737	−25.158	62	−219.804	66.224
15	114.764	−72.910	66	−312.465	236.463
16	22.988	−27.783	67	−73.084	10.594
17	7.122	−14.064	68	97.552	−208.446
18	−147.715	156.477	69	−125.206	99.179
19	2.519	−37.144	70	−9.784	14.268
20	70.520	−69.428	71	34.449	13.419
21	−184.355	164.391	72	−54.458	42.981
22	196.768	−197.963	73	97.987	−90.401
24	35.717	−46.227	74	−60.895	−37.753
25	15.407	−37.921	75	−80.600	−17.668
26	1.307	−28.504	76	98.618	−133.236
27	−29.789	23.790	77	−5.376	20.874
28	17.022	−34.449	78	199.837	−271.567
29	24.566	−38.275	79	−85.309	58.111
30	−172.451	124.232	81	27.729	−102.977
31	−69.566	57.977	83	91.825	−108.728
32	102.599	−127.133	84	−24.330	−26.998
33	−142.321	133.190	86	87.878	−83.598
34	−38.530	2.676	87	−31.514	27.565
35	3.326	−32.210	88	1.574	6.254
36	39.013	−24.129	89	58.509	−49.160
37	−29.267	67.043	90	1.551	−1.719
38	15.997	−28.475	91	74.386	−74.935
39	−32.394	42.356	94	109.924	−94.889
40	−19.191	7.663	95	−2.585	0.234
41	81.500	−75.874	96	152.798	−169.690
42	−253.751	196.069	99	28.582	−44.374
43	−172.380	194.056	106	22.699	−9.884
44	−112.645	93.604	107	84.576	−94.481
46	−5.733	−7.634	108	−272.909	214.790
47	−52.775	39.041	109	−167.539	136.562
48	−46.753	37.020	110	13.258	−86.145
49	−28.639	36.060	117	−31.768	−46.726
50	−50.621	1.494			

By fitting local and global information simultaneously, we were able to lower the *rms* error by a factor of 2 compared to the simple model I thereby validating our approach. We might anticipate further improvement through a better model of the 1-octanol phase beyond the *Grid* approximation. In addition, refinement of the cavity contributions is advisable. Based on the conclusions of several authors,[6,30,33] enhancement is to be expected from incorporation of descriptors related to shape and curvature of the molecular surface. The resulting correlation plot is shown in Figure 2, where calculated log P values are plotted against experimental log P data.

The list of the experimental and calculated log P values from eq 14 for the 400 compounds present in the training is shown in Table 3. Although we find good agreement in terms of the *rms* error of the transfer free energy when compared to other fragment approaches,[11,20] the model was not further optimized to accurately predict log P values. Instead, in the next section we will outline the use of 3D thermodynamic information given by the molecular interaction field for a more rigorous derivation of local hydrophobicity.

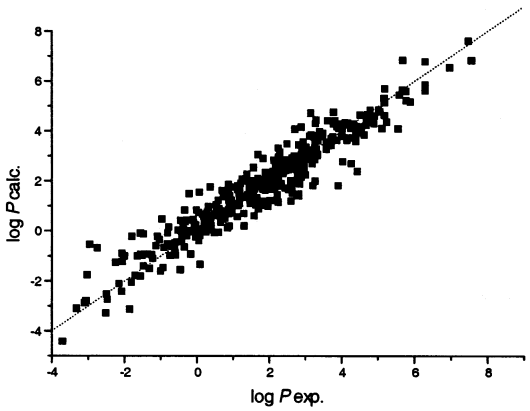

Figure 2. Log P values as obtained from our final model (eq 14) are plotted versus experimental data. The fit statistics are $R^2 = 0.95$ and $\sigma = 0.57$ log units.

SURFACE REPRESENTATION OF THE FINAL MODEL BY MEANS OF A MOLECULAR FREE ENERGY SURFACE DENSITY

Despite the success of the MolFESD approach in terms of a quantitative description of local hydrophobicity,[19] a more rigorous derivation of model parameters is desired. Because of its sound physical basis, the corresponding three-dimensional quantity, i.e., the 3D-FED, offers a convenient starting point to derive a free energy surface density. For a given expansion degree j of the Legendre polynomials, the summation over j in eq 11 can be carried out yielding a parameter f_i

$$f_i = \sum_j a_{ij} c_j = \frac{\sum_j a_{ij} c_j}{S_i} S_i = \langle \rho^{\mathrm{FESD}} \rangle_i S_i \qquad (15)$$

which has the meaning of an average FED, $\langle \rho^{\mathrm{FESD}} \rangle$, when divided by the area S_i of the associated patch i. For the sake of clarity, i is again taken as the index of surface patches of a molecule. Obviously, if two patches are associated with the same atom type, the f_i value is identical.

Such a surface-based representation is advantageous when partial surfaces are considered and one is interested in their contribution to global observables, e.g. the transfer free energy. The treatment of partial surface areas is already possible using our final model with the condensed description of the interaction part in terms of f_i values (see eq 15). The summation in eq 14 is to be taken over all patches present for the given partial surface, with S_i being the related surface area, yielding a partial thermodynamic quantity. When fractions of patches exist, the corresponding fraction of the f_i values has to be taken into account.

Fractional handling is avoided when a *continuous* surface representation is employed. Toward a description in terms of a free energy surface density, we accomplish a continuous representation of the patch-based f_i values by a weighted normalized summation for each surface point s

$$f_S(\mathbf{r}_s) = \frac{\omega \sum_i f_i g_i(d_{is})}{\sum_i g_i(d_{is})} \qquad (16)$$

with

$$g(d_{is}) = e^{-a d_{is}^2} \qquad (17)$$

where ω is a scaling factor, f_s is the interaction contribution of a surface point s given by vector $\mathbf{r}_s$, $g_i(d_{is})$ is a Gaussian-type function with parameter a, and d_{is} is the distance *along the surface* between a given surface point, s, and the reference point of patch i. While in principle the summation in eq 16 is carried out over all patches i, we consider only the neighboring patch values f_i to contribute to a given surface point. The model equation for the transfer free energy can now be written, without any loss of global information, as

$$\Delta G_{\mathrm{trans}} = \alpha \sum_s f_s + \beta S \qquad (18)$$

Surface Representation. While the f_i values from eq 15 were localized on the reference points of the surface patches, we have now available a continuous description of the transfer free energy in terms of contributions from all points of the molecular surface. Under the condition that the sum of all surface point areas O_s recovers the total surface area S, eq 18 is written as

$$\Delta G_{\mathrm{trans}} = \alpha \sum_s \frac{f_s}{O_s} O_s + \beta \sum_s O_s \qquad (19)$$

From eq 19 follows immediately

$$\Delta G_{\mathrm{trans}} = \sum_s (\alpha f'_s + \beta) O_s; \; f'_s = \frac{f_s}{O_s} \qquad (20)$$

The relation to the free energy surface density becomes now apparent as

$$\Delta G_{\mathrm{trans}} = \int_S \rho^{\mathrm{FESD}}(\mathbf{r})\, d\mathbf{r} \approx \sum_s \rho^{\mathrm{FESD}}(\mathbf{r}_s) O_s = \sum_s (\alpha f'_s + \beta) O_s \qquad (21)$$

where s is the index of all surface points, $\mathbf{r}_s$ are the corresponding point coordinates, and O_s is the surface point area.

While there is in principle no need for a surface-based representation of the parameters derived from the final model, such a representation is advantageous when partial surfaces are considered. If the free energy surface density is known for a molecule, integration of a surface fraction yields a partial contribution to the global observable, e.g. the transfer free energy. Since both local and global information were considered for the final model in the previous section, the patch-based f_i values represent a highly plausible distribution of local information gathered form the 3D-FED. The f_i values act as sampling points thereby preserving the local information for a continuous representation on the molecular surface (see eq 16). Additionally, global information is completely retained as the transfer free energies from eq 21 will be identical to the values obtained from eq 11 omitting the volume term.

Table 3. Experimental and Calculated log *P* Values as Obtained from the Final Model (Eq 14)

compound	log P_{exp}	log P_{calc}	compound	log P_{exp}	log P_{calc}	compound	log P_{exp}	log P_{calc}
1-bromopropane	2.10	2.18	cefmetazole	−0.60	−0.49	flunarizine	5.78	5.21
17-α-hydroxyprogesterone	3.17	3.01	cefoperazone	−0.74	−1.00	flunitrazepam	2.06	1.97
17-methyltestosterone	3.36	3.82	cefotaxime	−1.36	−0.94	fluorouracil	−1.00	−1.62
2-chloroaniline	1.90	1.87	cefuroxime	−0.16	−0.94	fluoxetine	3.82	4.41
2-chloropyridine	1.22	1.89	cetirizine	1.70	3.04	flupentixol	4.51	4.63
2-decanone	3.73	3.58	chloralhydrate	0.99	0.88	fluphenazine	4.36	4.21
2-iodoaniline	2.32	1.85	chloramphenicol	1.14	1.06	flurbiprofen	4.16	3.64
2-naphthylamine	2.28	2.30	chloroquine	4.63	3.84	flutamide	3.35	3.05
2-nitrophenol	1.79	1.83	chlorothiazide	−0.24	−0.05	ftorafur	−0.27	−0.16
3-chloroaniline	1.88	1.90	chlorpheniramine	3.17	3.84	furosemide	2.03	1.20
3-nitrophenol	2.00	1.90	chlorpromazine	5.19	4.61	ganciclovir	−2.07	−2.43
4-chloroaniline	1.88	1.90	chlorpropamide	2.27	2.12	gitoformate	3.16	2.89
4-methoxyacetophenone	2.03	2.17	chlorprothixene	5.18	5.30	gitoxin	1.68	1.79
4-nitrophenol	1.91	1.90	cimetidine	0.40	0.51	glipizide	1.91	1.71
6-azauridine	−2.14	−2.12	cinnarizine	5.77	5.58	griseofulvin	2.18	2.69
acebutolol	1.71	0.87	ciprofloxacin	−1.08	−0.23	haloperidol	3.23	3.76
acetazolamide	−0.26	0.42	clindamycin	2.16	1.69	hexobarbital	1.49	1.46
acetylsalicylate	1.19	1.09	clobazam	2.12	2.07	histidine	−3.32	−3.12
aciclovir	−1.56	−1.82	clofazimine	7.48	7.60	hydrochlorothiazide	−0.07	0.09
alclofenac	2.47	2.58	clomethiazole	2.12	2.24	hydroflumethiazide	0.36	0.50
α-bromotoluene	2.92	2.95	clomipramine	5.19	5.14	hydroxyurea	−1.80	−2.06
alprazolam	2.12	1.99	clonazepam	2.41	2.61	hymecromone	1.90	2.34
alprenolol	2.89	2.61	clonidine	1.57	2.52	ibuprofen	3.50	3.90
altretamine	2.73	1.98	clotiazepam	3.49	3.77	ifosfamide	0.86	0.77
amantadine	2.44	1.71	cloxacillin	2.43	2.81	imipramine	4.80	4.83
amben	−1.55	−0.11	cocaine	2.30	2.38	indinavir	2.92	2.89
aminocaproic acid	−2.95	−0.55	codeine	1.14	1.80	indometacin	4.27	2.69
aminosalicylic acid	1.32	0.19	colchicine	1.30	1.98	indoprofen	2.77	2.74
amiodarone	7.57	6.81	corticosterone	1.94	1.88	iodoethane	2.00	2.03
amitriptyline	4.92	4.59	cortisol	1.61	1.29	iodomethane	1.51	1.62
amoxicillin	−1.99	−1.02	cortisone	1.47	1.22	iohexol	−3.05	−2.81
ampicillin	−1.13	−0.60	coumarin	1.39	2.27	iopromide	−2.05	−1.23
anthracene	4.45	4.17	cromolyn	1.92	1.76	isocarboxazid	1.49	2.00
aprobarbital	1.37	1.34	cyclophosphamide	0.63	0.88	isoniazid	−0.70	−0.63
ascorbic acid	−1.64	−1.02	cysteine	−2.49	−2.55	isosorbide	1.31	0.69
atenolol	0.16	0.65	cytarabine	−2.51	−3.30	isosorbide-2-mononitrate	−0.40	−0.42
atropine	1.83	1.66	dapsone	0.97	1.22	isosorbide-5-mononitrate	−0.40	−0.35
azatadine	3.59	3.26	debrisoquine	0.75	1.36	isotretinoin	6.30	5.84
azelaic acid	1.57	1.37	desipramine	4.90	4.28	isoxicam	2.83	1.40
baclofen	−0.96	0.46	dexamethasone	2.01	2.24	ketamine	2.18	2.46
bendroflumethiazid	1.19	1.68	dexfenfluramine	3.36	3.59	ketanserin	3.29	3.00
benorylate	2.15	1.90	dextroamphetamine	1.76	1.10	ketoconazole	4.34	4.28
benznidazole	0.91	0.18	diamorphine	1.14	1.65	lacidipine	5.56	4.08
benzotrifluoride	3.01	3.03	diazepam	2.99	2.92	lamivudine	−0.93	−1.48
betamethasone	2.01	2.14	diazoxide	1.20	1.60	lanatosid c	0.07	−0.40
betaxolol	2.81	2.85	diclofenac	4.40	3.58	levamisole	1.84	2.89
biperiden	4.25	4.11	dicloxacillin	2.91	3.18	levobunolol	2.40	2.28
bisoprolol	1.87	2.03	didanosine	−1.24	−0.97	levodopa	−2.74	−0.70
bromazepam	1.69	2.09	diethylstilbestrol	5.07	4.77	levofloxacin	−0.28	−0.09
bromoethane	1.61	1.67	diflunisal	4.44	2.37	lidocaine	2.26	2.96
bromomethane	1.19	1.26	digitoxin	2.83	2.69	lidoflazine	5.60	5.44
bromopride	2.83	1.46	digoxin	1.26	2.07	lincomycin	0.20	0.56
brotizolam	2.79	2.79	diltiazem	2.80	2.22	lomefloxacin	−0.80	−0.09
bunitrolol	1.91	2.23	diphenhydramine	3.27	3.50	loratadine	5.20	5.69
bupranolol	2.97	3.14	disulfiram	3.88	4.16	lorazepam	2.51	2.51
buprenorphine	3.29	4.29	dosulepine	4.49	4.17	lovastatin	4.26	4.09
busulfan	−0.52	0.02	doxorubicin	0.10	−1.35	m-245-trichlorophenol	3.72	3.40
caffeine	−0.07	0.28	enoxacin	−0.91	−0.68	m-246-tribromophenol	4.13	4.25
carazolol	3.59	3.55	enprofylline	0.33	0.97	*m*-iodoaniline	2.86	1.88
carbamazepine	2.19	2.73	ephedrine	0.93	1.40	*m*-xylene	3.20	3.24
carbenicillin	1.13	0.55	erythromycin	2.54	1.64	mannitol	−3.10	−2.89
carbon tetrachloride	2.83	2.57	estradiol	4.01	3.77	mebendazole	2.83	3.43
carbutamide	1.01	1.01	estriol	2.45	2.83	medazepam	4.41	3.86
carfecillin	2.96	2.59	estrone	3.13	3.79	mefenamic acid	5.12	4.09
carmustine	1.53	1.55	ethanol	−0.31	0.36	mefoxine	−0.02	−0.26
carvedilol	4.11	4.32	ethinylestradiol	3.67	3.53	mefruside	1.54	1.13
cathine	0.67	0.34	etoposide	0.60	1.21	meloxicam	3.01	2.24
cefacetrile	−0.45	−1.57	fampridine	0.32	0.05	memantine	3.28	1.99
cefaclor	−1.79	−0.23	fenbufen	3.20	2.68	mephenesin	1.41	1.44
cefadroxil	−2.06	−0.91	fenfluramine	3.36	3.59	mepindolol	2.30	2.69
cefaloram	0.20	0.10	finasteride	3.03	3.35	meprobamate	0.70	0.46
cefalotine	−0.41	−0.22	fleroxacin	−0.55	−0.45	mercaptopurine	−0.18	−0.01
cefamandole	−1.47	−0.14	flucloxacillin	2.61	2.57	methadone	3.93	4.09
cefatrizine	−2.24	−1.27	fludrocortisone	1.67	1.32	methamphetamine	2.07	2.17
cefazolin	−0.58	−0.88	flufenamic acid	5.25	4.35	methaqualone	2.50	3.48
methimazole	−0.34	−0.66	phenazone	0.38	1.75	sulfadimethoxine	1.63	1.03
methotrimeprazine	4.68	4.49	phenobarbital	1.47	1.55	sulfaguanidine	−1.22	−1.11
methylphenobarbital	1.84	1.80	phenprocoumon	3.62	4.41	sulfameter	0.41	0.20
metipranolol	2.28	2.42	phentermine	1.90	1.62	sulfamethizole	0.54	1.04

Table 3. (Continued)

compound	log P_{exp}	log P_{calc}	compound	log P_{exp}	log P_{calc}	compound	log P_{exp}	log P_{calc}
metoclopramide	2.62	1.18	phenylbutazone	3.16	4.70	sulfamethoxazole	0.89	1.21
metoprolol	1.88	1.66	phenytoin	2.47	2.54	sulfaperine	0.34	0.93
metronidazole	−0.02	0.01	picricacid	0.89	1.84	sulfapyridine	0.00	0.82
mexiletine	2.15	1.79	pimozide	6.30	6.77	sulfinpyrazone	2.30	3.14
midazolam	3.27	2.85	pindolol	1.75	1.88	sulfisomidine	−0.33	0.82
moricizine	2.98	3.23	pipamperone	2.02	1.68	sulfisoxazole	1.01	1.74
morphine	0.76	1.42	piperacillin	0.50	−0.08	sulindac	3.05	2.88
moxalactam	−0.58	−0.98	piperazine	−1.50	−1.00	sulfadiazine	−0.09	0.45
n-heptylamine	2.57	1.67	piracetam	−1.54	−0.95	sulfadimidine	0.89	1.54
n-octylamine	2.90	2.14	pirenzepine	0.10	−0.20	temafloxacin	−0.20	1.47
nadolol	0.71	1.61	piretanide	3.92	1.80	temazepam	2.19	2.04
nalidixic acid	1.59	0.59	piroxicam	1.98	0.97	terbutaline	0.08	1.53
naloxone	2.09	1.66	practolol	0.79	0.50	terfenadine	5.69	6.83
naproxen	3.34	3.01	prazepam	3.73	3.69	testosterone	3.32	3.27
nebivolol	4.04	2.76	prednisolone	1.62	1.73	tetrachloroethene	3.40	3.96
nedocromil	2.22	2.15	prednisone	1.47	1.53	tetracycline	−1.47	−1.41
niacinamide	−0.37	0.13	primidone	0.91	1.06	tetrahydrocannabinol	6.97	6.53
nicotine	1.17	1.49	probenecid	3.21	1.80	tetrazepam	3.20	2.73
niflumic acid	4.43	4.04	procainamide	0.88	0.83	tetroxoprim	0.56	0.91
nimesulide	2.60	2.52	procarbazine	0.06	0.73	theophylline	−0.02	0.23
nitrazepam	2.25	2.20	progabide	2.97	2.97	thiamphenicol	−0.27	−0.21
nitromethane	−0.35	0.06	progesterone	3.87	3.62	thioanisole	2.74	2.74
nordazepam	2.93	2.68	proguanil	2.53	2.53	thioridazine	5.90	5.16
norethindrone	2.97	3.06	promazine	4.55	4.06	tiabendazol	2.47	2.39
norfloxacin	−1.03	−0.37	promethazine	4.81	4.35	tiapride	0.90	0.68
nortriptyline	4.04	4.25	propicillin	2.65	1.71	timolol	1.83	1.18
o-xylene	3.12	3.18	propofol	3.79	4.27	tinidazole	0.23	0.60
octan-1-ol	3.00	3.24	propoxyphene	4.18	4.31	tolazamide	1.45	1.71
ofloxacin	−0.28	−0.09	propranolol	2.98	2.64	tolazolin	2.65	1.50
omeprazole	2.23	1.44	protirelin	−2.47	−2.75	tolbutamide	2.34	2.36
ornidazole	0.60	0.73	proxyphylline	−0.77	0.14	toliprolol	1.93	2.25
orphenadrine	3.77	3.76	pyrazinamide	−0.60	−0.76	tolmetin	2.79	2.30
ouabain	−1.70	−1.80	pyrene	4.88	4.44	tramadol	2.63	2.57
oxacillin	2.38	2.34	pyrilamine	2.98	2.42	tranylcypromine	1.49	1.08
oxamniquine	2.24	1.78	pyrimethamine	2.69	2.93	tretinoin	6.30	5.60
oxazepam	2.24	2.08	quazepam	4.03	3.98	triamcinolone	1.16	1.00
oxprenolol	2.10	2.37	quinidine	2.64	2.92	triamterene	0.98	0.70
oxyphenbutazone	2.72	4.06	quinine	2.64	2.85	triazolam	2.42	2.52
p-iodoaniline	2.34	1.88	ranitidine	0.27	−0.06	trichlormethiazide	0.56	0.60
p-xylene	3.15	3.24	remoxipride	2.13	2.82	trichloroethene	2.61	2.00
papaverine	2.95	3.56	retinol	5.68	5.63	trichloromethane	1.97	2.20
paracetamol	0.51	0.66	ribavirin	−1.85	−3.15	trifluoperazine	5.03	4.84
paraldehyde	0.67	0.80	rifampicin	1.87	1.77	trihexyphenidyl	4.49	3.94
pefloxacin	0.27	0.24	ropivacaine	2.90	4.15	trimeprazine	4.71	4.65
penbutolol	4.15	3.62	saccharin	0.91	0.76	trimethoprim	0.91	1.08
penicillin_g	1.83	1.02	salicylate	2.26	1.14	trimetrexate	2.55	2.43
penicillin_v	2.09	1.20	secnidazole	0.22	0.39	triprolidine	3.92	4.27
penta-1−4-dien	2.47	2.44	secobarbital	1.97	2.05	valproate	2.75	2.44
pentazocine	3.31	4.43	selegiline	2.90	3.04	verapamil	3.79	4.76
pentobarbital	2.10	1.99	simvastin	4.68	4.22	warfarin	2.70	2.99
pentoxifylline	0.29	0.75	sotalol	−0.44	0.41	xamoterol	0.61	−0.07
perphenazine	4.20	3.72	spironolactone	2.26	3.30	xylose	−3.02	−1.77
pethidine	2.45	2.63	stavudine	−0.81	−0.52	zalcitabine	−1.30	−1.51
phenacemide	0.87	0.12	sucrose	−3.70	−4.44	zofenopril	4.40	3.94
phenacetin	1.58	1.40						

The free energy surface density as determined from eq 21 is now applied to a series of sucrose derivatives, and the results are compared to a model based on a set of incremental log *P* values.[19]

Application to the Relative Sweetness of a Series of Sucrose Derivatives. A series of sucrose halogens derivatives has been examined previously with the MolFESD model.[19] It was found that the partial transfer free energy, computed from the surface area belonging to the fructose portion of the sucrose, nicely correlates with the relative sweetness of these derivatives. Those findings are based on particular assumptions regarding the model of the sweetness receptor[34−36] and the proportionality of the sweetness to the equilibrium constant *K* for the association of a sweetener and its receptor.[37] In this earlier treatment, the MolFESD values were parametrized based on a set of established incremental log *P* values,[20] and a continuous surface representation was found employing a membership function satisfying eq 1. These investigations yielded a correlation coefficient of $R^2 = 0.94$ and a *rms* error of 0.81. We will now reexamine this correlation behavior using the parameters derived from our model according to eq 14. Both patch-based, f_i values, and surface point-based, f_s values, are employed to establish a correlation as indicated above.

Patch-Based Approach. To obtain partial thermodynamic information from incremental surface areas, our final model equation (see eq 14) is well suited for such a purpose. The sum in eq 14 is to be taken over all patches that are present on the given partial surface area. In this equation, the sum of the f_i values represents the interaction part of the 3D-FED even if the sum is not taken over all patches *i* of a complete surface. If integration boundaries cut patches into fractions, the corresponding fraction of the f_i value has to be considered in the summation. With this approach the

Table 4. Transfer Free Energies Obtained Form the Final Model (See Eq 12)[a]

	substitution positions	surface increment, [Å^2]	$\Delta G_{transfer}$ [kcal/mol]	relative sweetness[19]
1		157.36	2.69	1
2	1′-Cl	163.56	0.88	20
3	6′-Cl	167.49	0.84	20
4	1′-Cl, 4′-Cl	171.47	−0.91	30
5	1′-Cl, 6′-Cl	173.58	−0.86	80
6	1′-Cl, 4′-Cl, 6′-Cl	181.54	−2.66	100
7		157.56	−1.63	5
8	6′-Cl	167.06	−0.77	50
9	1′-Cl	163.88	−0.75	120
10	4′-Cl, 6′-Cl	174.66	−2.56	160
11	1′-Cl, 4′-Cl	171.60	−2.54	220
12	1′-Cl, 6′-Cl	173.79	−2.49	600
13	1′-6′-Cl, 4′-F	170.18	−3.35	1000
14	1′-Cl, 4′-Cl, 6′-Cl	181.52	−4.28	2200
15	1′-Cl, 6′-Cl, 4′-Br	184.90	−4.29	3000
16	1′-6′-Cl, 4′-I	188.79	−4.73	3500

[a] Summation was carried out over all patches *i* that are present on the surface area of the fructose portion. Compounds 2−6 are derivatives of sucrose (**1**), while compounds **8**−**16** are derivatives of 4-chloro-4-deoxy-*galacto*-sucrose (**7**).

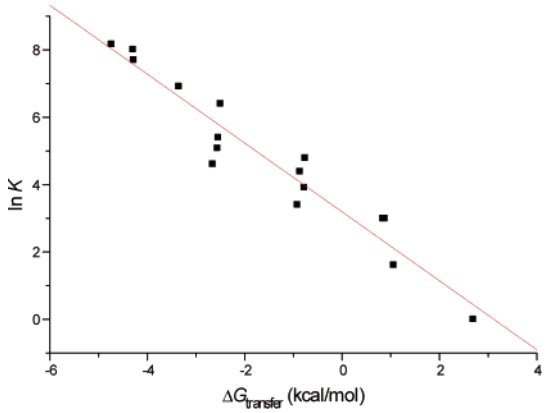

Figure 3. The relative sweetness, ln *K*, plotted as a function of the predicted partial transfer free energy, $\Delta G_{transfer}$; $R^2 = 0.96$, $\sigma = 0.64$.

results given in Table 4 for the sucrose derivatives were obtained. The sequence of compounds (**1**−**16**) is identical to the one used in the previous publication.[19] Correlating the partial transfer free energy with the relative sweetness yields $R^2 = 0.96$ and $\sigma = 0.64$. We consider this correlation excellent, because of its high correlation coefficient and the fact that none of the above listed compounds were present in the training set of the log *P* parametrization (except for pure sucrose). This selection of sweetners thus represents a small test set, and the obtained fit results render strong confirmation of our model approach.

Since the fit quality exceeds the one from ref 19, this is a clear indication of the plausibility of the f_i values in terms of local contributions. Such local information was already present in the parametrization process therefore leading to a more rigorous rationalization of local moieties. Assigning local or fragmental contributions from global information alone remains somewhat arbitrary, although its successful application[19] is undisputed. In Figure 3 we plot the natural logarithm of the relative sweetness versus the transfer free energy obtained from the partial surfaces of the sucrose derivatives.

Table 5. Resulting Partial Transfer Free Energies (kcal/mol) Obtained from Integration of the Surface Area of the Fructose Portion for Different Values of *a* of the Gaussian Function in Eq 17[a]

compound	$a =$ 0.5 Å^{-2}	$a =$ 0.125 Å^{-2}	$a =$ 0.055 Å^{-2}	$a =$ 0.0313 Å^{-2}
1	13.29	13.12	13.12	13.12
2	7.53	8.27	8.73	8.84
3	6.16	6.85	7.53	7.82
4	0.17	1.88	2.91	3.25
5	0.34	1.83	2.85	3.25
6	−7.36	−4.74	−3.08	−2.51
7	−6.80	−6.80	−6.80	−6.80
8	0.39	0.16	0.22	0.28
9	2.04	1.53	1.07	0.96
10	−7.71	−6.57	−5.66	−5.32
11	−5.54	−5.03	−4.80	−4.75
12	−6.97	−6.12	−5.37	−5.09
13	−10.57	−9.60	−8.74	−8.34
14	−14.62	−12.56	−11.31	−10.91
15	−14.22	−12.45	−11.25	−10.85
16	−19.12	−15.30	−13.36	−12.79
R^2	0.956	0.962	0.960	0.958
σ [log units]	0.70	0.65	0.67	0.68

[a] The ordering of compound is identical to Table 4. Correlation coefficients R^2 and *rms* errors σ for the correlation of ΔG_{trans} with the relative sweetness of compounds **1**−**16** (see Table 4) are also listed.

Continuous Representation. As it was shown above, depending on the available surface we were able to evaluate partial thermodynamic quantities from the discrete, patch-based f_i values alone. One might ask for a continuous representation in particular to benefit from visual inspection. Such a continuous function, i.e., the MolFESD (see eq 1), can be derived from the model parameters of eq 14. Although the local information inherent in our final model was used for the derivation of the MolFESD, an unambiguous surface representation is unattainable. According to eq 17, the Gaussian coefficient *a* can be adjusted to a particular application.

We now focus again on the series of sucrose derivatives. We determine the best possible surface representation with respect to an optimal correlation of partial surface integrals and the relative sweetness. The Gaussian parameter *a* of eq 17 was altered in such a way, as the turning point of the Gaussian function would increase from 1 Å to 4 Å in integer steps. $\rho^{FESD}(\mathbf{r}_s)$ was determined according to eq 21, and integration of the partial surface area of the fructose portion was performed yielding the related partial transfer free energy. Results are summarized in Table 5. We chose the parameter *a* such as to go from a very localized representation (turning point of the Gaussian function at 1 Å, $a = 0.5$ Å^{-2}) to a more delocalized picture allowing greater influence from more distant patches. As one can see from Table 5 the correlation coefficient has its maximum when the turning point of the Gaussian function is chosen around 2 Å. The resulting representation of the MolFESD on the molecular surface is presented in Figure 4. Global information is completely conserved, since surface integration of the MolFESD yields the log *P* value or the transfer free energy.

Since the fit quality for the continuous representation is comparable to the simple, patch-based approach, we conclude that both the continuous and discrete representations bear a consistent description of the concept of local hydrophobicity. The results obtained from both descriptions with respect to

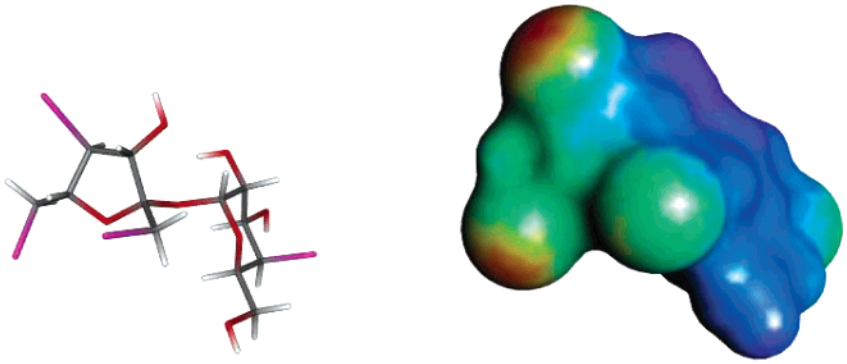

Figure 4. The molecular surface (right) of the 4,1′,6′-trichloro-4,1′,4′,6′-tetradeoxy-4′-iodo-*galacto*-sucrose (left) is shown. Color coding is used to represent the MolFESD on the surface (yellow − positive values, lipophilic areas; blue − negative values, hydrophilic areas). The Gaussian parameters for this representation is $a = 0.125$ Å^{-2}.

the application to sweetness recognition are in good agreement with our previous model,[19] where local information was absent for the model parametrization. Obviously, the MolFESD gains tremendous plausibility in terms of interpreting local, fragmental contributions when constructed from our final model given in eq 14. We are able to assess the mutual influences of neighboring substructures on the molecular surface using a normalized Gaussian function to distribute local, discrete information, i.e., the patch-based f_i values. Apparently, with the turning point of the gauss function set to 2 Å, the contributions of f_i values further away than 5 Å is below 5% thus yielding an optimal correlation for the application described above.

For the applied parametrization strategy the empirical character is notably reduced since the 3D-FED is in principle accessible through statistical-thermodynamic theory. In this work, we use empirical molecular interaction fields (derived from *Grid*) as a model for the 3D-FED and employ experimental data for a proper scaling of model parameters. Finally, we arrive at a model that allows a surface-based description of local hydrophobicity from 3D thermodynamic information, while at the same time consistency with experimental log *P* values is ensured.

Additionally, dependence on partial, atomic log *P* values[11] has become obsolete for our approach. The independence of previously determined fragmental log *P* contributions marks a significant departure from many other empirical works. We are optimistic that our approach and its enhanced plausibility in terms of local contributions can be efficiently used in QSAR studies, in particular for rational drug design where local aspects of the binding interface are of great importance.

CONCLUSION

Based on the *three-dimensional free energy density* (3D-FED) a parametrization strategy for the transfer free energy was presented. This strategy has previously been established[21] providing excellent results when applied to the prediction of hydration free energies. The interaction part of the 3D-FED is approximated by an appropriate molecular interaction field generated by the program *Grid*, while cavity contributions are modeled as a linear combination of surface and volume proportional terms. The parameters for the interaction part of the 3D-FED are related to the molecular structure allowing the fast prediction of transfer free energies by means of a group contribution method. The final model equation is derived considering both local (from the 3D-FED data) and global (experimental log *P* values) thermodynamic information *simultaneously*. For a set of 400 compounds with known experimental log *P* values, good correlation was found with $R^2 = 0.95$ and a *rms* error of 0.82 kcal/mol.

This log *P* model provides a convenient starting point for a physically motivated derivation of the established MolFESD. From the 3D-FED we find surface-based parameters capturing this local information and acting as sampling points when a continuous distribution (the MolFESD) is determined on the surface. Consequently, the MolFESD concept provides the representation of local hydrophobicity on the solvent accessible surface in a quantitative manner. The incorporation of local information to the parametrization procedure marks the essential feature of our strategy allowing for reasonable estimates of partial thermodynamic properties.

To further validate the presented model, a correlation was established between the partial free energy surface density taken from incremental surfaces of a series of sucrose derivatives and their relative sweetness. Comparison of the results to an earlier model showed improvement thereby rendering justification of the presented approach.

ACKNOWLEDGMENT

The authors like to thank Peter Goodford for many valuable comments on the use of the program *Grid*. We are also grateful to F. Javier Luque for presenting us material prior to publication and to Jens Sadowski for providing us experimental log *P* values. We also thank Bernd Schilling for many helpful discussions. This work has been supported by the Deutsche Forschungsgemeinschaft, Bonn.

REFERENCES AND NOTES

(1) Blokzijl, W.; Engberts, J. B. F. N. Hydrophobic effects: opinion and fact. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1545−1579.
(2) Scheraga, H. A. Theory of hydrophobic interactions. *J. Biomol. Struct. Dyn.* **1998**, *16*, 447−460.
(3) Lazaridis, T. Solvent reorganization energy and entropy in hydrophobic hydration. *J. Phys. Chem. B* **2000**, *104*, 4964−4979.
(4) Ruelle, P.; Kesselring, U. The Hydrophobic Effect. 1. A Consequence of the Mobile Order in H-Bonded Liquids. *J. Pharm. Sci.* **1998**, *87*, 987−997.
(5) Besseling, N. A. M.; Lyklema, J. Molecular Thermodynamics of Hydrophobic Hydration. *J. Phys. Chem. B* **1997**, *101*, 7604−7611.
(6) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales. *J. Phys. Chem. B* **1999**, *103*, 4570−4577.
(7) Hummer, G. Hydrophobic Force Field as a Molecular Alternative to Surface-Area Models. *J. Am. Chem. Soc.* **1999**, *121*, 6299−6305.

(8) Sangster, J. *Octanol−Water Partition Coeffients: Fundamentals and Physical Chemistry*; John Wiley & Sons: Chichester, 1997; Vol. 2.
(9) Leo, A. J.; Hansch, C. Role of hydrophobic effects in mechanistic QSAR. *Perspect. Drug Discovery Des.* **1999**, *17*, 1−25.
(10) *Star list Database*; Biobyte Corporation; Pomona, CA, 1997.
(11) (a) Buchwald, P.; Bodor, N. Octanol−water partition: searching for predictive models *Curr. Med. Chem.* **1998**, *5*, 353−380. (b) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.
(12) (a) Mannhold, R.; Rekker, R. F.; Sonntag, C.; ter Laak, A. M.; Dross, K.; Polymeropoulos, E. E. Comparative evaluation of the predictive power of calculation procedures of molecular lipophilicity. *J. Pharm. Sci.* **1995**, *84*, 1410−1419. (b) Mannhold, R.; Rekker, R. F. The hydrophobic fragmental constant approach for calculating log P in octanol/water and aliphatic hydrocarbon/water systems *Perspect. Drug Discovery Des.* **2000**, *18*, 1−18.
(13) Oprea, T. I.; Waller, C. L. Theoretical and Practical Aspects of Three-Dimensional Quantitative Structure Activity Relationships. In *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1997; Vol. 11, pp 127−182.
(14) Vedani, A.; McMasters, D. R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: reducing the bias associated with ligand alignment. *Quant. Struct.-Act. Relat.* **2000**, *19*, 149−161.
(15) Waller, C. L.; Kellogg, G. E. Adding Chemical Information to CoMFA Models with Alternative 3D QSAR Fields. *Network Sci*. [Electronic Publications] **1996**, *2*, (URL: http://www.netsci.org/Science/Compchem/feature10.html).
(16) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 545−552.
(17) Carrupt, P.-A.; Testa, B.; Gaillard, P. Computational Approaches to Lipophilicity: Mehtods and applications. In *Reviews in Computational Chemistry*, Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1997; Vol. 11, pp 241−315.
(18) Pixner, P.; Heiden, W.; Merx, H.; Möller, A.; Moeckel, G.; Brickmann, J. Empirical Method for the Quantification and Localization of Molecular Hydrophobicity. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309−1319.
(19) Jäger, R., Schmidt, F.; Schilling, B.; Brickmann, J. Localization and quantification of hydrophobicity: The molecular free energy density (MolFESD) concept and its application to the sweetness recognition. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 631−646.
(20) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Anlaysis of ALOGP and CLOGP methods. *J. Phys. Chem. A* **1998**, *102*, 3762−3772.
(21) Jäger, R.; Kast, S. M. Fast prediction of hydration free energies from molecular interaction fields. *J. Mol. Graph. Modell.* **2001**, *20*, 123−131.
(22) No, K. T.; Kim, S. G.; Cho, K.-H.; Scheraga, H. A. Description of hydration free energy density as a function of molecular physical properties. *Biophys. Chem.* **1999**, *78*, 127−145.
(23) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*, 2nd ed.; Academic Press: San Diego, 1990; Chapter 6.7.
(24) Luque, F. J., Barril, X. and Orozco, M., Fractional description of free energies of solvation. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 139−152.
(25) Connolly, M. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709−713.
(26) Goodford, P. Multivariate characterization of molecules for QSAR analysis. *J. Chemometrics* **1996**, *10*, 107−117.
(27) Best, S. A.; Merz, K. M., Jr.; Reynolds, C. H. Free energy pertubation study of octanol/water partition coefficients: comparison with continuum GB/SA calculations. *J. Phys. Chem B* **1999**, *103*, 714−726.
(28) DeBolt, S. E.; Kollman, P. A. Investigation of Structure, Dynamics, and Solvation in 1-Octanol and Ist Water-Saturated Solution: Molecular Dynamics and Free-Energy Perturbation Studies. *J. Am. Chem. Soc.* **1995**, *117*, 5316−5340.
(29) Ben-Naim, A.; Lovett, R. Solvation free energy of a hard sphere solute in a square well solvent as a function of solute size. *J. Phys. Chem. B* **1997**, *101*, 10535−10541.
(30) Shimizu, S.; Ikeguchi, M.; Nakamura, S.; Shimizu, K. Size dependence of transfer free energies: A hard-sphere-chain-based formalism. *J. Chem. Phys.* **1999**, *110*, 2971−2982.
(31) Vitha, M. F.; Carr, P. W. The Chemical Meaning of the Standard Free Energy of Transfer: Use of the van der Waals' Equation of State To Unravel the Interplay between Free Volume, Volume Entropy, and the Role of Standard States. *J. Phys. Chem. B* **2000**, *104*, 5343−5349.
(32) Mannhold, R.; Cruciani, G.; Dross, K. and Rekker, R. Multivariate analysis of experimental and computational descriptors of molecular lipophilicity. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 573−581.
(33) Southhall. N. T.; Dill, K. A. The Mechanism of Hydrophobic Solvation Depends on Solute Radius. *J. Phys. Chem. B* **2000**, *104*, 1326−1331.
(34) Shallenberger, R. S.; Acree, T. E. Molecular theory of sweet taste. *Nature* **1967**, *216*, 480−482.
(35) Kier, L. B. Molecular theory of sweet taste. *J. Pharm. Sci.* **1972**, *61*, 1394−1397.
(36) Lichtenthaler, F. W.; Immel, S.; Kreis, U. Evolution of the structural representation of sucrose[1]. *Starch* **1991**, *43*, 121−132.
(37) Lee, C. K. The chemistry and biochemistry of the sweetness of sugars. *Adv. Carbohydr. Chem. Biochem.* **1987**, *45*, 199−351.

CI025576H