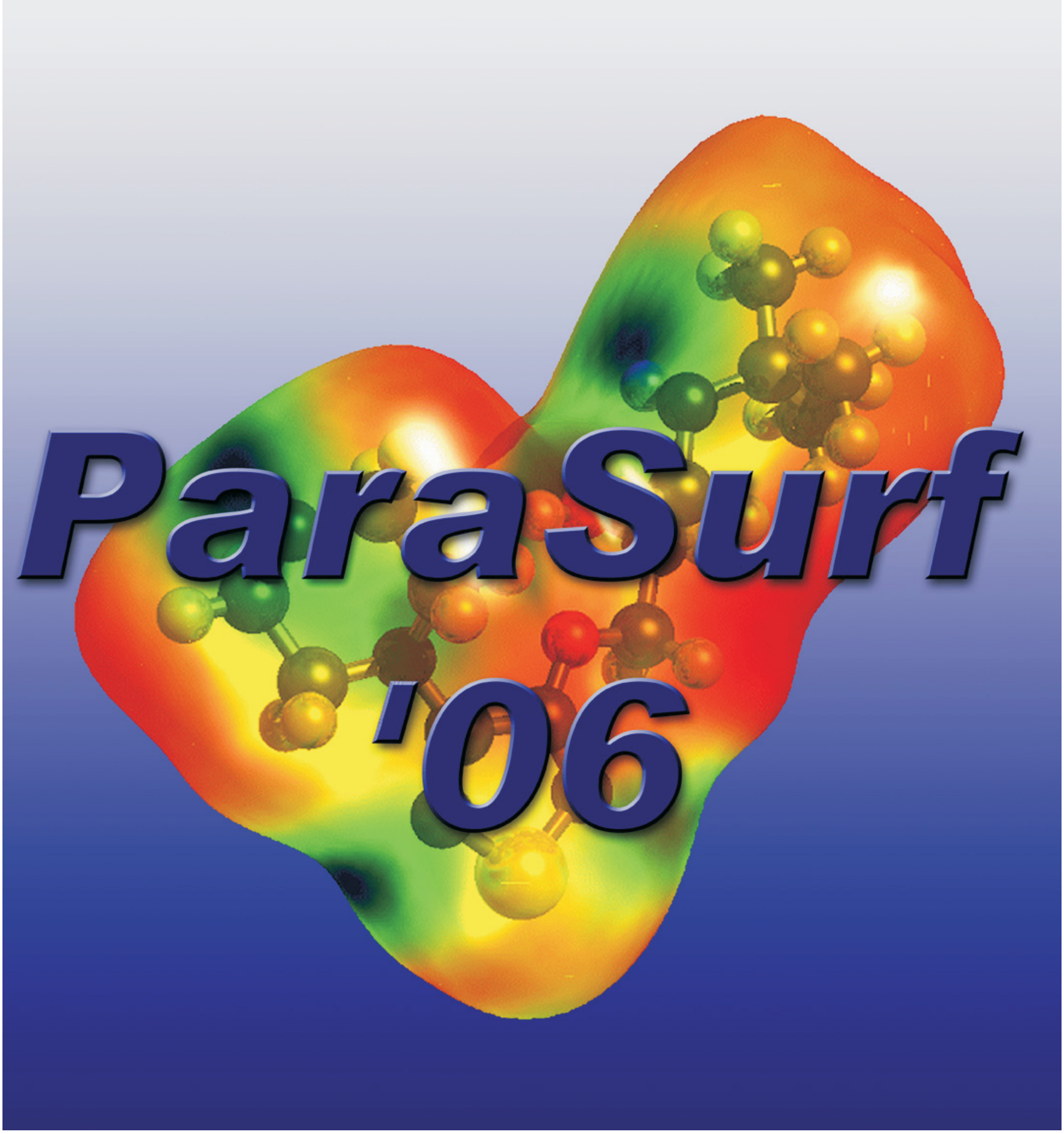
CEPOS
InSilico



ParaSurf '06

Users´ Manual

26 Brookfield Gardens Ryde
ISLE OF WIGHT PO33 3NP
www.ceposinsilico.com

Manual Tim Clark

Software Tim Clark
Jr-Hung Lin
Anselm H. C. Horn

Layout **www.eh-bitartist.de**

TABLE OF CONTENTS

PROGRAM HISTORY

Release Date	Version	Platforms
1st July 2005	ParaSurf´05 initial release (Revision A1)	32-bit Windows 32-bit Linux Irix
1st January 2006	ParaSurf´05 Revision B1 (customer-feedback release)	
1st July 2006	ParaSurf´06 Revision A1	32-bit Windows 32-bit Linux 64-bit Linux Irix

PROGRAM HISTORY

1 INTRODUCTION

ParaSurf™ is a program to generate isodensity or solvent-excluded surfaces from the results of semiempirical molecular orbital calculations, either from VAMP [1] or a public-domain version of MOPAC modified and made available by Cepos InSilico. [2] The surface may be generated by shrink-wrap [3] or marching-cube [4] algorithms and the former may be fit to a spherical harmonic series. [5] The principles of these two techniques are explained below, but for comparison figure 1 shows default isodensity surfaces calculated by ParaSurf™ for a tetracycline derivative. The surfaces are color-coded according to the electrostatic potential at the surface.

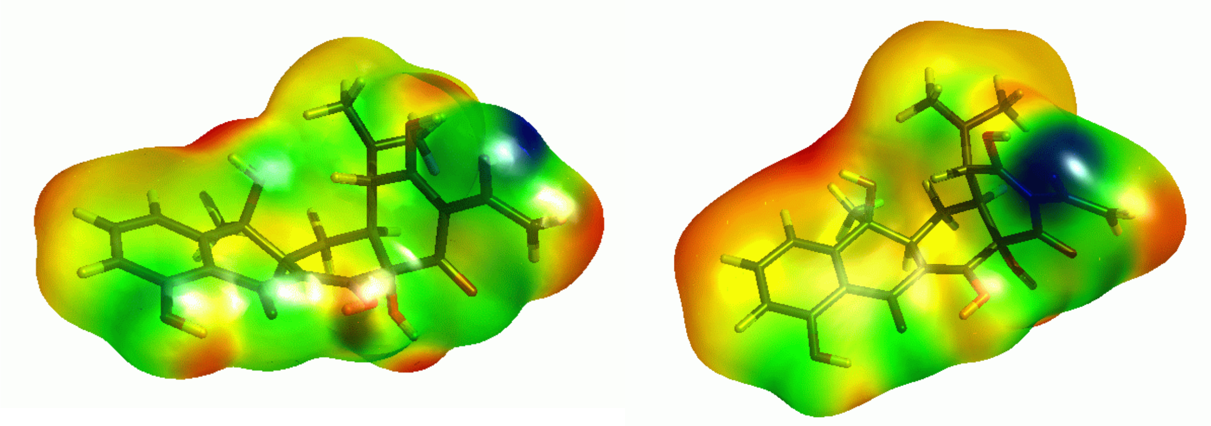


Figure 1: Marching-cube (left) and shrink-wrap (right, fitted to a spherical-harmonic approximation) isodensity surfaces calculated with ParaSurf™ using the default settings.

Four local properties, the molecular electrostatic potential (MEP), [6] the local ionization energy (IE_L), [7] the local electron affinity (EA_L), [8] and the local polarizability (α_L) [8] are calculated at the points on the surface. Two further properties, the local hardness (η_L), [8] and the local electronegativity (χ_L) [8] can be derived from IE_L and EA_L.

The local properties can be used to generate a standard set of 40 descriptors [9] appropriate for quantitative structure-property relationships (QSPRs) for determining physical properties.

ParaSurf™ can also generate local enthalpies and free energies of solvation [10] and integrate them over the entire molecular surface to give the enthalpy or free energy of solvation. ParaSurf™ can read so-called *Surface-Integral Model* (SIM) files that allow it to calculate properties such as, for instance, the enthalpy and free energy of hydration and the free energies of solvation in *n*-octanol and chloroform. The surface-integral models are expressed as summations of local solvation energies over the molecular surface. These local solvation energies can be writted to the ParaSurf™ surface file.

ParaSurf™ is the first program to emerge from the ParaShift project, a collaboration between the Universities of Erlangen, Portsmouth, Southampton, Oxford and Aberdeen. It is intended to provide the molecular surfaces for small molecules (i.e. non-proteins) for subsequent quantitative structure-activity relationship (QSAR), QSPR, high-throughput virtual screening (HTVS), docking and scoring, pattern-recognition and simulation software that will be developed in the ParaShift project.

1.1 Isodensity surfaces

Isodensity surfaces [11] are defined as the surfaces around a molecule at which the electron density has a constant value. Usually this value is chosen to approximate the van der Waals' shape of the molecule. ParaSurf™ allows values of the isodensity level down to 0.00001 $e^{-}Å^{-3}$. Lower values than this may result in failures of the surface algorithms for very diffuse surfaces.

1.2 Shrink-wrap surface algorithm

Shrink-wrap surface algorithms [3] are used to determine single-valued molecular surfaces. Single-valued in this case means that for any given radial vector from the center of the molecule the surface is only crossed once (vectors **A** and **B** in figure 2) and not multiply (vectors **C** and **D** in figure 2):

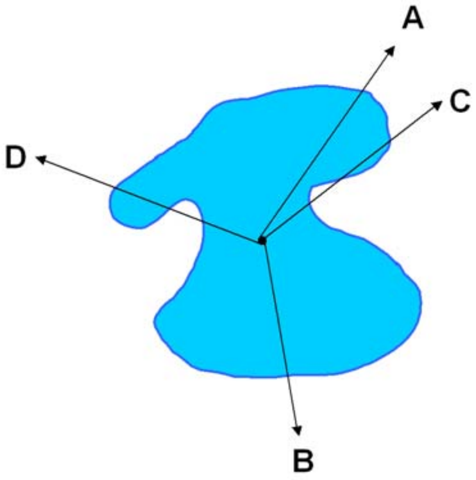


Figure 2: 2D-representation of a molecular surface with single-valued (**A** and **B**) and multiply valued (**C** and **D**) radial vectors from the center

Single-valued surfaces are necessary for spherical-harmonic fitting (see **1.4**). Thus, spherical-harmonic fitting is only available for shrink-wrap surfaces in ParaSurf™. The shrink-wrap algorithm works by starting outside the molecule (point **a** in **figure 3**) and moving inwards along the radial vector until it finds the surface (in our case defined by the predefined level of the electron density, point **b** in **figure 3**). Thus, the shrink-wrapped surface may contain areas (marked by dashed lines in **figure 3**) for which the surface deviates from the true isodensity surface.

These areas of the surface, however, often have little consequence as they are situated above indentations in the molecule that are poorly accessible to solvents or other molecules. The shrink-wrapped surfaces generated by ParaSurf™ should normally be fitted to a spherical-harmonic series for use in HTVS, similarity, pattern-recognition or high-throughput docking applications. The default molecular center in Parasurf™ is the center of gravity (CoG). In special cases in which the CoG lies outside the molecule, another center may be chosen.

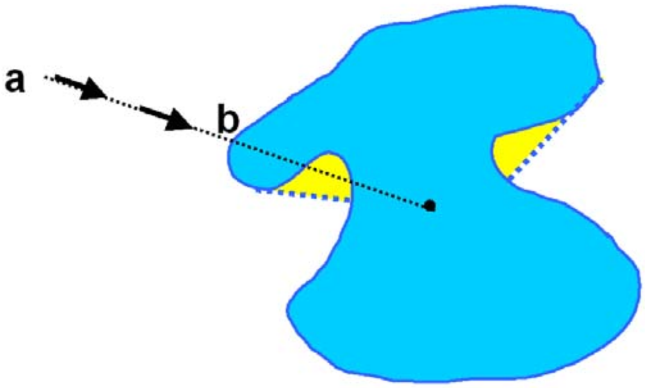


Figure 3: 2D-representation of the shrink-wrap algorithm. The algorithms scans along the vector from point **a** towards the center of the molecule until the electron density reaches the preset value (point **b**). The algorithm results in enclosures (marked yellow) for multi-valued radial vectors.

Figure 4 shows a spherical-harmonically fitted shrink-wrap surface for a difficult molecule. The areas shown schematically in figure 3 are clearly visible.

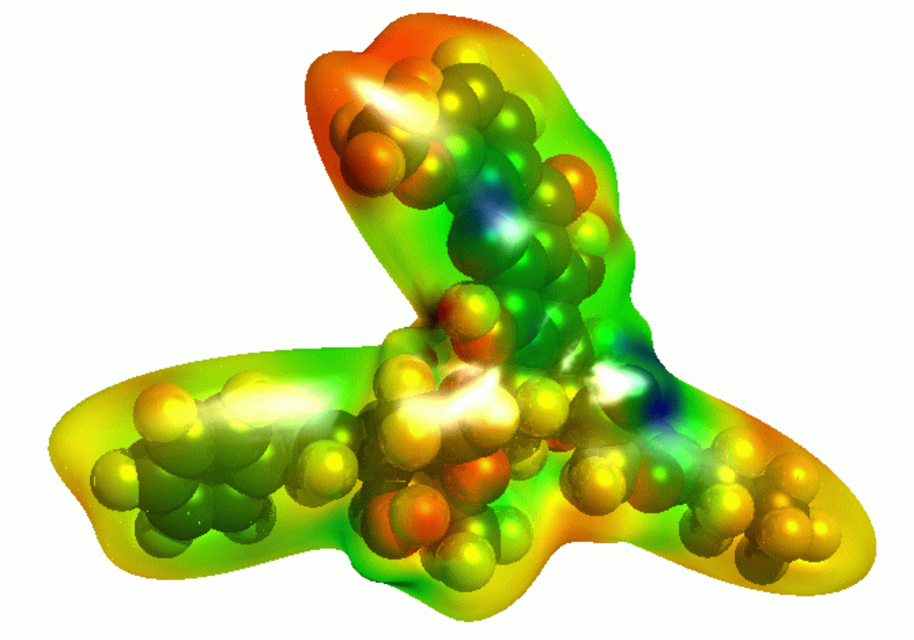


Figure 4: Spherical-harmonic approximation of a shrink-wrap isodensity surface. Note the areas where the surface does not follow the indentations of the moecule.

1.3 Marching-cube algorithm

The marching-cube algorithm [4] implemented in Parasurf™ does not have the disadvantage of being single-valued like the shrink-wrap surface. It cannot, therefore, be fitted to a spherical harmonic series and is used as a purely numerical surface primarily for QSPR applications or surface-integral models. [10] The algorithm works by testing the electron density at the corners of cubes on a cubic lattice laid out through the molecular volume. The corners are divided into those “inside” the molecule (i.e. with a higher electron density than the preset value) and those “outside”. The surface triangulation is then generated for each surface cube and the positions of the surface points corrected to the preset electron density.

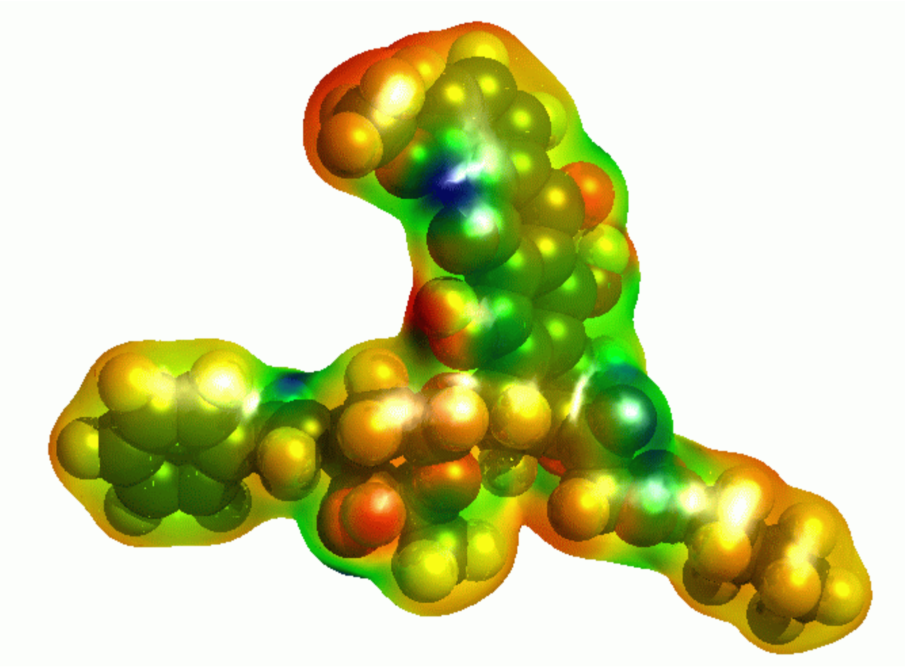


Figure 5: Marching-cube isodensity surface for the molecule shown in figure 4. This surface is better suited for QSPR and surface-integral models

1.4 Spherical-harmonic fitting

Complex surfaces can be fitted to spherical harmonic series to give analytical approximations of the surface. [5] The surfaces are fit to a series of distances $r_{\alpha,\beta}$ from the center along the radial vector defined by the angles α and β as:

$$r_{\alpha,\beta} = \sum_{l=0}^{N} \sum_{m=-l}^{l} c_l^m Y_l^m \qquad (1)$$

Where the distances $r_{\alpha,\beta}$ are linear combinations of spherical harmonics Y_l^m defined as:

$$Y_l^m(\alpha,\beta)=\sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}P_l^m(\cos\alpha)e^{im\beta} \tag{2}$$

where P_l^m *(cos α)* are associated Legendre functions and l and m are integers such that $-l \leq m \leq l$. In the above form, spherical harmonics are complex functions. Duncan and Olson [12] have used the real functions

$$Y_l^m(\alpha,\beta)=N_{lm}P_l^m(\cos\alpha)\cos|m|\beta \tag{3}$$

where N_{lm} are normalization factors, to describe molecular surfaces using spherical harmonics.

Parasurf™ not only fits the surface itself (i.e. the radial distances) to spherical harmonic expansions, but also the four local properties (see 1.5). In this way, a completely analytical description of the shape of the molecule and its intermolecular binding properties is obtained. [13] This description can be truncated at different orders l depending on the application and the precision needed. Thus, a simple description of the molecular properties (shape, MEP, IE_L, EA_L and α_L) to order 2 consists of only five sets of nine coefficients each, or 45 coefficients. These coefficients can be rotated, overlaps calculated etc. [5] to give fast scanning of large numbers of compounds.

Note that, because of the approximate nature of the spherical-harmonic fits, the default isodensity level for the shrink-wrapped surface (0.0003 e^-Å^{-3}) is lower than that (0.008 e^-Å^{-3}) appropriate for an approximately van der Waals' surface using the marching-cube algorithm. The lower value avoids the surface coming too close to atoms. Note also that the fits are incremental, which means that the order chosen for a given application can be obtained by ignoring coefficients of higher order in the spherical-harmonic series.

In some cases, the default resolution of the molecular surface does not allow fitting the spherical-harmonic expansion to very high orders without introducing noise ("ripples") on the fitted surface. In this case, the calculated RMSD becomes larger at higher orders of the spherical-harmonic expansion. ParaSurf'06 recognizes this condition and truncates the fitting procedure at the optimum value. This can be recognized in the output because the RMSD for later cycles remains constant and the coefficients of the higher order spherical harmonics are all zero. This guarantees the optimum fit in each case and is important for applications that use either the spherical-harmonic coefficients themselves or the hybridization coefficients.

1.5 Local Properties

The local properties calculated by Parasurf™ are those related to intermolecular interactions. Local properties, sometimes inaccurately called fields in QSAR work, are properties that vary in space around the molecule and therefore have a distribution of values at the molecular surface. The best known and most important local property in this context is the molecular electrostatic potential, which governs Coulomb interactions, but the MEP only describes a part of the intermolecular interaction energy, so that further local properties are needed.

1.5.1 Molecular electrostatic potential

The MEP is defined in Parasurf™ as the energy of interaction of a single positive electronic charge at the position ***r*** with the molecule. Within quantum mechanical (semiempirical or *ab initio* molecular orbital (MO) theory, density functional theory (DFT)) the MEP (V(***r***)) is described [6] as:

$$MEP(\mathbf{r}) = \sum_{i=1}^{n} \frac{Z_i}{|\mathbf{R_i} - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \tag{4}$$

where n is the number of atoms in the molecule, Z_i is the nuclear charge of atom i located at $\mathbf{R_i}$ and $\rho(\mathbf{r})$ is the electron-density function of the molecule. This expression, however, involves integrating the electron density, a time-consuming calculation. Parasurf™ therefore uses two different approximate models for calculating the MEP.

1.5.1.1 The natural atomic orbital/PC (NAO-PC) model

The NAO-PC model [14,15] uses a total of nine point charges, one positive charge at the nucleus and eight negative ones distributed around it, to describe the electrostatics of a non-hydrogen atom with a valence-only *s*- and *p*-basis set for the semiempirical Hamiltonians MNDO, [16] AM1 [17] and PM3. [18] The negative charges are located at the charge centers of each lobe of the natural atomic orbitals, which are obtained by diagonalizing the one-atom blocks of the density matrix. [14] The NAO-PC charges are calculated by VAMP and output in the .sdf file for use in Parasurf™. The NAO-PC model is therefore only available when using Parasurf™ with VAMP .sdf input. NAO-PC charges are also not available for semiempirical Hamiltonians such as MNDO/d [19] or AM1* [20] that use *d*-orbitals in the basis set.

1.5.1.2 The multipole model

The integrals needed to evaluate **equation (4)** in MNDO-type methods use a multipole approximation [16,19] that extends to quadrupoles. We can therefore also use this approximation to calculate atom-centered monopoles, dipoles and quadrupoles for each atom in the molecule. [21] This multipole model is applicable to all methods, including those with *d*-orbitals, and can be used with MOPAC output files as input to Parasurf™.

1.5.2 Local ionization energy, electron affinity, hardness and electronegativity

The local ionization energy $IE_L(\mathbf{r})$ is defined [7] as a density-weighted Koopmans' ionization potential at a point $\mathbf{r}$ near the molecule:

$$IE_L(\mathbf{r}) = \frac{-\sum_{i=1}^{HOMO} \rho_i(\mathbf{r})\varepsilon_i}{\sum_{i=1}^{HOMO} \rho_i(\mathbf{r})} \qquad (5)$$

where $HOMO$ is the number of the highest occupied MO, $\rho_i(\mathbf{r})$ is the electron density at point $\mathbf{r}$ due to MO i and ε_i is its Eigenvalue. The local ionization energy describes the tendency of the molecule to interact with electron acceptors (Lewis acids) in a given region in space. [7, 8]

The definition of the local electron affinity is a simple extension of **equation (5)** to the virtual MOs: [8]

$$EA_L(\mathbf{r}) = \frac{-\sum_{i=LUMO}^{norbs} \rho_i(\mathbf{r})\varepsilon_i}{\sum_{i=LUMO}^{norbs} \rho_i(\mathbf{r})} \qquad (6)$$

The local electron affinity is the equivalent of the local ionization energy for interactions with electron donors (Lewis bases). [8]

Two further, less fundamental local properties have been defined. [8] These are the local hardness, η_L:

$$\eta_L = \frac{(IP_L - EA_L)}{2} \qquad (7)$$

and the local electronegativity, χ_L:

$$\chi_L = \frac{(IP_L + EA_L)}{2} \qquad (8)$$

1.5.3 Local polarizability

Within the NDDO, the molecular electronic polarizability is easily accessible using the parameterized version [22] of the variational technique introduced by Rivail, [23] which can also be partitioned into an additive polarizability scheme. [24] This allows us to define the local polarizability, α_L, at a point near the molecule as

$$\alpha_L(\mathbf{r}) = \frac{\sum_{j=1}^{norbs} \rho_j^1(\mathbf{r}) q_j \bar{\alpha}_j}{\sum_{j=1}^{norbs} \rho_j^1(\mathbf{r}) q_j} \qquad (9)$$

where q_j is the Coulson occupation and $\bar{\alpha}_j$ the isotropic polarizability attributed to atomic orbital *j*. The density ρ_j^1 is defined as the electron density at the point in question due to an exactly singly occupied atomic orbital *j*. The sum is now over atomic orbitals, rather than MOs as for the other local properties. Thus, the local polarizability is a simple occupation-weighted sum of the orbital polarizabilities in which the contribution of each AO is determined by the density of the individual AO at the point being considered.

1.6 Descriptors

A set of 40 molecular descriptors derived from the MEP, local ionization energy, IE_L, electron affinity, EA_L, electronegativity, χ_L, hardness, η_L, and polarizability, α_L has been defined for QSPR-studies. [9] These and several related descriptors calculated and output by ParaSurf™ are defined in the following table.

Table 1: The descriptors calculated by Parasurf™.

Descriptor	Description	Formula/Reference
μ	Dipole moment	
μ_D	Dipolar density	[25]
α	Molecular electronic polarizabilty	[24]
MW	Molecular weight	
G	Globularity	[26]
A	Molecular surface area	
VOL	Molecular volume	
V_{max}	Maximum (most positive) MEP	[27]
V_{min}	Minimum (most negative) MEP	[27]
$\bar{V}_+$	Mean of the positive MEP values	[27]
$\bar{V}_-$	Mean of the negative MEP values	[27]
$\bar{V}$	Mean of all MEP values	[27]
ΔV	MEP-range	[27]
σ_+^2	Total variance in the positive MEP values	[27]
σ_-^2	Total variance in the negative MEP values	[27]

Descriptor	Description	Formula/Reference
σ^2_{tot}	Total variance in the MEP	[27]
ν	MEP balance parameter	[27]
$\sigma^2_{tot}\nu$	Product of the total variance in the MEP and the balance parameter	[27]
$IE_L^{\max}$	Maximum value of the local ionization energy	
$IE_L^{\min}$	Minimum value of the local ionization energy	
$\overline{IE_L}$	Mean value of the local ionization energy	$\overline{IE_L} = \frac{1}{N}\sum_{i=1}^{N} IE_L^i$
ΔIE_L	Range of the local ionization energy	$\Delta IE_L = IE_L^{\max} - IE_L^{\min}$
σ^2_{IE}	Variance in the local ionization energy	$\sigma^2_{IE} = \frac{1}{N}\sum_{i=1}^{N}\left[IE_L^i - \overline{IE_L}\right]^2$
$EA_L^{\max}$	Maximum of the local electron affinity	
$EA_L^{\min}$	Minimum of the local electron affinity	
$\overline{EA_{L+}}$	Mean of the positive values of the local electron affinity	$\overline{EA_{L+}} = \frac{1}{N^+}\sum_{i=1}^{N^+} EA_{L+}^i$
$\overline{EA_{L-}}$	Mean of the negative values of the local electron affinity	$\overline{EA_{L-}} = \frac{1}{N^-}\sum_{i=1}^{N^-} EA_{L-}^i$
$\overline{EA_L}$	Mean value of the local electron affinity	$\overline{EA_L} = \frac{1}{N}\sum_{i=1}^{N} EA_L^i$
ΔEA_L	Range of the local electron affinity	$\Delta EA_L = EA_L^{\max} - EA_L^{\min}$
σ^2_{EA+}	Variance in the local electron affinity for all positive values	$\sigma^2_{EA+} = \frac{1}{m}\sum_{i=1}^{m}\left[EA_i^+ - \overline{EA^+}\right]^2$
σ^2_{EA-}	Variance in the local electron affinity for all negative values	$\sigma^2_{EA-} = \frac{1}{n}\sum_{i=1}^{n}\left[EA_i^- - \overline{EA^-}\right]^2$
σ^2_{EAtot}	Sum of the positive and negative variances in the local electron affinity	$\sigma^2_{EAtot} = \sigma^2_{EA+} + \sigma^2_{EA-}$
ν_{EA}	Local electron affinity balance parameter	$\nu_{EA} = \frac{\sigma^2_{EA+}\cdot\sigma^2_{EA-}}{\left[\sigma^2_{EA}\right]^2}$
δA^+_{EA}	Fraction of the surface area with positive local electron affinity	$\delta A^+_{EA} = \frac{A^+_{EA}}{A}$, A = total surface area
$\overline{\chi_L}$	Mean value of the local electronegativity	$\overline{\chi_L} = \frac{1}{N}\sum_{i=1}^{N}\chi_L^i$
$\alpha_L^{\max}$	Maximum value of the local polarizability	

Descriptor	Description	Formula/Reference
$\alpha_L^{\min}$	Minimum value of the local polarizability	
$\overline{\alpha_L}$	Mean value of the local polarizability	$\overline{\alpha_L} = \frac{1}{N}\sum_{i=1}^{N}\alpha_L^i$
$\Delta\alpha_L$	Range of the local polarizability	$\Delta\alpha_L = \alpha_L^{\max} - \alpha_L^{\min}$
σ_α^2	Variance in the local polarizability	$\sigma_\alpha^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\alpha_L^i - \overline{\alpha_L}\right]^2$

1.7 Surface-integral models

The surface-integral models that can be calculated by ParaSurf™ are defined [10] using the expression

$$P = \sum_{i=1}^{ntri} f\left(V^i, IE_L^i, EA_L^i, \alpha_L^i, \eta_L^i\right) \cdot A^i \tag{10}$$

where P is the target property, usually a free energy, f is a non-linear function of the electrostatic potential V, the local ionization energy, IE_L, the local electron affinity, EA_L, the local polarizability, α_L and the local hardness, η_L. A^i is the area of the surface triangle i.

The molecular property P is printed to the output file and to the `<filename>_p.sdf` ParaSurf™ output SD-file. The individual values of the function f are added to the list of local properties written for each surface point to the .psf file if the surface details are output.

The surface-integral models themselves are not implemented directly in ParaSurf™, but are read in general form from the SIM file, whose format is given in 3.9. Thus, the users' own surface-integral models can be added to ParaSurf™. Data for generating surface-integral models can be derived simply from the .psf surface output for a normal ParaSurf™ run. Note that the program options given in the SIM file must be the same for all the models included in the file and that they override conflicting command-line options.

1.8 Spherical Harmonic "Hybrids"

Once the molecular shape or a local property have been fitted to a spherical-harmonic expansion, [11] the shape or property can be described succinctly as a series of spherical-harmonic "hybridization" coefficients analogous to the concept of hybrid atomic orbitals. Thus, for each value of l in **equation (1)** the "hybridization" coefficient H_l is given by:

$$H_l = \sum_{i=-m}^{m} \left(c_l^m\right)^2 \tag{11}$$

The hybridization coefficients H_l can be used as additional descriptors for fast QSPR screening.

1.9 Descriptors and moments based on surface-integral models

ParaSurf'06™ uses local properties defined in a surface-integral model (SIM, **see 1.7**) to calculate descriptors analogous to those listed in **Table 1**. Additionally, “dipolar moments of the local property are calculated. These are gauge-independent moments calculated by first shifting values of the local property so that their sum is zero and then calculating moments according to

$$\mu = \sum_{i=1}^{ntri} P_i \mathbf{r_i} \tag{12}$$

where μ is the dipolar moment, P_i the value of the local property i situated at position $\boldsymbol{r_i}$.

The output for these properties derived from a SIM for $\log P_{OW}$ is shown below:

```
Descriptors calculated for logP:

 Dipolar moment      x:   -549.2     y:   -247.9     z:   -937.0
                                                   Sum:    1114.
 Most positive value  :      1.407
 Most negative value  :      0.8325E-01
 Range                :      1.324
 Mean                 :      0.1874
 Mean positive        :      0.1874
 Mean negative        :      0.000
 Total variance       :      0.2376E-01
 Positive variance    :      0.2376E-01
 Negative variance    :      0.000
 Balance parameter    :      0.000
 Balance*variance     :      0.000
```

The values of these descriptors are often useful for deriving models directly related to the property modeled by the SIM. Note that no units are given in the output because they depend on the property modeled by the SIM.

1.10 Shannon entropy

The information content at the surface of the molecule can be defined based on the distribution of the four local properties over the surface using an approach analogous to that introduced by Shannon. [28]

Shannon defined the Shannon entropy, H, which corresponds to the amount of information (in bits) as

$$H = -\sum_{i=1}^{n} p_i \log_2 \left(p_i \right) \tag{13}$$

where n is the number of possible characters and p_i is the probability that character i will occur. Note that, importantly, this definition of the amount of information is local (*i.e.* it only depends on the value of the probability of character i).

For a continuous property, X, **equation (1)** becomes

$$H = -\int_{-\infty}^{\infty} p(X) \log_2 p(X) dX \tag{14}$$

If we now assume that the Shannon entropy at a point in space near a molecule is defined by the values of the four continuous local properties described above, we obtain

$$H = -\iint\iint p\left(V,I,E,\alpha\right) \log_2 \left(V,I,E,\alpha\right) dVdIdEd\alpha \tag{15}$$

where $p\left(V,I,E,\alpha\right)$ is the probability of finding the values V,I,E and α. However, we can simplify this expression because the four properties are essentially independent of each other, [8,9] so that we can write

$$\begin{aligned} H = &-\int p\left(V\right) \log_2 p\left(V\right) dV - \int p\left(I\right) \log_2 p\left(I\right) dI \\ &-\int p\left(E\right) \log_2 p\left(E\right) dE - \int p\left(\alpha\right) \log_2 p\left(\alpha\right) d\alpha \end{aligned} \tag{16}$$

Transferring this definition to a molecule for which a triangulated surface of k triangles, where triangle i has area A_i and average values of the four local properties V_i, I_i, E_i and α_i we obtain

$$H = -\sum_{i=1}^{k} \left[p(V_i) \log_2 p(V_i) + p(I_i) \log_2 p(I_i) + p(E_i) \log_2 p(E_i) + p(\alpha_i) \log_2 p(\alpha_i) \right] \cdot A_i \tag{17}$$

where $p(X_i)$ is the probability that the value X_i of the property X, where X may be V, I, E or α, will occur.

ParaSurf’06™ offers two alternatives as sources for the probabilities $p(X_i)$. The first, known as the “external” Shannon entropy, is to use probabilities taken from an external dataset and defined in a separate statistics file. The default "external" statistics file is called *bins.txt* and is read from the ParaSurf root directory. The statistics defined in *bins.txt* were derived from AM1 calculations of all the bound ligands defined in the PDBbind database [**29**] in their correct protonation states and at geometries obtained by optimizing with AM1 starting from the bound conformation. [**30**]

Alternatively, the user can define a custom "external" statistics file using the ParaSurf module *binner* (**see appendix**). The "external" Shannon entropy is useful for relating a series of molecules to each other, but is sensitive, for instance, to the total charge of the molecule.

The "internal" Shannon entropy is calculated using probabilities determined from the surface properties of the molecule itself, and therefore corresponds more closely to Shannon's classical definition than the "external" Shannon entropy and the probabilities used are individual for each molecule. The "internal" Shannon entropy can be considered to represent the information content of the molecule. The properties of the two types of Shannon entropy will be described in a forthcoming paper. [**30**]

1.11 Surface autocorrelations

Gasteiger et al. [**31**] introduced the concept of surface autocorrelations as powerful descriptions of molecular binding properties for quantitative structure-activity relationships (QSARs). In ParaSurf'06™, autocorrelations *A(R)* are defined as:

$$A(R) = \frac{1}{ntri} \sum_{i=1}^{ntri} \sum_{j=i+1}^{ntri} \omega_{ij} \, e^{-\sigma\left(R - r_{ij}\right)^2} \tag{18}$$

where r_{ij} is the distance between surface points *i* and *j* and ω_{ij} is a function of one or more local properties at the points *i* and *j*. The smoothing factor σ determines the steepness of the exponential function.

Four different autocorrelation functions are calculated by ParaSurf'06™. These are:

Shape autocorrelation	$\omega_{ij} = 1.0$	
Plus-plus MEP autocorrelation (V1)	$\omega_{ij} = V_i \times V_j$ $\omega_{ij} = 0.0$	$(V_i > 0$ and $V_j > 0)$ $(V_i < 0$ or $V_j < 0)$
Minus-minus MEP autocorrelation (V1)	$\omega_{ij} = V_i \times V_j$	$(V_i < 0$ and $V_j < 0)$
Plus-minus MEP autocorrelation (V2)	$\omega_{ij} = -V_i \times V_j$ $\omega_{ij} = 0.0$	$(V_i \times V_j < 0)$ $(V_i \times V_j > 0)$

Autocorrelation functions based on the other three local properties correlate very strongly with the shape autocorrelation and are therefore not calculated.

ParaSurf'06™ calculates autocorrelations as vectors of *A(R)* values 128 elements long starting at an *R*-value of 2.5 Å and increasing in steps of 0.06 Å (i.e. up to a maximum value of 10.12 Å). **Figure 6** shows the four autocorrelation functions for trimethoprim calculated with AM1.

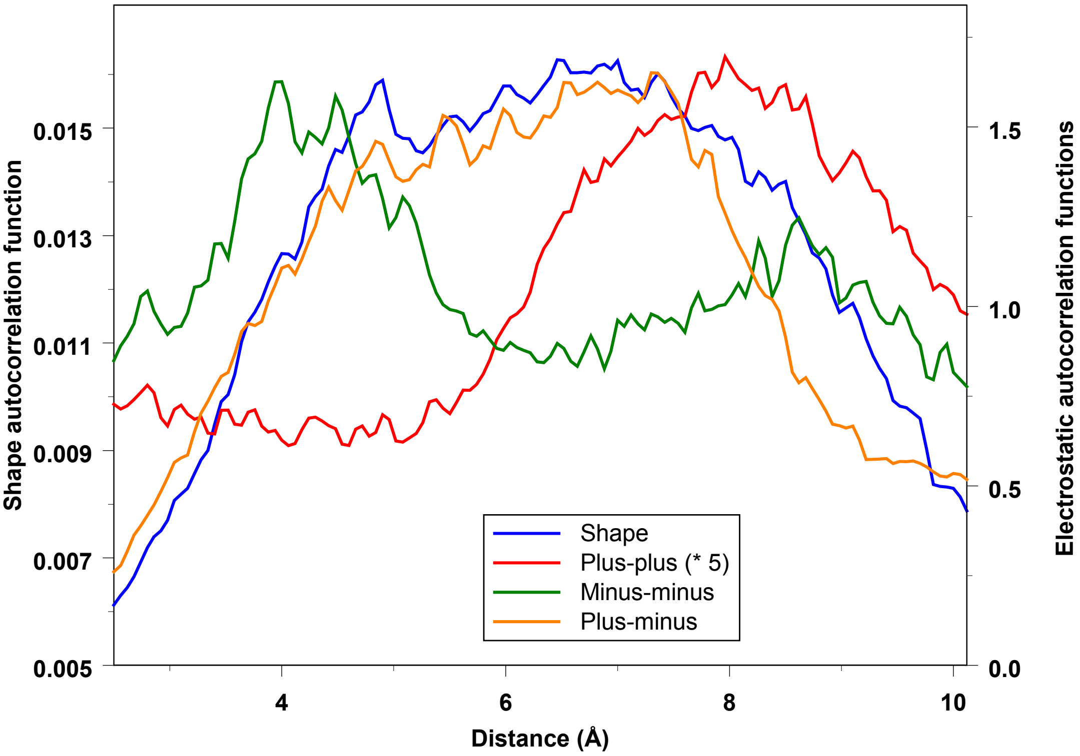


Figure 6: The four autocorrelation functions calculated using the AM1 Hamiltonian for trimethoprim.

The command-line argument **`autocorr=<filename>`** requests that similarities in the autocorrelation functions with the molecule described in **`<filename>`**, where **`<filename>`** must be a **`ParaSurf'06™.sdf`** output file. The similarities *S* are defined as:

$$S = \frac{1}{N} \sum_{i=1}^{N} \frac{2 \cdot \min\left(A_1\left(R_i\right), A_2\left(R_i\right)\right)}{\left(A_1\left(R_i\right) + A_2\left(R_i\right)\right)} \qquad (19)$$

where $A_1\left(R_i\right)$ is the value of the autocorrelation function for molecule 1 at distance R_i etc. To avoid division by zero, the summation ignores values of *i* for which the sum $A_1\left(R_i\right) + A_2\left(R_i\right)$ is zero. *N* is therefore the number of points within the defined range of *R* for which this sum is non-zero.

These similarities are calculated for the entire range of each of the three autocorrelation functions and also for the first, second, third and fourth quartal of the distance range for each of the autocorrelation functions (i.e. 2.5-4.42 Å, 4.426-6.34 Å, 6.346-8.26 Å and 8.266-10.12 Å). These individual similarities can be written to a table file **(see 3.11)** and are printed in the output file **(see 3.4.4).**

2 PROGRAM OPTIONS

ParaSurf™ program options are given as command-line arguments. Arguments are separated by blanks, so that no single argument may contain a blank character. Arguments may be written in any combination of upper and lower case. The options are:

Table 2: ParaSurf™ command-line options

`<name>`		Base name for the input file (must be the first argument)
		
		Using this option, the input file is assumed to be
		if a file with this name exists.
		
		Otherwise the file
		will be used as input.
		
		If neither of these files are found, the program will use an .sdf file written by the Cepos version of Mopac 6.
		These files are called
		
		The output files are
			
			`<name>_v.sdf`
			
			`<name>.sdf`
			
			`<name>_m.sdf`
			
			`<name>_p.out`
			`<name>_p.sdf`
			`<name>.psf` (optional)
			`<name>.asd` (optional)
			`<name>_p.vmp` (optional)

`surf=`	`wrap` `cube`	Shrink-wrap surface (default) Marching-cube surface	
`contour=`	`isoden` `solvex`	The surface is defined by the electron density A solvent-excluded surface is used (only active for `surf=wrap` in this version)	

`fit=`	`sphh`
	`isod`
	`none`
		Spherical-harmonic fitting (default for `surf=wrap`)
		Smooth to preset isodensity value (default for `surf=cube`)
		No fitting	

`iso=`	*`n.nn`*	Isodensity value set to `n.nn` $e^{-}\text{Å}^{-3}$ (default for shrink-wrap surface = `0.0002`; default for marching-cube surface = `0.003`; minimum possible value = `0.00001`)	
`estat=`	`naopc` `multi`	Use NAO-PC electrostatics Use multipole electrostatics (default)	
`psf=`	`on` `off`	Write .psf surface file Do not write .psf surface file (default)	
`asd=`	`on` `off`	Write anonymous SD (.asd) file Do not write .asd file (default)	

`vmp=`	`on`	Write .vmp file for debugging. Map the MEP onto the surface
	`off`	Do not write .vmp file (default)
	`mep`	Write .vmp file for debugging. Map the MEP onto the surface
	`iel`	Write .vmp file for debugging. Map IE$_L$ onto the surface
	`eal`	Write .vmp file for debugging. Map EA$_L$ onto the surface
	`pol`	Write .vmp file for debugging. Map α_L onto the surface
	`har`	Write .vmp file for debugging. Map η_L onto the surface
	`eng`	Write .vmp file for debugging. Map χ_L onto the surface
	`<MOD>`	Write .vmp file for debugging. Map the local property with the three-character designator `<MOD>` defined in the SIM file onto the surface
`grid=`	`<filename>`	Read the Cartesian coordinates at which to calculate a grid of the four properties (MEP, IE$_L$, EA$_L$, α_L). See 3.8
`sim=`	`<filename>`	One or more surface-integralmodels will be read from the file `<filename>.sim` in the ParaSurf™ executable directory. `<filename>` can be upper or lower case or any mixture but must be exactly three characters long.

`center=`
or
`centre=`
	`on`	The atomic and surface coordinates in the `.psf` output file will be centered for calculations that use spherical-harmonic fitting. Note that this means that the atomic coordinates in the `SDF`-output file (which are the input coordinates) will be different to those in the `PSF`-output file. This option is default.

	`off`	The atomic and surface coordinates in the `.psf` output file will not be centered and will correspond to the input coordinates and those in the `SDF`-output file.
`shannon=`	`<filename>`	Requests that Shannon entropies (both internal and external) be calculated. If no statistics file `<filename>` is given, the default file (`bins.txt` in the ParaSurf Root directory) will be used. If a statistics file is given that either does not exist, contains errors or is derived from ParaSurf runs using different options to the current one, only the internal Shannon entropy is calculated.
`autocor=`	`<filename>`	Requests that the surface autocorrelation functions be calculated and written to the output .sdf file. `<filename>` must be a ParaSurf™ output `.sdf file` that contains the autocorrelation functions. In this case, similarities between the two molecules will be calculated and printed (see also `aulist=`).
`table=`	`<filename>`	An ASCII table of the ParaSurf™ descriptors will be written to the file `<filename>`. If `<filename>` exists, the values for the current molecule will be appended to the existing table, otherwise the file will be created.
`aulist=`	`<filename>`	An ASCII table of the calculated autocorrelation similarities will be written to the file `<filename>`. If `<filename>` exists, the values for the current molecule will be appended to the existing table, otherwise the file will be created.

Examples:

```
parasurf test surf=wrap fit=sphh iso=0.03 psf=on estat=naopc
```

Use the input file `est_v.sdf`, `test.sdf` or `test_m.sdf` to calculate a shrink-wrap surface with an isodensity value of 0.03 e^{-}Å^{-3}, perform a spherical-harmonic fit, use NAO-PC electrostatics and write the spherical-harmonic coefficients to `test_P.sdf` and the entire surface to `test_P.psf`.

```
parasurf test surf=cube fit=none
```

Use the file `test_v.sdf`, `test.sdf` or `test_m.sdf` as input to perform a marching-cube surface determination without fitting and to calculate the descriptor set.

3 INPUT AND OUTPUT FILES

ParaSurf™ uses the following files for input and output:

Table 3: ParaSurf™ input and output files

File	Name	Description

Input	`<filename>_v.sdf`
	or
	`<filename>.sdf`
	(if available) or
	`<filename>_m.sdf`
		VAMP .sdf file output. VAMP must be run with the **`ALLVECT`** option to be able to calculate all the properties. The VAMP version used must be able to calculate AO-polarizabilities.
		If no VAMP .sdf file is found, ParaSurf™ defaults to a Cepos Mopac 6 .sdf file. It is strongly recommended to use the EF option for geometry optimizations in Mopac.

Hamiltonian	`Vhamil.par`	The VAMP parameters file (found in the VAMP executable directory). This file must be copied to the ParaSurf™ executable directory.
Output	`<filename>_p.out`	Always written.
SD-file	`<filename>_p.sdf`	Always written.
ASD-file	`<filename>.asd`	Anonymous SD-file. Requested by the option `asd=on`
PSF-file	`<filename>.psf`	ParaSurf™ surface file. Requested by the option `psf=on`
VMP-file	`<filename>_p.vmp`	Debug file.
SIM-file	`<filename>.sim`	Surface-integral model definition. `<filename>` must have exactly three characters and the file must reside in the ParaSurf™ executable directory.
Descriptor table file	`User defined`	An ascii, comma-separated file that contains a line of descriptors for each molecule. This file will be created if it does not exist or an extra line will be appended if it does exist.
Autocorrelation similarity file	`User defined`	An ascii, comma-separated file that contains a line of autocorrelation similarities for each molecule. This file will be created if it does not exist or an extra line will be appended if it does exist.

3.1 The VAMP .sdf file as input

VAMP .sdf files, an extension of the MDL .sdf file format, [32] are the primary communication channel between VAMP and ParaSurf™. The atomic coordinates and bond definitions are given in the MDL format as shown in figure 7.The remaining fields are indicated by tags with the form:

`<FIELD_NAME>`

FIELD_NAME is a predefined text tag used to locate the relevant data within the .sdf file.

Only the important fields for a ParaSurf™ calculation will be described here:

```
1-Bromo-3,5-difluorobenzene
OMVAMP81A04250313563D 1   0.00000     0.00000     0

 12 12  0  0  0  0              1 V2000
   -2.6274    0.2410    0.0003 F
   -1.2738    0.2410    0.0003 C
   -0.5810    1.4623    0.0003 C
    0.8231    1.4389    0.0003 C
    1.5096    2.6055    0.0004 F
    1.5266    0.2198    0.0001 C
    0.8142   -0.9793    0.0001 C
    1.7431   -2.6055   -0.0004 Br
   -0.5805   -0.9840    0.0002 C
   -1.1264    2.4167   -0.0003 H
    2.6274    0.2339    0.0003 H
   -1.1515   -1.9253    0.0001 H
  1  2  1
  2  3  4
  3  4  4
  4  5  1
  4  6  4
  6  7  4
  7  8  1
  2  9  4
  7  9  4
  3 10  1
  6 11  1
  9 12  1
M  END
```

Figure 7: The headers and titles, atomic coordinates and bond definitions from a VAMP .sdf file. The format follows the MDL definition. [29]

<HAMILTONIAN>

The Hamiltonian field defines the semiempirical Hamiltonian (model and parameters) used for the calculation. The Hamiltonian must be defined for ParaSurf™ to be able to calculate the electrostatics and the local polarizabilities. NAO-PC electrostatics and the local polarizability are not available for all methods. Quite generally, the multipole electrostatics model is to be preferred over the NAO-PC model, which can only be used if the VAMP .sdf file contains a block with the tag:

<NAO-PC>

NAO-PCs cannot be calculated for methods with *d*-orbitals. The local polarizability calculation has not yet been extended to these methods, but will be in a future release.

The following table gives an overview of the methods and their limitations:

Table 4: Hamiltonians and the available electrostatic and polarizability models.

Hamiltonian	Reference	Electrostatics		Local Polarizability
		NAO-PC	Multipole	
MNDO	[16]	YES	YES	YES
AM1	[17]	YES	YES	YES
PM3	[18]	YES	YES	YES
MNDO/c	[32]	YES	YES	NO
MNDO/d	[19]	NO	YES	NO
AM1*	[20]	NO	YES	NO

<VAMPBASICS>

The VAMPBASICS block contains the following quantities (`FORTRAN format 6f13.6`):

Heat of Formation	kcal mol^{-1}
HOMO energy	eV
LUMO energy	eV
Dipole moment	
x-component	Debye
y-component	Debye
z-component	Debye

<TOTAL COULSON CHARGE>

The total charge of the molecule.

<DENSITY MATRIX ELEMENTS>

The DENSITY MATRIX ELEMENTS block contains the one-atom blocks of the density matrix for the non-hydrogen atoms. For an *sp*-atom, there are ten elements, for an *spd*-atom 45. The squares of the diagonal elements for hydrogen atoms are included in the **`<CHARGE ON HYDROGENS>`** block that follows the density matrix. The density-matrix elements are used in ParaSurf™ to calculate the local properties and are essential.

<ORBITAL VECTORS>

The ORBITAL VECTORS block contains the MO-eigenvecxtors and related information and is essential for calculating the local properties. VAMP must be run with the keyword **`ALLVECT`** in order to write all the MO vectors to the SDF file.

The entire SDF input file is echoed to the **`<filename>_p.sdf`** output file and the properties calculated by ParaSurf™ are added in additional blocks at the end.

3.1.1 Multi-structure SD-files

ParaSurf'06™ can read SD-files containing more than one molecule (e.g. those produced by the VAMP-QSAR model engine) and process them in one run. The command-line arguments apply to each molecule in the SD-file and the same semiempirical Hamiltonian must be used for each molecule or an error message will be printed and the program terminated.

As part of this enhancement, ParaSurf'06 Revision A1™ can use SD-files that do not contain the one-atom blocks of the density matrix explicitly. Thus, SD-files that only contain the molecular-orbital Eigenvectors and Eigenvalues give full ParaSurf'06 functionality within the previous restrictions that:

- Polarizabilities are not yet available for Hamiltonians that use *d*-orbitals (MNDO/d and AM1*).
- NAO-PC electrostatics are only available if the NAO-PCs are present in the SD-file. Multipole electrostatics are available for all Hamiltonians.

The output SD-file written by ParaSurf™ also contains multiple molecules as in the input file. Other ParaSurf™ output files (.asd, .vmp etc.) are also concatenated.

Multiple SD-files can be used with a SIM file exactly as single molecules.

3.2 The Cepos MOPAC 6.sdf file as input

Cepos Mopac 6 writes an .sdf file containing the above blocks with the exception that the MOPACBASICS block replaces VAMPBASICS. No additional keywords are required to request the correct .sdf output for ParaSurf™.

3.3 The Vhamil.par file

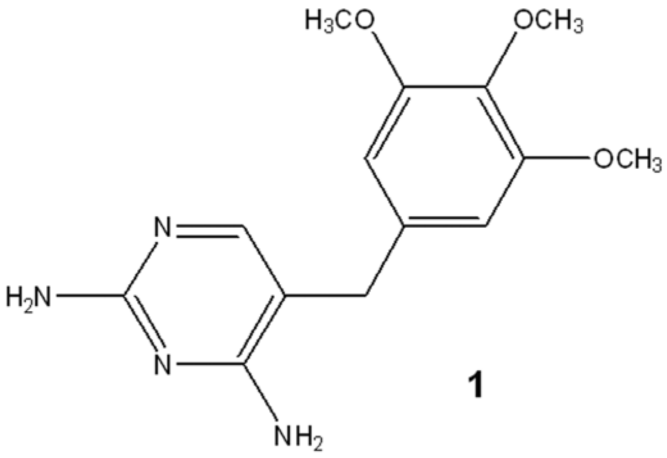
The file Vhamil.par is used by VAMP to define the available Hamiltonians and elements and supply the parameters. This file is also used by ParaSurf™ for the same purpose. A Vhamil.par file for standard Hamiltonians and elements is supplied with the ParaSurf™ program. In order to be sure that all Hamiltonians and elements available to VAMP can also be handled by ParaSurf™, however, the Vhamil.par file from the VAMP executable directory should be copied into the ParaSurf™ executable directory.

3.4 The ParaSurf™ output file

The ParaSurf™ output file provides the user with information about the calculation and the results. It is, however, not intended as the primary means of communication between ParaSurf™ and other programs. Thus, the essential information contained in the output file is also available from the ParaSurf™ output .sdf file.

3.4.1 For a spherical-harmonic surface

Figure 8 shows the output for a calculation using the options `surf=wrap fit=sphh` for trimethoprim, **1**.



1

```
<> ParaSurf'06, Revision A1
 <> Copyright (c) 2006, Cepos InSilico Ltd. All rights
reserved.

 <> Input = trimethoprim.sdf

 <> Program options :

    Using shrink-wrap isocontour surface
    Fitting surface to spherical harmonics
    Using an isodensity surface contour
    Isodensity value = 0.2000E-04 electrons/Angstrom**3
    Using multipole electrostatics
```

Figure 8: ParaSurf® output for trimethoprim, **1**, using a spherical-harmonic surface.

```
<> AM1   calculation for  Trimethoprim
 <> Fitting surface to spherical harmonics
 <> Order(l)   RMSD

       0       1.95263386
       1       2.00052993
       2       1.57640277
       3       1.13683010
       4       0.95340420
       5       0.74540464
       6       0.66926274
       7       0.58942331
       8       0.54355194
       9       0.51760442
      10       0.50656564
      11       0.48555581
      12       0.46284202
      13       0.45019006
      14       0.44009732
      15       0.43261027

 <> Spherical harmonic fit for MEP:
 <> Order(l)   RMSD

       0      11.98388200
       1      12.05977747
       2       8.94196171
       3       7.10750676
       4       5.65374058
       5       4.72443217
       6       4.22936592
       7       3.40471958
       8       3.00459770
       9       2.70025262
      10       2.37948183
      11       2.08585451
      12       1.89670665
      13       1.67167599
      14       1.55026696
      15       1.35608719
      16       1.18594737
      17       0.99808358
      18       0.89576936
      19       0.82750580
      20       0.76423977
```

Figure 8: continued

```
<> Spherical harmonic fit for IE(l):
 <> Order(l)   RMSD

       0      56.43504813
       1      51.09884711
       2      45.60899136
       3      44.66107072
       4      40.40371330
       5      35.94498069
       6      33.23502674
       7      27.59565679
       8      23.61792252
       9      20.28010728
      10      18.90880404
      11      17.11492314
      12      15.40818538
      13      14.15504388
      14      13.02995375
      15      12.41905890
      16      11.84905676
      17      10.87231348
      18      10.08600883
      19       9.88754170
      20       9.73512067

 <> Spherical harmonic fit for EA(l):
 <> Order(l)   RMSD

       0      12.79556998
       1      12.35517173
       2      11.90899797
       3       9.50286120
       4       8.67783125
       5       7.10627646
       6       6.94283382
       7       6.55691258
       8       6.16378320
       9       5.67694852
      10       5.33536823
      11       5.03567792
      12       4.76494701
      13       4.15529924
      14       3.74615266
      15       3.57989712
      16       3.21344347
      17       3.08789530
      18       2.83823821
      19       2.67737233
      20       2.50100870
```

Figure 8: continued

```
<> Spherical harmonic fit for Alpha(l):
 <> Order(l)   RMSD

       0       0.02314245
       1       0.01698452
       2       0.01434977
       3       0.01198993
       4       0.00986702
       5       0.00906865
       6       0.00841069
       7       0.00796426
       8       0.00751901
       9       0.00702683
      10       0.00649850
      11       0.00607376
      12       0.00531069
      13       0.00512745
      14       0.00481518
      15       0.00472809
      16       0.00458103
      17       0.00428100
      18       0.00419188
      19       0.00413632
      20       0.00410800

 <> Property ranges:
    Density   :   0.2773E-05 to   0.1368E-03
    IE(l)     :       391.01 to       669.50
    EA(l)     :      -108.57 to       -38.17
    MEP       :       -44.47 to        15.29
    Alpha(l)  :       0.2372 to       0.3375
```

Figure 8: continued

```
<> Descriptors :

    Dipole moment        :       1.2467 Debye
    Dipolar density      :   0.1890E-02 Debye.Angstrom**-3
    Molecular pol.       :     128.5408 Angstrom**3
    Molecular weight     :       290.32
    Globularity          :       0.7705
    Total surface area   :       475.65 Angstrom**2
    Molecular volume     :       659.71 Angstrom**3

    Most positive MEP    :        15.29 kcal/mol
    Most negative MEP    :       -44.47 kcal/mol
    Mean +ve MEP         :         5.26 kcal/mol
    Mean -ve MEP         :       -12.63 kcal/mol
    Mean MEP             :        -4.45 kcal/mol
    MEP range            :        59.75 kcal/mol
    MEP +ve Variance     :        11.08 kcal/mol
    MEP -ve Variance     :       107.42 kcal/mol
    MEP total variance   :       118.50 kcal/mol
    MEP balance parameter:       0.0848
    MEP balance*variance :      10.0464 kcal/mol

    Maximum IE(l)        :       669.50 kcal/mol
    Minimum IE(l)        :       391.01 kcal/mol
    Mean IE(l)           :       478.95 kcal/mol
    IE(l) range          :       278.49 kcal/mol
    IE(l) variance       :      3184.77 kcal/mol


    Maximum EA(l)        :       -38.17 kcal/mol
    Minimum EA(l)        :      -108.57 kcal/mol
    Mean +ve EA(l)       :         0.00 kcal/mol
    Mean -ve EA(l)       :       -93.51 kcal/mol
    Mean EA(l)           :       -93.51 kcal/mol
    EA(l) range          :        70.39 kcal/mol
    EA(l) +ve variance   :         0.00 kcal/mol
    EA(l) -ve variance   :       160.81 kcal/mol
    EA(l) total variance :       160.81 kcal/mol
    EA(l) balance param. :       0.0000
    Fraction pos. EA(l)  :       0.0000 ( =    0.00 Angstrom**2)

    Mean eneg(l)         :       385.44 kcal/mol

    Maximum alpha(l)     :   0.3375     Angstrom**3
    Minimum alpha(l)     :   0.2372     Angstrom**3
    Mean alpha(l)        :   0.2831     Angstrom**3
    Alpha(l) range       :   0.1003     Angstrom**3
    Variance in alpha(l) :   0.5276E-03 Angstrom**3
```

Figure 8: continued

```
<> Spherical-Harmonic Hybridization:

    Shape hybrids          :
   17.720383    1.137854    3.465499    2.855941    1.400140    1.618042
    0.739926    0.695812    0.470178    0.385822    0.382195    0.318993
    0.238589    0.203220    0.213801    0.188207

    MEP hybrids            :
   12.652756    4.616972   25.325368   18.030298   13.236618   10.466215
    6.582408    7.740968    4.274007    4.477487    4.010893    3.271562
    2.394703    2.521394    1.619796    1.978157    1.824135    1.810851
    1.389207    0.902731    0.854934

    IE(l) hybrids          :
   1696.4489     92.7699     78.6441     63.0086     59.2524     53.9940
     52.5311     54.0722     40.0586     44.7139     23.4674     24.2601
     19.7150     17.2182     16.8091     13.8518     13.5376     14.8130
     11.8342     12.0944     10.8561

    EA(l) hybrids          :
    325.4167      6.2664     12.0332     25.0295     13.7078     16.7659
      8.5380      9.3463      7.8402      7.4368      6.2604      5.1981
      5.1017      7.1861      5.4323      4.1630      4.4779      3.6446
      3.2960      2.9795      2.8237

    Alpha(l) hybrids       :
  1.01346091  0.05267412  0.03336797  0.02955989  0.02150438  0.01504621
  0.01164927  0.00870566  0.00742029  0.00844974  0.00751906  0.00587211
  0.00688754  0.00489269  0.00635104  0.00464445  0.00537937  0.00639075
  0.00437520  0.00556036  0.00491094
```

Figure 8: continued

```
<> Atomic surface properties:


   Atom    Area         MEP            IE(l)           EA(l)        mean
                     max    min     max     min     max     min     pol.
 C     1   0.000
 O     2   0.000
 C     3   3.197   -6.36  -43.62  585.24  481.90  -43.59  -91.66   0.310
 C     4   1.110   -6.33  -15.47  586.02  496.37  -71.95  -92.45   0.323
 C     5   0.618   -8.50  -14.47  584.41  533.04  -85.69  -93.35   0.322
 C     6   0.000
 C     7   0.910   -9.48  -15.35  559.51  515.74  -69.07  -91.62   0.319
 C     8   4.094   -1.40  -21.83  584.59  484.64  -51.15  -98.73   0.294
 N     9   3.539  -14.64  -29.38  535.69  458.71  -80.63 -105.08   0.280
 C    10  10.194   -1.76  -27.31  633.23  532.80  -38.86  -87.62   0.284
 N    11   0.000
 N    12   1.379  -14.42  -32.82  539.14  469.91  -78.29 -102.75   0.266
 C    13   6.416   -8.41  -25.74  645.00  514.26  -38.17  -82.91   0.286
 N    14   0.000
 C    15   1.899   -7.13  -15.28  588.19  493.69  -60.68  -92.40   0.316
 C    16   3.923  -10.57  -37.69  575.64  479.10  -38.37  -99.13   0.314
 O    17   0.000
 C    18   0.000
 C    19   6.529  -10.29  -44.47  573.17  465.56  -46.33  -99.90   0.310
 O    20   1.569  -30.66  -41.88  534.29  446.11  -75.31  -93.26   0.247
 C    21   0.000
 H    22  33.709   12.62  -39.38  559.91  405.88  -81.69  -99.73   0.297
 H    23  21.732   13.89  -19.69  562.67  408.42  -82.57  -95.35   0.294
 H    24  24.156   13.92  -31.15  568.19  407.11  -68.85  -95.26   0.292
 H    25   7.010   11.08   -6.40  519.22  425.97  -83.77  -96.24   0.289
 H    26  18.849    7.77   -8.84  534.47  400.18  -88.51 -100.67   0.300
 H    27  17.993    7.31  -24.11  597.65  399.66  -68.44 -100.82   0.302
 H    28  27.541    9.00  -27.14  535.48  413.54  -60.86 -103.75   0.283
 H    29  34.230   15.29  -27.63  669.50  464.72  -65.39 -107.85   0.249
 H    30  33.502   15.16  -28.64  668.15  480.57  -66.40 -107.93   0.242
 H    31  35.285   11.51  -29.34  662.46  473.61  -62.49 -108.57   0.245
 H    32  10.409   11.42  -26.05  643.01  482.53  -73.15  -99.44   0.262
 H    33   5.733    9.44  -14.04  525.92  434.76  -73.49  -96.54   0.295
 H    34  29.960   11.28  -37.39  568.06  406.83  -83.23 -100.30   0.291
 H    35  24.218   11.42  -33.01  567.74  407.68  -69.15  -95.69   0.293
 H    36  23.158   11.37  -12.90  568.25  407.57  -81.12  -99.24   0.295
 H    37  34.521    4.42  -37.89  554.53  394.89  -89.71 -108.49   0.292
 H    38  19.106    4.28  -37.25  550.62  396.52  -86.84 -108.23   0.298
 H    39  26.171    3.92  -26.54  550.83  391.01  -90.02 -107.59   0.296

 Total   472.660

 <> ParaSurf used   38.27 seconds CPU time
```

Figure 8: continued

After printing the program options and fitting the calculated shrink-wrap surface, ParaSurf™ lists the root-mean-square deviations (RMSDs) for the surface points as a function of the order of the spherical-harmonic expansion, first for the geometry of the surface and then for each of the four local properties. The RMSD values give an idea of how well each order of the spherical-harmonic expansion fits the calculated shrink-wrap surface or the relevant property. The highest order used by ParaSurf™ is 15 for the surface itself and 20 for each property.

The descriptor table is then printed. For molecules with no surface areas with positive EA_L, $\sigma^2_{EA_{L+}}$ is set to zero. The descriptors are those described in **table 1**.

The spherical-harmonic hybridization coefficients are then listed for the shape and the four local properties. The coefficients are listed by increasing *l* starting from zero.

The table of atomic surface properties is derived by first finding the atom that contributes most (according to a Coulson analysis) to the electron density for each surface point. The point is then assigned to this atom and the maxima and minima in the MEP, IE_L and EA_L as well as the mean local polarizability for the points assigned to each atom are calculated. Note that, because of the fitting procedure, the values reported in this table may contain spurious values if the fitted surface comes particularly close to an atom (or does not approach it). This situation is generally recognisable from the RMSD values printed for the fit. The surface used to calculate the descriptors and atomic-surface properties is the fitted spherical-harmonic surface of order 15.

3.4.2 For a marching-cube surface

Figure 9 shows the output for a calculation using the options `surf=cube` for trimethoprim.

```
<> ParaSurf'06, Revision A1
 <> Copyright (c) 2006, Cepos InSilico Ltd. All rights
reserved.

 <> Input = trimethoprim.sdf

 <> Program options :

    Using marching-cube isodensity surface
    Surface fitting turned off
    Using an isodensity surface contour
    Isodensity value = 0.3000E-03 electrons/Angstrom**3
    Using multipole electrostatics


 <> AM1   calculation for  Trimethoprim
 <> Number of triangles =  15024
 <> Number of unique points :   7517

 <> Property ranges:
    Density    :   0.2881E-03 to   0.3099E-03
    IE(l)      :       392.35 to       654.76
    EA(l)      :      -109.82 to       -29.09
    MEP        :       -69.88 to        24.82
    Alpha(l)   :       0.2288 to       0.3301
```

Figure 9: ParaSurf™ output for trimethoprim using a marching-cube surface

```
<> Descriptors :

    Dipole moment          :       1.2467 Debye
    Dipolar density        :   0.3155E-02 Debye.Angstrom**-3
    Molecular pol.         :     128.5408 Angstrom**3
    Molecular weight       :       290.32
    Globularity            :       0.7042
    Total surface area     :       369.79 Angstrom**2
    Molecular volume       :       395.13 Angstrom**3

    Most positive MEP      :        24.82 kcal/mol
    Most negative MEP      :       -69.88 kcal/mol
    Mean +ve MEP           :         9.05 kcal/mol
    Mean -ve MEP           :       -18.72 kcal/mol
    Mean MEP               :        -4.94 kcal/mol
    MEP range              :        94.70 kcal/mol
    MEP +ve Variance       :        31.60 kcal/mol
    MEP -ve Variance       :       239.92 kcal/mol
    MEP total variance     :       271.53 kcal/mol
    MEP balance parameter:         0.1028
    MEP balance*variance :        27.9261 kcal/mol

    Maximum IE(l)          :       654.76 kcal/mol
    Minimum IE(l)          :       392.35 kcal/mol
    Mean IE(l)             :       486.30 kcal/mol
    IE(l) range            :       262.41 kcal/mol
    IE(l) variance         :      3584.97 kcal/mol


    Maximum EA(l)          :       -29.09 kcal/mol
    Minimum EA(l)          :      -109.82 kcal/mol
    Mean +ve EA(l)         :         0.00 kcal/mol
    Mean -ve EA(l)         :       -89.08 kcal/mol
    Mean EA(l)             :       -89.08 kcal/mol
    EA(l) range            :        80.74 kcal/mol
    EA(l) +ve variance     :         0.00 kcal/mol
    EA(l) -ve variance     :       276.47 kcal/mol
    EA(l) total variance :         276.47 kcal/mol
    EA(l) balance param. :         0.0000
    Fraction pos. EA(l)    :       0.0000 ( =    0.00 Angstrom**2)

    Mean eneg(l)           :       397.21 kcal/mol

    Maximum alpha(l)       :   0.3301     Angstrom**3
    Minimum alpha(l)       :   0.2288     Angstrom**3
    Mean alpha(l)          :   0.2830     Angstrom**3
    Alpha(l) range         :   0.1013     Angstrom**3
    Variance in alpha(l) :     0.4898E-03 Angstrom**3
```

Figure 9: continued

```
<> Atomic surface properties:


   Atom    Area         MEP            IE(1)          EA(1)        mean
                     max     min     max     min     max     min    pol.
 C     1   0.257  -23.83  -46.81  569.04  546.04  -81.50  -92.65   0.268
 O     2   3.658  -15.50  -69.70  594.61  456.71  -63.66  -81.53   0.269
 C     3   6.490   -7.36  -64.78  643.17  499.72  -30.09  -99.43   0.304
 C     4   2.166   -3.31  -19.76  632.00  493.47  -39.95 -100.75   0.316
 C     5   1.600   -3.37  -18.28  633.56  547.00  -53.42 -100.30   0.313
 C     6   0.000
 C     7   2.042   -4.01  -22.33  605.70  512.44  -49.80  -91.06   0.317
 C     8   5.665    4.11  -28.25  638.27  488.34  -35.80  -88.30   0.288
 N     9   6.693  -19.84  -58.79  571.34  417.76  -54.71 -103.24   0.260
 C    10   9.411   -0.86  -46.23  654.76  543.32  -41.17  -81.09   0.279
 N    11   0.537  -46.53  -53.00  615.72  593.86  -60.28  -78.98   0.276
 N    12   6.122  -16.63  -55.66  571.30  417.23  -51.01  -98.93   0.247
 C    13   7.570  -10.11  -44.07  644.87  527.82  -37.47  -82.80   0.284
 N    14   0.713  -41.89  -57.27  618.93  590.90  -64.88  -81.21   0.287
 C    15   4.127   -8.07  -22.33  640.32  494.13  -30.91 -100.89   0.314
 C    16   5.886  -15.50  -60.73  641.06  507.48  -29.09  -94.41   0.307
 O    17   1.261  -18.70  -69.88  567.04  464.95  -64.27  -87.64   0.252
 C    18   0.289  -16.74  -56.71  573.39  531.54  -74.34  -94.89   0.267
 C    19   5.580  -15.52  -60.03  617.79  492.33  -39.25  -96.42   0.314
 O    20   3.960  -31.04  -63.94  579.08  438.53  -69.08  -94.47   0.265
 C    21   0.543  -26.28  -54.82  563.68  530.83  -90.92 -106.12   0.269
 H    22  20.848   22.24  -43.97  561.11  407.48  -83.45  -97.55   0.297
 H    23  16.018   22.21  -47.65  566.01  408.34  -70.47  -97.22   0.294
 H    24  16.235   22.21  -45.75  567.60  407.90  -66.77  -96.72   0.290
 H    25   7.143   16.35   -8.05  537.67  429.84  -70.41  -97.59   0.288
 H    26  13.545   13.09   -5.83  579.22  401.23  -85.87 -100.66   0.299
 H    27  13.114   11.65  -38.33  611.02  400.48  -74.22 -100.87   0.301
 H    28  17.462   13.34  -29.83  533.62  415.38  -56.04 -100.30   0.282
 H    29  20.093   24.36  -44.05  639.92  488.09  -72.40 -107.69   0.247
 H    30  20.380   24.82  -48.21  644.36  488.15  -70.42 -107.77   0.241
 H    31  20.025   22.73  -51.85  642.47  483.89  -67.45 -108.35   0.245
 H    32  10.792   22.31  -49.42  644.46  478.52  -79.21 -102.13   0.259
 H    33   7.935   15.01  -21.60  523.98  429.34  -65.58  -98.60   0.294
 H    34  20.353   17.51  -31.87  560.07  408.27  -87.93  -99.07   0.290
 H    35  16.221   17.81  -47.09  565.56  408.31  -66.99  -96.49   0.291
 H    36  16.250   17.77  -37.52  557.41  408.14  -69.22  -96.81   0.294
 H    37  20.708    8.06  -41.68  545.33  396.15  -95.89 -109.82   0.294
 H    38  16.217    8.06  -54.42  595.55  394.15  -82.57 -109.54   0.296
 H    39  18.651    7.90  -40.13  586.15  392.35  -74.89 -109.41   0.295

 Total   366.558

 <> ParaSurf used    8.39 seconds CPU time
```

Figure 9: continued

The table of RMSD values is no longer printed and the range of the electron-density values for the surface points (a test for the quality of the surface) is closer to the target isodensity value (in this case 0.0003 e^{-}Å^{-3}) than for the fitted surface. The internal precision used by the program is ± 2% of the target isodensity value. The values of the descriptors and the atomic-surface properties are more consistent using the marching-cube surface and are recommended for QSPR and surface-integral applications.

3.4.3 For a job with Shannon entropy

Figures **10** and **11** show the relevant sections of the output for a calculation using the options surf=cube for trimethoprim with the extra *shannon* option, which requests internal and external Shannon entropies using the default *bins.txt* statistical background file from the PARASURF_ROOT directory. The output is identical to that shown in **figure 8** except that an additional Shannon entropy block is printed after the descriptors, as shown in **figure 9**:

```
                              internal    external
Maximum Shannon H      :       0.4467      0.4786     bits
Angstrom**-2
Minimum Shannon H      :       0.0448      0.1570     bits
Angstrom**-2
Mean Shannon H         :       0.2296      0.3624     bits
Angstrom**-2
Variance Shannon H     :       0.0107      0.0068     bits
Angstrom**-2
Molecular Shannon H    :        86.24      135.04     bits
Information density    :       0.2332      0.3652     bits
Angstrom**-2
```

Figure 10: Shannon entropy section of the ParaSurf® output for trimethoprim, 1, using a marching-cube isodensity surface.

If the statistical background file is not found or does not have the correct format, only the “internal” Shannon entropy appears in this table.

The Shannon entropy is also analyzed based on the surfaces assigned to the individual atoms to give the table shown in **figure 11**:

```
   Shannon-entropy analysis :


                                         Shannon Entropy
                            Internal                        External
  Atom    Area     max     min    mean   total     max     min    mean   total
C     1   0.257  0.2201  0.0790  0.1374  0.0353  0.3125  0.2335  0.2746  0.0707
O     2   3.658  0.2033  0.0587  0.1090  0.3986  0.3492  0.1882  0.2617  0.9572
C     3   6.490  0.1802  0.0583  0.1021  0.6629  0.3454  0.1696  0.2534  1.6442
C     4   2.166  0.1736  0.0780  0.1054  0.2283  0.3700  0.1607  0.2203  0.4773
C     5   1.600  0.1543  0.0847  0.1067  0.1707  0.3462  0.1729  0.2385  0.3815
C     6   0.000
C     7   2.042  0.1729  0.0868  0.1117  0.2282  0.3693  0.1673  0.2290  0.4677
C     8   5.665  0.2372  0.0859  0.1395  0.7901  0.4121  0.2460  0.3178  1.8003
N     9   6.693  0.1664  0.0789  0.1058  0.7084  0.3193  0.2207  0.2687  1.7987
C    10   9.411  0.1532  0.0539  0.1038  0.9767  0.3477  0.2101  0.2768  2.6047
N    11   0.537  0.0855  0.0588  0.0665  0.0358  0.2473  0.2256  0.2351  0.1263
N    12   6.122  0.2108  0.0756  0.1072  0.6566  0.2916  0.2197  0.2479  1.5178
C    13   7.570  0.1873  0.0591  0.1093  0.8277  0.3198  0.2162  0.2721  2.0596
N    14   0.713  0.1746  0.0629  0.1019  0.0727  0.2921  0.2318  0.2724  0.1943
C    15   4.127  0.1749  0.0879  0.1089  0.4492  0.3316  0.1837  0.2344  0.9674
C    16   5.886  0.1718  0.0600  0.0991  0.5832  0.3231  0.1854  0.2415  1.4211
O    17   1.261  0.1353  0.0614  0.0877  0.1105  0.3207  0.1928  0.2303  0.2904
C    18   0.289  0.2010  0.0734  0.1431  0.0414  0.3602  0.2089  0.2725  0.0788
C    19   5.580  0.1590  0.0558  0.0891  0.4971  0.2937  0.1669  0.2086  1.1643
O    20   3.960  0.1842  0.0648  0.0969  0.3836  0.3222  0.2104  0.2630  1.0415
C    21   0.543  0.1637  0.1057  0.1404  0.0762  0.3424  0.2627  0.2993  0.1624
H    22  20.848  0.4039  0.0796  0.3038  6.3337  0.4648  0.2564  0.4181  8.7168
H    23  16.018  0.4239  0.0765  0.3107  4.9767  0.4713  0.2101  0.4125  6.6078
H    24  16.235  0.4248  0.0749  0.3254  5.2820  0.4712  0.2231  0.4158  6.7503
H    25   7.143  0.3288  0.1404  0.2347  1.6761  0.4577  0.3210  0.4113  2.9376
H    26  13.545  0.3942  0.1235  0.2664  3.6089  0.4739  0.3284  0.4407  5.9700
H    27  13.114  0.3249  0.0891  0.2069  2.7131  0.4695  0.2296  0.4208  5.5177
H    28  17.462  0.3825  0.1071  0.2862  4.9975  0.4400  0.3017  0.3967  6.9266
H    29  20.093  0.2782  0.0527  0.1926  3.8688  0.3792  0.2132  0.3098  6.2241
H    30  20.380  0.2756  0.0540  0.1882  3.8360  0.3613  0.2121  0.2954  6.0205
H    31  20.025  0.3054  0.0541  0.2015  4.0361  0.3729  0.2176  0.3069  6.1449
H    32  10.792  0.2809  0.0715  0.1551  1.6743  0.4371  0.2328  0.3175  3.4264
H    33   7.935  0.3473  0.1138  0.2468  1.9586  0.4682  0.2585  0.4191  3.3252
H    34  20.353  0.4467  0.0991  0.3458  7.0389  0.4712  0.2812  0.4326  8.8037
H    35  16.221  0.4453  0.0718  0.3357  5.4460  0.4712  0.2381  0.4174  6.7712
H    36  16.250  0.4390  0.0836  0.3349  5.4415  0.4718  0.2623  0.4245  6.8976
H    37  20.708  0.3625  0.1214  0.2774  5.7443  0.4786  0.2938  0.4325  8.9554
H    38  16.217  0.3297  0.0776  0.2428  3.9373  0.4784  0.2562  0.4243  6.8809
H    39  18.651  0.3467  0.0890  0.2854  5.3230  0.4786  0.2533  0.4290  8.0007

<> ParaSurf used    8.34 seconds CPU time
```

Figure 11: Shannon entropy analysis from the ParaSurf® output for trimethoprim, 1, using a marching-cube isodensity surface.

3.4.4 For a job with autocorrelation similarity

In order to calculate, for instance, the autocorrelation similarities between captopril and trimethoprim, first calculate the reference compound (in this case captopril) and request that the autocorrelation functions be written to the ParaSurf™ SDF-output file:

```
parasurf captopril surf=cube autocorr
```

The calculate the autocorrelations for trimethoprim and their similarities to those of captopril:

```
parasurf trimethoprim surf=cube autocorr=captopril_p.sdf
```

This leads to the following additional output from ParaSurf™:

```
<> Calculating autocorrelation similarities to captopril_p.sdf
<> Lead molecule =  OC(=O)C1CCCN1C(=O)C(C)CS

Similarities :     Shape        +/+         -/-          +/-

  Total      :    0.784647    0.107028    0.393519    0.334957

  1. Quartal :    0.807435   -0.243055    0.266812    0.426333
  2. Quartal :    0.772026   -0.257466    0.352233    0.127342
  3. Quartal :    0.869377    0.458815    0.630442    0.319388
  4. Quartal :    0.689751    0.469819    0.324591    0.466765
```

Figure 12: Similarity output using autocorrelation functions. The lead molecule is captopril, which is defined in captopril_p.sdf using the SMILES string.

Similarities are calculated over the entire distance range (Total) and for each of the four quartals using the four different types of autocorrelation defined in **1.11**. Often, for small molecules, the 4th quartal similarities are unity because the autocorrelations peter out at long range. Either the total similarities or the individual similarities for the quartals can be used for QSAR studies.

3.5 ParaSurf™ SDF-output

The SDF output file (a fixed-format file) contains additional blocks with the information generated by ParaSurf™. These are:

<ParaSurf OPTIONS>

The ParaSurf™ OPTIONS block consists of one line giving the options used in the ParaSurf™ calculation. These are:

`<surface> <fit> <electrostatic model> <isodensity level>` `(a4,2x,a4,2x,a5,2x,f8.3)`

Where the individual variables can be:

<surface>	**WRAP**	Shrink-wrap surface
	CUBE	Marching-cube surface
<fit>	**NONE**	No fitting, unsmoothed marching-cube surface
	ISO	Marching-cube surface corrected to ± 2% of the preset isodensity value
	SPHH	Spherical-harmonic surface fit
<electrostatic model>	**NAOPC**	NAO-PC electrostatics
	MULTI	Multipole electrostatics
<isodensity level>	***n.nn***	The target isodensity value in e^{-}Å^{-3}

<MOLECULAR_CENTERS>

The molecular centers block appears only for calculations that use spherical harmonic fits. It includes two lines of the form:

```
"Spherical harmonic center = ", 3f12.6

"Center of gravity         = ", 3f12.6
```

These blocks give the *x*, *y* and *z* coordinates of the center of the molecule used for the spharical-harmonic fit and the center of gravity, respectively. These two centers are usually identical, but may be different if the center of gravity lies outside the molecule (e.g. for U-shaped molecules).

<SPHERICAL_HARMONIC_......>

The spherical harmonic fits are described in **<SPHERICAL_HARMONIC_…..>** blocks. These blocks all have the same format and vary only in the property described. Each block has the form:

The spherical harmonic fits are described in <SPHERICAL_HARMONIC_…..> blocks. These blocks all have the same format and vary only in the property described. Each block has the form:

Order = *nn*	`("Order = ",i4)`
l (c_l^m) ***m = -l to l***	`(I5, 10f8.4/5x,10f8.4/5x,10f8.4/5x,10f8.4)` (One set of coefficients each for `l` = 1 to 15)

RMSDs:
***l*, *RMSD*^{1}, *RMSD*^{2}**
	`("RMSDs:")`
	`(i8, 2f12.8)`
	(One line for each `l` for `l` = 1 to 15, where RMSD^1 is the area-weighted RMSD and RMSD^2 the simple RMSD)

There are five such blocks, indicated by the tags:

`<SPHERICAL_HARMONIC_SURFACE>`	The fitted molecular surface (radial distances) in Ångstrom
`<SPHERICAL_HARMONIC_MEP>`	The MEP values at the spherical-harmonic surface (`l` = 20) in kcal mol^{-1}
`<SPHERICAL_HARMONIC_IE(l)>`	The IE$_L$ values at the spherical-harmonic surface (`l` = 20) in kcal mol^{-1}
`<SPHERICAL_HARMONIC_EA(l)>`	The EA$_L$ values at the spherical-harmonic surface (`l` = 20) in kcal mol^{-1}
`<SPHERICAL_HARMONIC_ALPHA(l)>`	The α_L values at the spherical-harmonic surface (`l` = 20) in kcal mol^{-1}

<ParaSurf Descriptors>

The ParaSurf™ descriptors block lists the calculated descriptors in the following groups:

Molecular:	μ, μ_D, α, MW, G, A, VOL `("Molecular ",5f10.4,2f10.2)`
MEP:	$V_{\max}$, $V_{\min}$, $\overline{V}_+$, $\overline{V}_-$, $\overline{V}$, ΔV , σ_+^2, σ_-^2, σ_{Tot}^2 , ν , $\sigma_{tot}^2\nu$ `("MEP        ",7f10.2/10x, f10.2,3f10.4)`
IE(l):	$IE_L^{\max}$, $IE_L^{\min}$,$\overline{IE_L}$, ΔIE_L , σ_{IE}^2 `("IE(l)      ",4f10.2,f10.4)`
EA(l):	$EA_L^{\max}$, $EA_L^{\min}$,$\overline{EA_{L+}}$, $\overline{EA_{L-}}$, $\overline{EA_L}$, ΔEA_L , σ_{EA+}^2, σ_{EA-}^2, σ_{EA}^2, ν_{EA}, $\delta \mathrm{A}_{EA}^+$, A_{EA}^+ `("EA(l)      ",7f10.2/2f10.2,2f10.4,f10.2)`
Eneg(l):	$\overline{\chi_L}$ `("Eneg(l)    ",f10.2)`
Alpha(l):	$\alpha_L^{\max}$, $\alpha_L^{\min}$, $\overline{\alpha_L}$, $\Delta\alpha_L$, σ_α^2 `("Alpha(l)   ",5f10.4)`

For calculations using a spherical-harmonic fit, the hybridization coefficients are printed to the .sdf file as follows (tag line followed by as many lines with the coefficients as necessary):

`<SHAPE HYBRIDS>`	`(15 coefficients, 6f12.6)`
`<MEP HYBRIDS>`	`(20 coefficients, 6f12.6`
`<IE(L) HYBRIDS>`	`(20 coefficients, 6f12.2)`
`<EA(L) HYBRIDS>`	`(20 coefficients, 6f12.2)`
`<ALPHA(L) HYBRIDS>`	`(20 coefficients, 6f12.8)`

The hybridization coefficients are listed in order of increasing *l* from zero, exactly as in the output file.

The atomic surface properties are listed in the atomic order according to the following headings (tag line followed by as many lines with the surface properties as necessary):

<ATOMIC SURFACE AREAS>	**Areas**	(10f8.4)
<ATOMIC SURFACE MEP MAXIMA>	**MEP maxima**	(10f8.2)
<ATOMIC SURFACE MEP MINIMA>	**MEP minima**	(10f8.2)
<ATOMIC SURFACE IE(L) MAXIMA>	**IE(l) maxima**	(10f8.2)
<ATOMIC SURFACE IE(L) MINIMA>	**IE(l) minima**	(10f8.2)
<ATOMIC SURFACE EA(L) MAXIMA>	**EA(l) maxima**	(10f8.2)
<ATOMIC SURFACE EA(L) MINIMA>	**EA(l) minima**	(10f8.2)
<ATOMIC SURFACE MEAN POL>	**Mean pol.**	(10f8.4)

The properties correspond exactly to those printed in the table of surface properties in the output file.

3.5.1 Optional blocks in the SDF-output file

A calculation including Shannon entropy gives two extra lines in the descriptors block of the SDF-output file:

The maximum, minimum, mean, variance and total “internal” Shannon entropies.

“Shannon(i)”	(4f10.4,f10.2,f10.4)

The maximum, minimum, mean, variance and total “external” Shannon entropies (if these are calculated).

“Shannon(e)”	(4f10.4,f10.2,f10.4)

Additionally, extra blocks for the atomic Shannon entropy-related variables are added to the SDF-output after the other atomic-property blocks:

<ATOMIC SURFACE MAXIMUM H (internal)>

Maximum “internal” Shannon entropies	(10f8.4)

<ATOMIC SURFACE MINIMUM H (internal)>

Minimum “internal” Shannon entropies	(10f8.4)

<ATOMIC SURFACE MEAN H (internal)>

Mean “internal” Shannon entropies	`(10f8.4)`

<ATOMIC SURFACE TOTAL H (internal)>

Total “internal” Shannon entropies	`(10f8.4)`

If the external Shannon entropy is also calculated, the following blocks are also written:

<ATOMIC SURFACE MAXIMUM H (external)>

Maximum “external” Shannon entropies	`(10f8.4)`

<ATOMIC SURFACE MINIMUM H (external)>

Minimum “external” Shannon entropies	`(10f8.4)`

<ATOMIC SURFACE MEAN H (external)>

Mean “external” Shannon entropies	`(10f8.4)`

<ATOMIC SURFACE TOTAL H (external)>

Total “external” Shannon entropies	`(10f8.4)`

For calculations that include surface autocorrelations, these are written in the following blocks:

<SURFACE AUTOCORRELATION PARAMETERS>

The initial (lowest) value of the autocorrelation range in Å	`("rlow = ", f12.6)`
The number of autocorrelation points	`("ncorr    = ",i6)`
The autocorrelation step length in Å	`("corrstep = ",f12.6)`
The smoothing parameter σ	`("smooth   = ",f12.6)`

This block is followed by the following tags and blocks, each of which contains **`ncorr`** values for the appropriate autocorrelation function:

<SHAPE AUTOCORRELATION>	`(10f8.1)`
<PLUS-PLUS MEP AUTOCORRELATION>	`(10f8.4)`
<MINUS-MINUS MEP AUTOCORRELATION>	`(10f8.4)`
<PLUS-MINUS MEP AUTOCORRELATION>	`(10f8.4)`

3.6 The surface (.psf) file

The .psf file can be used to derive properties and descriptors from the ParaSurf™ results. It includes the coordinates and properties of the atoms, surface points and surface triangles in the following format:

Number of atoms	`(i6)`

One line per atom with the atomic surface properties:

Atomic number, *x*-coordinate, *y*-coordinate, *z*-coordinate, atomic surface area, V_{max}, V_{min}, IE_L^{min}, EA_L^{max}, mean polarizability	`(i2,3f10.5,f8.3,4f8.2,f8.3)`

Number of surface points	`(i6)`

One line per point with the local properties:

***x*-coordinate, *y*-coordinate, *z*-coordinate, MEP, IE_L, EA_L, α_L, $atom_L$**	`(3f10.5,3f8.2,f8.4,i6)`

(where $atom_L$ is the atom to which the surface point is assigned)

Number of surface triangles	`(i6)`

One line per triangle with the ID of the triangle and the local properties:

point #1, point #2, point #3, area, $atom_{tri}$	`(3i6,f10.5,i6)`

(where point #1, 2 and 3 are the numbers of the surface points that make up the triangle and $atom_{tri}$ is the atom to which the triangle is assigned)

3.7 Anonymous SD (.asd) files

The .asd file contains only those blocks from the ParaSurf™ output SD file that do not pertain directly to the 2D-molecular structure. Its purpose is to allow a full descriptions of the intermolecular bonding properties of the molecule without revealing its structure. The .asd file can only be written from a ParaSurf™ calculation using spherical-harmonic fitting. Its form is:

The SD header line	(A molecular ID number etc.)
The program identifier line	(The normal second line of the SD-file)

And the blocks defined by the following tags:

```
<SPHERICAL_HARMONIC_SURFACE>
<SPHERICAL_HARMONIC_MEP>
<SPHERICAL_HARMONIC_IE(l)
```

<SPHERICAL_HARMONIC_EA(l)

<SPHERICAL_HARMONIC_ALPHA(l)>

<SHAPE HYBRIDS>

<MEP HYBRIDS>

<IE(L) HYBRIDS>

<EA(L) HYBRIDS>

<ALPHA(L) HYBRIDS>

<ParaSurf Descriptors>

(The molecular weight and the atomic surface properties are not included because they would allow the molecular formula to be reconstructed. The atoms assigned to each surface point or triangle are also not given.) The format of the descriptors is:

Molecular	μ, μ_D, α, G, A, VOL `("Molecular ",4f10.4,2f10.2)`
MEP	$V_{\max}$, $V_{\min}$, $\overline{V}_+$, $\overline{V}_-$, $\overline{V}$, ΔV , σ_+^2, σ_-^2, σ_{Tot}^2 , ν , $\sigma_{tot}^2\nu$ `("MEP         ",7f10.2/10x, f10.2,3f10.4)`
IE(l)	$IE_L^{\max}$, $IE_L^{\min}$,$\overline{IE_L}$, ΔIE_L , σ_{IE}^2 `("IE(l)      ",4f10.2,f10.4)`
EA(l)	$EA_L^{\max}$, $EA_L^{\min}$,$\overline{EA_{L+}}$, $\overline{EA_{L-}}$, $\overline{EA_L}$, ΔEA_L , σ_{EA+}^2, σ_{EA-}^2, σ_{EA}^2, ν_{EA}, δA_{EA}^+, A_{EA}^+ `("EA(l)      ",7f10.2/2f10.2,2f10.4,f10.2)`
Eneg(l)	$\overline{\chi_L}$ `("Eneg(l)    ",f10.2)`
Alpha(l)	$\alpha_L^{\max}$, $\alpha_L^{\min}$, $\overline{\alpha_L}$, $\Delta\alpha_L$, σ_α^2 `("Alpha(l)   ",5f10.4)`

3.7.1 Optional blocks

For calculations that include surface autocorrelations, these are written in the following blocks:

<SURFACE AUTOCORRELATION PARAMETERS>

The initial (lowest) value of the autocorrelation range in Å	`("rlow = ", f12.6)`
The number of autocorrelation points	`("ncorr    = ",i6)`
The autocorrelation step length in Å	`("corrstep = ",f12.6)`
The smoothing parameter σ	`("smooth   = ",f12.6)`

This block is followed by the following tags and blocks, each of which contains *ncorr* values for the appropriate autocorrelation function:

`<SHAPE AUTOCORRELATION>`	`(10f8.1)`
`<PLUS-PLUS MEP AUTOCORRELATION>`	`(10f8.4)`
`<MINUS-MINUS MEP AUTOCORRELATION>`	`(10f8.4)`
`<PLUS-MINUS MEP AUTOCORRELATION>`	`(10f8.4)`

3.8 Grid calculations with ParaSurf™

The command

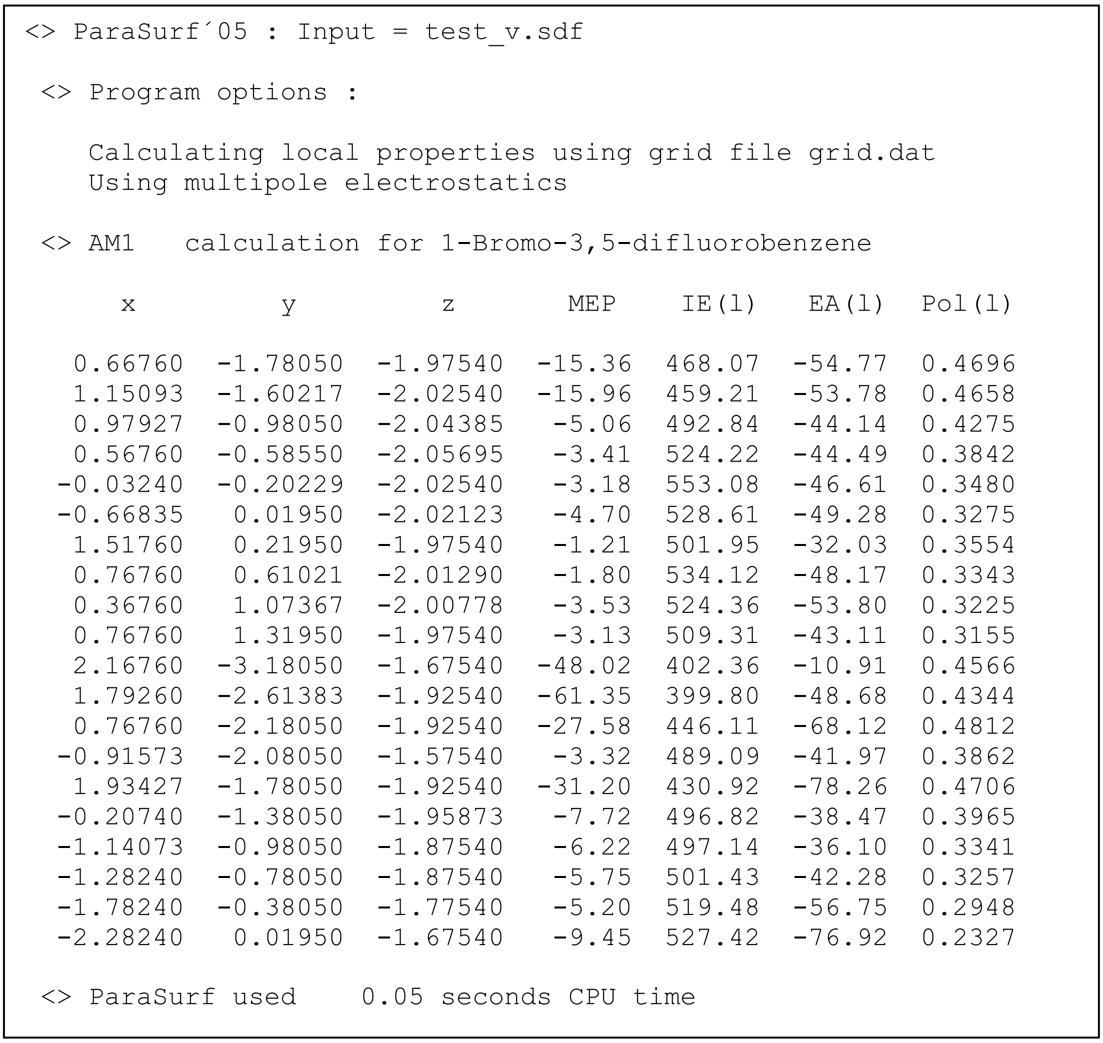
```
parasurf <filename> estat=multi grid=grid.dat
```

instructs ParaSurf™ to read a set of Cartesian coordinates from the file `grid.dat` and to calculate the four local properties (MEP, IE_L, EA_L, α_L). The format of the file `grid.dat` (which must be in the same directory as the input) is one line per atom containing the *x*, *y* and *z* coordinates in free format, comma-separated, maximum line length 80. For instance, the following grid file:

```
0.667600 , -1.780500 , -1.975400
1.150933 , -1.602167 , -2.025400
0.979267 , -0.980500 , -2.043852
0.567600 , -0.585500 , -2.056948
-0.032400 , -0.202286 , -2.025400
-0.668352 ,  0.019500 , -2.021233
1.517600 ,  0.219500 , -1.975400
0.767600 ,  0.610214 , -2.012900
0.367600 ,  1.073667 , -2.007781
0.767600 ,  1.319500 , -1.975400
2.167600 , -3.180500 , -1.675400
1.792600 , -2.613833 , -1.925400
0.767600 , -2.180500 , -1.925400
-0.915733 , -2.080500 , -1.575400
1.934267 , -1.780500 , -1.925400
-0.207400 , -1.380500 , -1.958733
-1.140733 , -0.980500 , -1.875400
-1.282400 , -0.780500 , -1.875400
-1.782400 , -0.380500 , -1.775400
-2.282400 ,  0.019500 , -1.675400
```

Figure 13: Sample grid file

Gives the output shown in **figure 14**.



```
<> ParaSurf´05 : Input = test_v.sdf

 <> Program options :

    Calculating local properties using grid file grid.dat
    Using multipole electrostatics

 <> AM1   calculation for 1-Bromo-3,5-difluorobenzene

      x         y         z        MEP     IE(l)   EA(l)   Pol(l)

   0.66760  -1.78050  -1.97540  -15.36  468.07  -54.77  0.4696
   1.15093  -1.60217  -2.02540  -15.96  459.21  -53.78  0.4658
   0.97927  -0.98050  -2.04385   -5.06  492.84  -44.14  0.4275
   0.56760  -0.58550  -2.05695   -3.41  524.22  -44.49  0.3842
  -0.03240  -0.20229  -2.02540   -3.18  553.08  -46.61  0.3480
  -0.66835   0.01950  -2.02123   -4.70  528.61  -49.28  0.3275
   1.51760   0.21950  -1.97540   -1.21  501.95  -32.03  0.3554
   0.76760   0.61021  -2.01290   -1.80  534.12  -48.17  0.3343
   0.36760   1.07367  -2.00778   -3.53  524.36  -53.80  0.3225
   0.76760   1.31950  -1.97540   -3.13  509.31  -43.11  0.3155
   2.16760  -3.18050  -1.67540  -48.02  402.36  -10.91  0.4566
   1.79260  -2.61383  -1.92540  -61.35  399.80  -48.68  0.4344
   0.76760  -2.18050  -1.92540  -27.58  446.11  -68.12  0.4812
  -0.91573  -2.08050  -1.57540   -3.32  489.09  -41.97  0.3862
   1.93427  -1.78050  -1.92540  -31.20  430.92  -78.26  0.4706
  -0.20740  -1.38050  -1.95873   -7.72  496.82  -38.47  0.3965
  -1.14073  -0.98050  -1.87540   -6.22  497.14  -36.10  0.3341
  -1.28240  -0.78050  -1.87540   -5.75  501.43  -42.28  0.3257
  -1.78240  -0.38050  -1.77540   -5.20  519.48  -56.75  0.2948
  -2.28240   0.01950  -1.67540   -9.45  527.42  -76.92  0.2327

 <> ParaSurf used    0.05 seconds CPU time
```

Figure 14: Sample grid outputfile

The name and the extension (if any) of the grid file are free. Only the output file is written. The units of the local properties are those used in the normal output (i.e. V, IE_L, and EA_L in kcal mol^{-1}, α_L in Ångstrom^3.

3.9 The SIM file format

SIM files must reside in the ParaSurf™ executable directory and are strictly fixed format. SIM files must be called **`<filename>.sim`**, where **`<filename>`** must have exactly three characters. A sample SIM file for a single model (the free energy of solvation in octabol) is shown in **figure 15**:

```
> <OPTIONS>
surf=cube
fit=isod
estat=multi
iso=0.05
> <MODELS>
   1   3
> <DGO>
   3   1.61058
DeltaG(n-Octanol)
kcal/mol
-0.01107     F  1.0      0.0      0.0      1.0      0.0      1.0
 1.6793d-9   F  1.0      0.0      3.0      0.0      0.0      1.0
-2.0407d-10  T  1.0      0.0      1.0      0.0      1.0      1.5
```

Figure 15: Sample surface-integral model (SIM) file.

The first line, the OPTIONS tag, is compulsory and takes the form:

<OPTIONS>

The second to fifth lines, also compulsory in the order shown above, give the ParaSurf™ options to be used for the surface-integral model. These options are given in lower case and override conflicting command-line options.

Line 6 must be the MODELS tag with the format

<MODELS>

Line 7 contains the two integers (***Nmodels*** and ***Maxterms***) that define the number of models given in the file and the maximum number of terms for any one model. The format is:

Nmodels ***Maxterms***	(2i4)

The remainder of the SIM file consists of ***Nmodels*** blocks, each of which defines a single model and has the following format:

Model identifier tag

<MOD>

where MOD is a three-letter unique identifier for the model.

Nterms (the number of terms in the model), ***constant*** (the constant in the regression equation)	(i4,g12.6)

Model name (for output, maximum 20 characters)	`(a20)`
Units of the property P (for output, maximum 20 characters)	`(a20)`
`Nterms` lines, one per term, giving the definition of the model: ***`Coeff   Abs   m   n   o   p   q   r`***	`(d12.6,l3,6f8.4)`

where each term is defined as:

$\left[MEP^m \cdot IE_L^n \cdot EA_L^o \cdot \alpha_L^p \cdot \eta_L^q\right]^r$ if ***`Abs`*** is **false** and $\left[\left|MEP^m \cdot IE_L^n \cdot EA_L^o \cdot \alpha_L^p \cdot \eta_L^q\right|\right]^r$ if ***`Abs`*** is **true**.

SIM files are only intended to be created by expert users.

3.10 Output Tables

The command-line argument **`"table=<filename>"`** requests that the 41 descriptors written in the **`<ParaSurf DESCRIPTORS>`** block of the ParaSurf™ SD-file output are written, one line per molecule, in the file **`<filename>.`** If **`<filename>`** already exists, the line for the new molecules will be appended, otherwise a new file will be created and a header line including designations of the descriptors will be written as the first line. All lines in the table file are comma-separated with all blanks (including those in the Molecule ID) removed. The Descriptors in order are:

Table 5

Column Header	Symbol	Descriptor
MolID	Molecular ID taken from the first line of the entry for each molecule with all blanks eliminated.	
dipole	μ	Dipole moment
dipden	μ_D	Dipolar density
polarizability	α	Molecular electronic polarizability
MWt	MW	Molecular weight
globularity	G	Globularity
totalarea	A	Molecular surface area
volume	VOL	Molecular volume
MEPmax	V_{max}	Maximum (most positive) MEP
MEPmin	V_{min}	Minimum (most negative) MEP
meanMEP+	$\bar{V}_+$	Mean of the positive MEP values
meanMEP-	$\bar{V}_-$	Mean of the negative MEP values
meanMEP	$\bar{V}$	Mean of all MEP values
MEP-range	ΔV	MEP-range

Column Header	Symbol	Descriptor
MEP-range	ΔV	MEP-range
MEPvar-	σ_{+}^{2}	Total variance in the positive MEP values
MEPvar+	σ_{-}^{2}	Total variance in the negative MEP values
MEPvartot	σ_{tot}^{2}	Total variance in the MEP
MEPbalance	ν	MEP balance parameter
var*balance	$\sigma_{tot}^{2}\nu$	Product of the total variance in the MEP and the balance parameter
IELmax	$IE_{L}^{\max}$	Maximum value of the local ionization energy
IELmin	$IE_{L}^{\min}$	Minimum value of the local ionization energy
IELbar	$\overline{IE_{L}}$	Mean value of the local ionization energy
IEL-range	ΔIE_{L}	Range of the local ionization energy
IELvar	σ_{IE}^{2}	Variance in the local ionization energy
EALmax	$EA_{L}^{\max}$	Maximum of the local electron affinity
EALmin	$EA_{L}^{\min}$	Minimum of the local electron affinity
EALbar+	$\overline{EA_{L+}}$	Mean of the positive values of the local electron affinity
EALbar-	$\overline{EA_{L-}}$	Mean of the negative values of the local electron affinity
EALbar	$\overline{EA_{L}}$	Mean value of the local electron affinity
EAL-range	ΔEA_{L}	Range of the local electron affinity
EALvar+	σ_{EA+}^{2}	Variance in the local electron affinity for all positive values
EALvar-	σ_{EA-}^{2}	Variance in the local electron affinity for all negative values
EALvartot	σ_{EAtot}^{2}	Sum of the positive and negative variances in the local electron affinity
EALbalance	ν_{EA}	Local electron affinity balance parameter
EALfraction+	$\delta \mathrm{A}_{EA}^{+}$	Fraction of the surface area with positive local electron affinity
EALarea+	A_{EA}^{+}	Surface area with positive local electron affinity
ENEGbar	$\overline{\chi_{L}}$	Mean value of the local electronegativity
POLmax	$\alpha_{L}^{\max}$	Maximum value of the local polarizability
POLmin	$\alpha_{L}^{\min}$	Minimum value of the local polarizability
POLbar	$\overline{\alpha_{L}}$	Mean value of the local polarizability
POL-range	$\Delta\alpha_{L}$	Range of the local polarizability
POLvar	σ_{α}^{2}	Variance in the local polarizability

[a] Symbols as used 1.6.

If the Shannon entropy is calculated, the following additional descriptors are added:

Column Header	Symbol	Descriptor
SHANImax	$H_{in}^{\max}$	Maximum internal Shannon entropy
SHANImin	$H_{in}^{\min}$	Minimum internal Shannon entropy
SHANIvar	$\sigma_{H(in)}^{2}$	Variance of the internal Shannon entropy
SHANIbar	$\overline{H}_{in}$	Mean internal Shannon entropy
SHANItot	H_{in}^{tot}	Total internal Shannon entropy
and if the external Shannon entropy is also calculated		
SHANEmax	$H_{ex}^{\max}$	Maximum external Shannon entropy
SHANEmin	$H_{ex}^{\min}$	Minimum external Shannon entropy
SHANEvar	$\sigma_{H(ex)}^{2}$	Variance of the external Shannon entropy
SHANEbar	$\overline{H}_{ex}$	Mean external Shannon entropy
SHANEtot	H_{ex}^{tot}	Total external Shannon entropy

3.11 Autocorrelation similarity tables

If the option **“aclist=<filename>”** is used, a user-defined file with the autocorrelation similarities is written. If this file does not exist, it is created and the header line written, otherwise entries are appended. The ASCII file is comma-separated with the following header line:

```
MolID,shape,shapeQ1,shapeQ2,shapeQ3,shapeQ4,Vpp,VppQ1,VppQ2,
VppQ3,VppQ4,Vmm,VmmQ1,VmmQ2,VmmQ3,VmmQ4,Vpm,VpmQ1,VpmQ2,VpmQ3,
VpmQ4
```

The MolID column contains the name of the molecule as given in its SDF-file and similarities for each of the four types of autocorrelation (shape, +/+ MEP, indicated as Vpp, -/- MEP, indicated as Vmm, and +/- MEP, indicated as Vpm) in the order total range – 1st quartal (Q1), second quartal (Q2), third (Q3) and fourth (Q4) quartals.

3.12 Shared files

The Vhamil.par and SIM files are now accessed in shared, read-only mode so that multiple ParaSurf jobs can access the same files.

4 SUPPORT

4.1 Contact

Questions regarding ParaSurf™ should be sent directly to:

support@ceposinsilico.com

4.2 Error reporting

Some of the routines in ParaSurf™ may detect error conditions that have not yet been encountered in our tests. In this case, an error message will be printed requesting that the input and output files be sent to the programming team at the above e-mail address. We realize that this will not always be possible for confidentiality reasons, but if the details can be sent, we will be able to treat the exception and improve the program.

5 APPENDIX

Format of the external statistics file

The statistics file used to define the probabilities for the “external” Shannon entropy as histograms of 200 bins each and has the following format:

Line	Content		Format
1	**Header**		“> <ParaSurf OPTIONS>”
2	**Options used by ParaSurf™:**		a4,2x,a4,2x,a5,2x,g12.4
		Surface type	“CUBE” or “WRAP”
		Fit type	“ISO ” or “SPHH”
		Electrostatic method	“NAOPC” or “MULTI”
		Isodensity value	G12.4
3	**Minimum values of the local properties** (MEP_{min}, IEL_{min}, EAL_{min}, α_{min})		4G12.6
4	**Bin sizes of the local properties** (Δ_{MEP}, Δ_{IEL}, Δ_{EAL}, Δ_{α})		4G12.6
5	**Maximum probabilities for the local properties** $(P^{\max}_{MEP}, P^{\max}_{IEL}, P^{\max}_{EAL}, P^{\max}_{\alpha})$		4G12.6
6-205	**Probabilities for bin *i* (*i*= 1-200)** $(i, P^{i}_{MEP}, P^{i}_{IEL}, P^{i}_{EAL}, P^{i}_{\alpha})$		i6,4g13.6

6 REFERENCES

[1] T. Clark, A. Alex, B. Beck, F. Burkhardt, J. Chandrasekhar, P. Gedeck, A. H. C. Horn, M. Hutter, B. Martin, G. Rauhut, W. Sauer, T. Schindler, and T.Steinke, VAMP 8.2, Erlangen **2002**; available from Accelrys Inc., San Diego, USA (**http://www.accelrys.com/mstudio/vamp.html**).

[2] J. J. P. Stewart, MOPAC2000, 1999, Fujitsu, Ltd, Tokyo, Japan. MOPAC 6.0 was once available as: J. J. P. Stewart, QCPE # 455, Quantum Chemistry Program Exchange, Bloomsville, Indiana, **1990**.

[3] W. Cai, M. Zhang und B. Maigret, *New approach for representation of molecular surface, J. Comput. Chem.*, **1998**, *19*, 1805-1815.

[4] W. Heiden, T. Goetze, and J. Brickmann, *Fast generation of molecular surfaces from 3D data fields with an enhanced "marching cube" algorithm. J. Comput. Chem.* **1993**, *14*, 246-50.

[5] D. W. Ritchie und G. J. L. Kemp, *Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces, J. Comput. Chem.*, **1999**, *20*, 383.

[6] *Chemical Applications of Atomic and Molecular Electrostatic Potentials. Reactivity, Structure, Scattering, and Energetics of Organic, Inorganic, and Biological Systems*, Politzer P; Truhlar DG; (Eds), Plenum Press, New York, NY, **1981**.

[7] P. Sjoberg, J. S. Murray, T. Brinck and P. A. Politzer , *Average local ionisation energies on the molecular surfaces of aromatic systems as guides to chemical reactivity, Can. J. Chem.* **1990**, *68*, 1440-3.

[8] B. Ehresmann, B. Martin, A. H. C. Horn and T. Clark, *Local molecular properties and their use in predicting reactivity*, *J. Mol. Model.,* **2003**, *9*, 342-347.

[9] B. Ehresmann, M. J. de Groot, A. Alex and T. Clark, *New Molecular Descriptors Based on Local Properties at the Molecular Surface and a Boiling-Point Model Derived from Them.*, *J. Chem. Inf. Comp. Sci.*, **2004**, *43*, 658-668.

[10] B. Ehresmann, M. J. de Groot and T. Clark, *A Surface-Integral Solvation Energy Model : The Local Solvation Energy, J. Chem. Inf. Comp. Sci.,* **2005**, *45,* 1053-1060.

[11] L. M. Loew and W. R. MacArthur, *A molecular orbital study of monomeric metaphosphate. Density surfaces of frontier orbitals as a tool in assessing reactivity, J. Am. Chem. Soc.,***1977**, *99*,1019-25.

[12] B. S. Duncan and A. J. Olson, *Approximation and Characterization of Molecular Surfaces*; Scripps Insitute, San Diego, California, **1995**.

[13] J.-H. Lin and T. Clark, *An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties*, *J. Chem. Inf. Model.*, **2005**, *45,* 1010-1016.

[14] G. Rauhut and T. Clark, *Multicenter Point Charge Model for High Quality Molecular Electrostatic Potentials from AM1 Calculations, J. Comput. Chem.*, **1993**, *14*, 503 – 509.

[15] B. Beck, G. Rauhut and T. Clark, *The Natural Atomic Orbital Point Charge Model for PM3: Multipole Moments and Molecular Electrostatic Potentials, J. Comput. Chem.*, **1994**, *15*, 1064 – 1073.

[16] M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, **1977**, *99*, 4899 - 4907; 4907- 4917; *MNDO*, W. Thiel, *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer,III and P. R. Schreiner (Eds), Wiley, Chichester, **1998**, *3*, 1599 –.

[17] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc.* **1985**, *107*, 3902-3909;
A. J. Holder, AM1, *Encyclopedia of Computational Chemistry*, Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R. (Eds), Wiley, Chichester, **1998**, *1*, 8-11.

[18] J. J. P. Stewart, *J. Comput. Chem.*, **1989**, *10*, 209 - 220; 221 - 264; PM3, J. J. P. Stewart, *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer,III and P. R. Schreiner (Eds), Wiley, Chichester, **1998**, *3*, 2080 - .

[19] W. Thiel and A. A. Voityuk, *Extension of the MNDO formalism to d orbitals: integral approximations and preliminary numerical results*, *Theoret. Chim. Acta*, **1992**, *81*, 391 - 404;
W. Thiel and A. A. Voityuk, *Extension of MNDO to d orbitals: parameters and results for the halogens*, **1996**, *93*, 315 - 315;
W. Thiel and A.A. Voityuk, *Extension of MNDO to d orbitals: parameters and results for silicon*, *J. Mol. Struct.*, **1994**, *313*, 141 - 154;
W. Thiel and A. A. Voityuk, *Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group*, *J. Phys. Chem.*, **1996**, *100*, 616 - 626;
MNDO/d, W. Thiel, *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer,III and P. R. Schreiner (Eds), Wiley, Chichester, **1998**, *3*, 1604.

[20] P. Winget, C. Selçuki, A. H. C. Horn, B. Martin and T. Clark, *AM1* Parameters for Phosphorous, Sulfur and Chlorine, J. Mol. Model.* **2003**, *9*, 408-414.

[21] A. H. C. Horn, J.-H. Lin and T. Clark, *A Multipole Electrostatic Model for NDDO-based Semiempirical Molecular Orbital Methods, Theor. Chem. Accts,* **2005**, *113*, 159-168.

[22] G. Schürer, P. Gedeck, M. Gottschalk and T. Clark, *Accurate Parametrized Variational Calculations of the Molecular Electronic Polarizability by NDDO-Based Methods, Int. J. Quant. Chem.*, **1999**, *75*, 17.

[23] D. Rinaldi and J.-L. Rivail, Calculation of molecular electronic polarizabilities. Comparison of different methods, *Theor. Chim. Acta* 1974, *32*, 243-251; J.-L. Rivail and D. Rinaldi, *Variational calculation of multipole electric polarizabilities*, Comptes Rendus, Serie B: Sciences Physiques **1976**, *283*, D. Rinaldi and J.-L. Rivail, *Molecular polarizabilities and dielectric effect of the medium in the liquid state. Theoretical study of the water molecule and its dimers*, *Theor. Chim. Acta* **1973**, *32*, 57.

[24] B. Martin, P. Gedeck and T. Clark, *An Additive NDDO-Based Atomic Polarizability Model. Int. J. Quant. Chem.*, **2000**, *77*, 473-497.

[25] C. J. Cramer, G. R. Famini and A. H. Lowrey, *Acc. Chem. Res.*, **1993**, *26*, 599 – 605.

[26] A. Y. Meyer, *The Size of Molecules*, *Chem. Soc. Rev.,* **1986**, *15*, 449-475.

[27] J. S. Murray and P. Politzer. *Statistical analysis of the molecular surface electrostatic potential: an approach to describing noncovalent interactions in condensed phases, J. Mol. Struct. (Theochem),* **1998**, *425*, 107-114;
J. S. Murray, S. Ranganathan and P. Politzer, *Correlations between the solvent hydrogen bond acceptor parameter* β *and the calculated molecular electrostatic potential, J. Org. Chem.*, **1991**, *56*, 3734-3747;
P. Politzer, P. Lane, J. S. Murray and T. Brinck, *Investigation of relationships between solute molecule surface electrostatic potentials and solubilities in supercritical fluids, J. Phys. Chem.*, **1992**, *96*, 7938-7643;
J. S. Murray, P. Lane, T. Brinck, K. Paulsen, M. E. Grice and P. Politzer*, Relationships of critical constants and boiling points to computed molecular surface properties, J. Phys. Chem.*, **1993**, *97*, 9369-9373.;
T. Brinck, J. S. Murray and P. Politzer, *Quantitative determination of the total local polarity (charge separation) in molecules, Mol. Phys.,* **1992**, *76*, 609.

[28] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication, University of Illinois Press, Chicago*, **1949**.

[29] Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. *The PDBbind Database: Methodologies and updates, J. Med. Chem.,* **2005**; *48*(12); 4111-4119; Wang, R.; Fang, X.; Lu, Y.; Wang, S. *The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures, J. Med. Chem.*, **2004**; *47*(12); 2977-2980; **http://www.pdbbind.org/**

[30] M. Wagener, J. Sadowski and J. Gasteiger, *Autocorrelation of Molecular Surface Properties for Modeling Corticosteroid Binding Globulin and Cytosolic Ah Receptor Activity by Neural Networks, J. Am. Chem. Soc.,* **1995**, *117*, 7769-7775.

[31] **http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp**

[32] W. Thiel, *The MNDOC method, a correlated version of the MNDO model*, *J. Am. Chem. Soc.*, **1981**, *103*, 1413-1420.