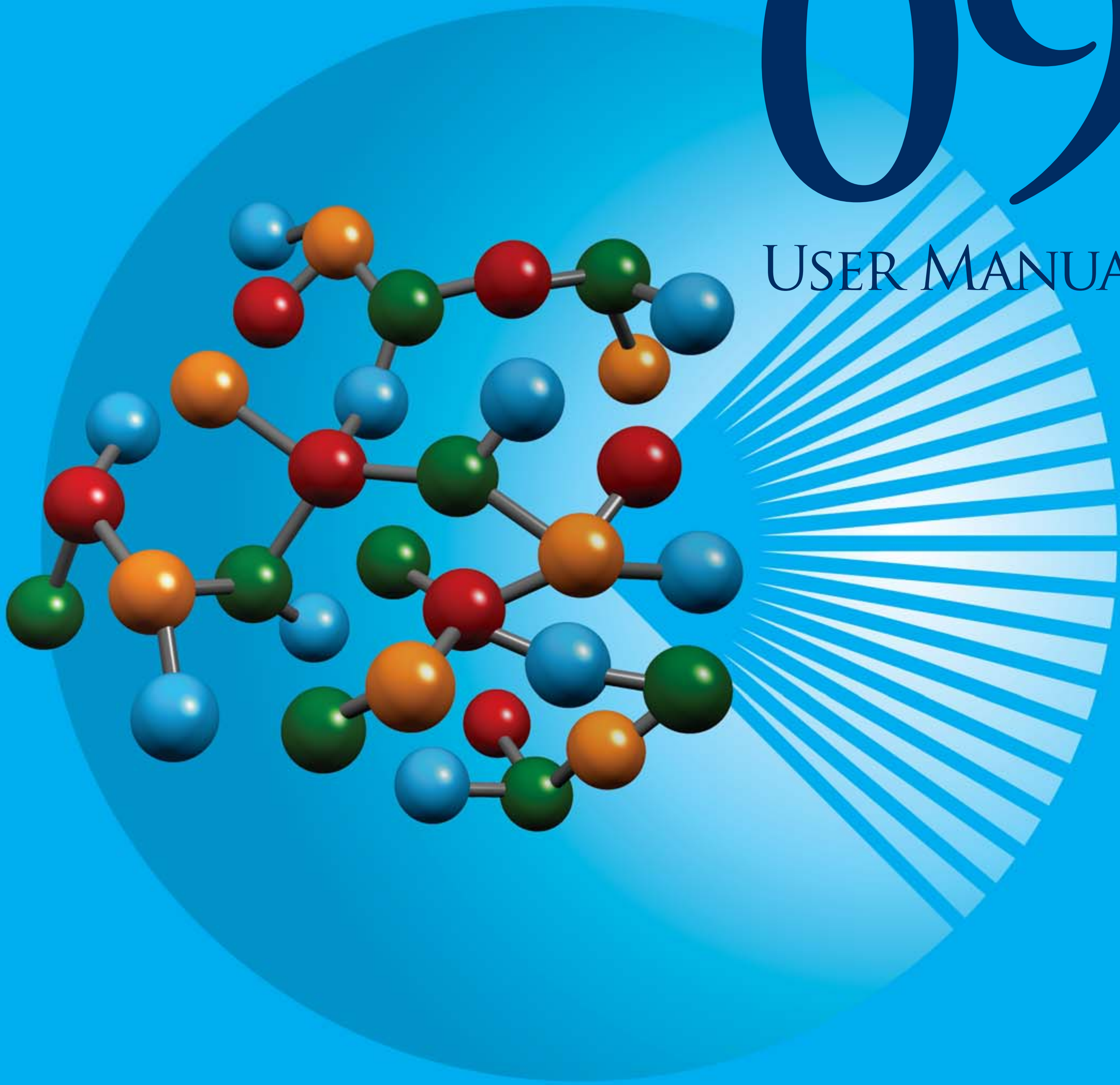
CEPOS INSILICO
SciENCE FOR SciENTISTS



09
USER MANUAL

PARASURF™

Impressum

The Old Vicarage
132 Bedford Road
Kempston
BEDFORD, MK42 8BQ
www.ceposinsilico.com

Manual Timothy Clark

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

CEPOS

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
1st July 2007	ParaSurf´07™ Revision A1	32-bit Windows 32-bit Linux 64-bit Linux Irix
1st July 2008	ParaSurf´08™ Revision A1	32-bit Windows 64-bit Windows 32-bit Linux 64-bit Linux
22nd August 2008	ParaSurf´08™ Revision A2 (minor bug fix release)	
16th December 2008	ParaSurf´08™ Revision A3 (minor bug fix release)	
1st July 2009	ParaSurf´09™ Revision A1	32-bit Windows 64-bit Windows 32-bit Linux 64-bit Linux

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.

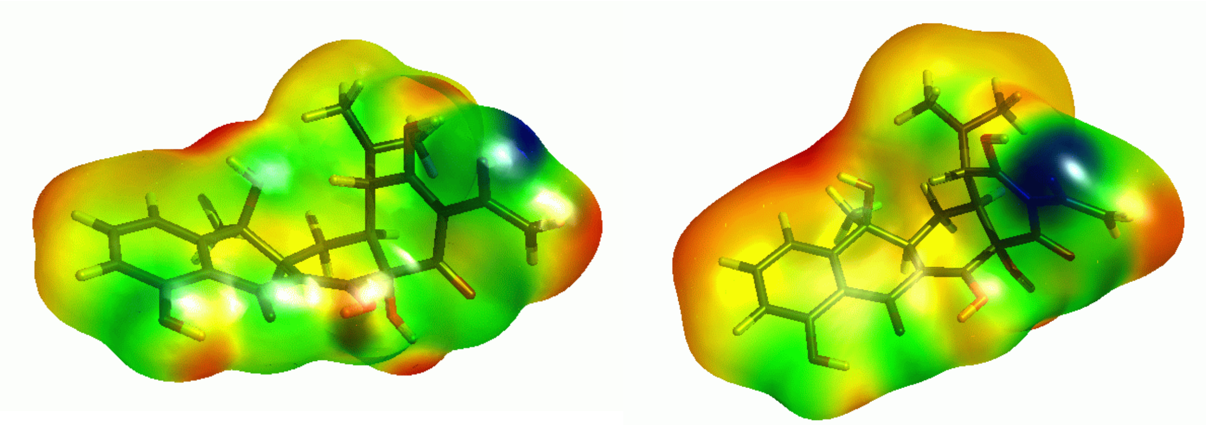


Figure 1: Marching-cube (left) and shrink-wrap (right, fitted to a spherical-harmonic approximation) isodensity surfaces calculated with ParaSurf™ using the default settings

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 written to the ParaSurf™ surface file.

ParaSurf™ is the first program to emerge from the ParaShift collaboration between researchers at 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 Changes relative to ParaSurf´08™

ParaSurf'09™ has been enhanced relative to its predecessor in order to provide better (=faster) performance, improved flexibility and a more comprehensive range of descriptors and features. The changes are outlined below:

1.1.1 Spherical-harmonic fitting (performance enhancements)

The algorithm used to fit spherical-harmonic expansions to shrink-wrapped surfaces has been optimized in ParaSurf'09™. Fitting is now approximately twice as fast as in ParaSurf'08™.

1.1.2 Translation option

ParaSurf'09™ can now calculate the dependence of the spherical-harmonic fit on the position of the molecular center. This will allow ParaFit™ to translate molecules as well as rotate them when overlaying.

1.1.3 New property

The electrostatic field normal to the surface has been introduced as a new property in ParaSurf'09™. A series of additional descriptors based on this property is calculated. Similarly, the surface autocorrelation utility also now includes the field normal to the surface. Grid calculations also now output the vector components of the electrostatic field in addition to the local properties output by ParaSurf'08™. The standard RIFs have also been extended to include coefficients for the field normal to the surface.

1.1.4 Improved error handling

ParaSurf'09™ exits cleanly for each failed molecule and moves on to the next when processing multi-molecule SDF files.

1.1.5 Failsafe procedure for molecular centers

In the event of the standard procedures for finding a suitable molecular centre for spherical-harmonic expansions failing, a new procedure is applied that finds a suitable centre for almost all molecules. For ring-shaped molecules such as macrocycles, however, ParaSurf'09 will still not find a suitable centre.

1.1.6 Cypscore descriptors and scores

ParaSurf'09 calculates and outputs the descriptors and raw (unscaled) scores for the five published CypScore models for predicting sites of metabolization by cytochrome P450 enzymes. [11]

1.2 Isodensity surfaces

Isodensity surfaces [12] 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.

1.3 Solvent-excluded surfaces

The solvent-excluded surface is obtained by rolling a spherical solvent molecule of radius r_{solv} over the surface of the molecule as shown in Figure 2. The surface of the solvent molecule defines the molecular surface, so that the yellow volume in Figure 2 becomes part of the molecule.

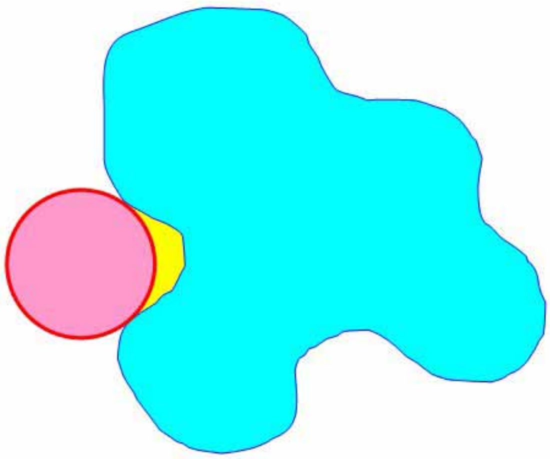


Figure 2: 2D-representation of a solvent-excluded surface.

1.4 Solvent-accessible surfaces

Solvent-accessible surfaces are obtained in the same way as solvent-excluded surfaces but the outer surface of the solvent sphere is used to define the molecular surface, as shown in Figure 3.

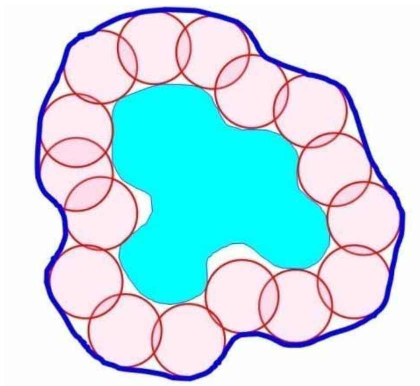


Figure 3: The solvent-excluded surface is obtained by rolling a spherical "solvent molecule".

1.5 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 4) and not multiply (vectors **C** and **D** in Figure 4):

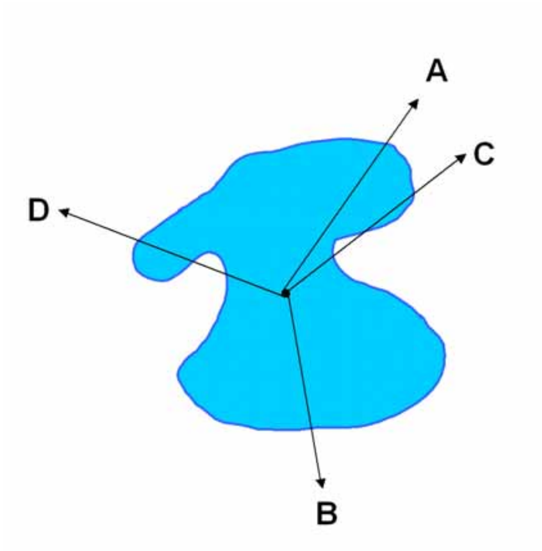


Figure 4: 2D-representation of a molecular surface with single-valued (A and B) and multiply valued (C and D) radial vectors from the center

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 5) 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 5). Thus, the shrink-wrapped surface may contain areas (marked by dashed lines in Figure 5) 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.

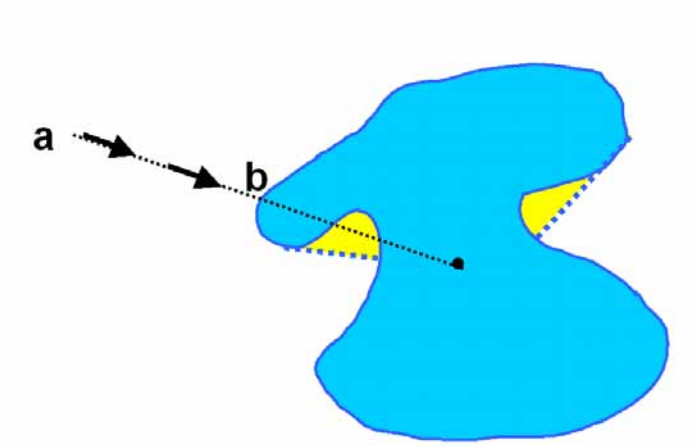


Figure 5: 2D-representation of the shrink-wrap algorithm. The algorithms scans along the vector from point a towards the center of the molecule until the electron density reaches the preset value (point b). The algorithm results in enclosures (marked yellow) for multi-valued radial vectors.

Figure 6 shows a spherical-harmonically fitted shrink-wrap surface for a difficult molecule. The areas shown schematically in Figure 5 are clearly visible.

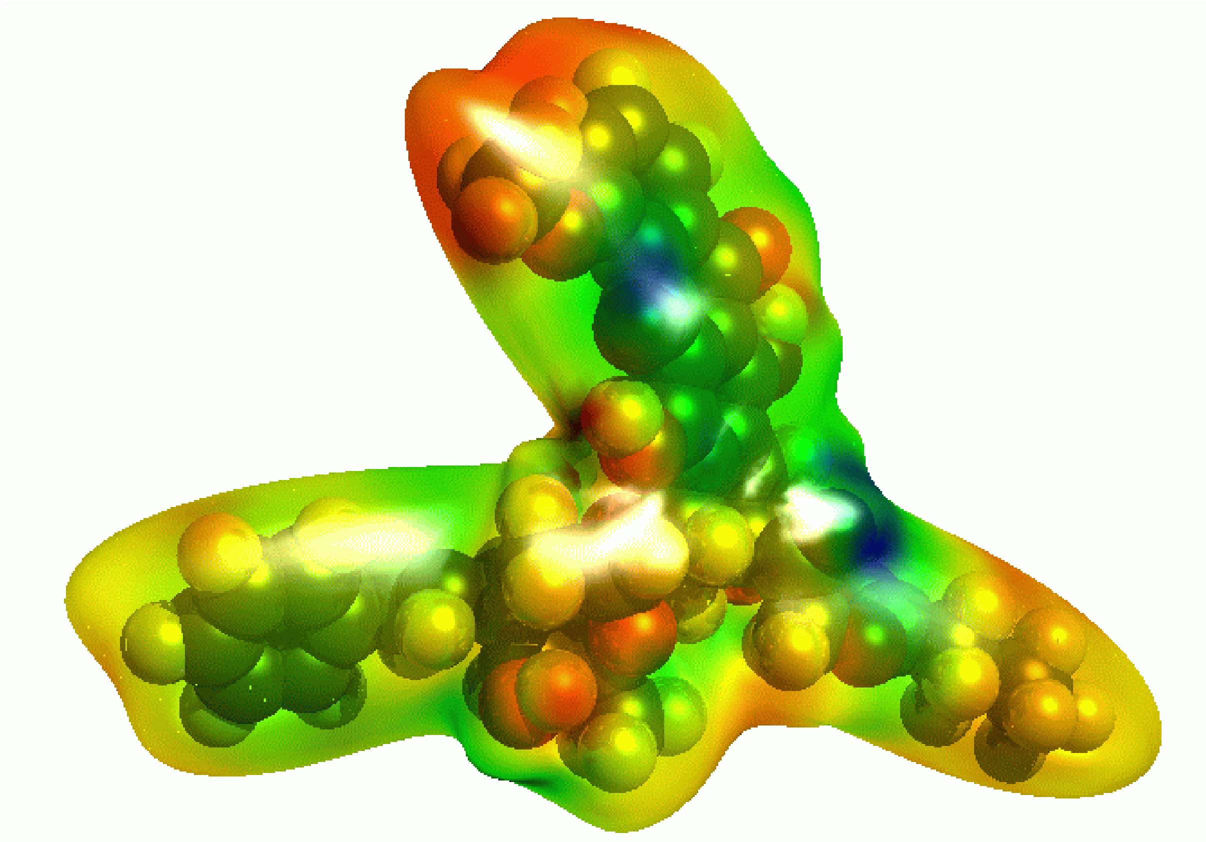


Figure 6: Spherical-harmonic approximation of a shrink-wrap isodensity surface. Note the areas where the surface does not follow the indentations of the moecule.

1.6 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.

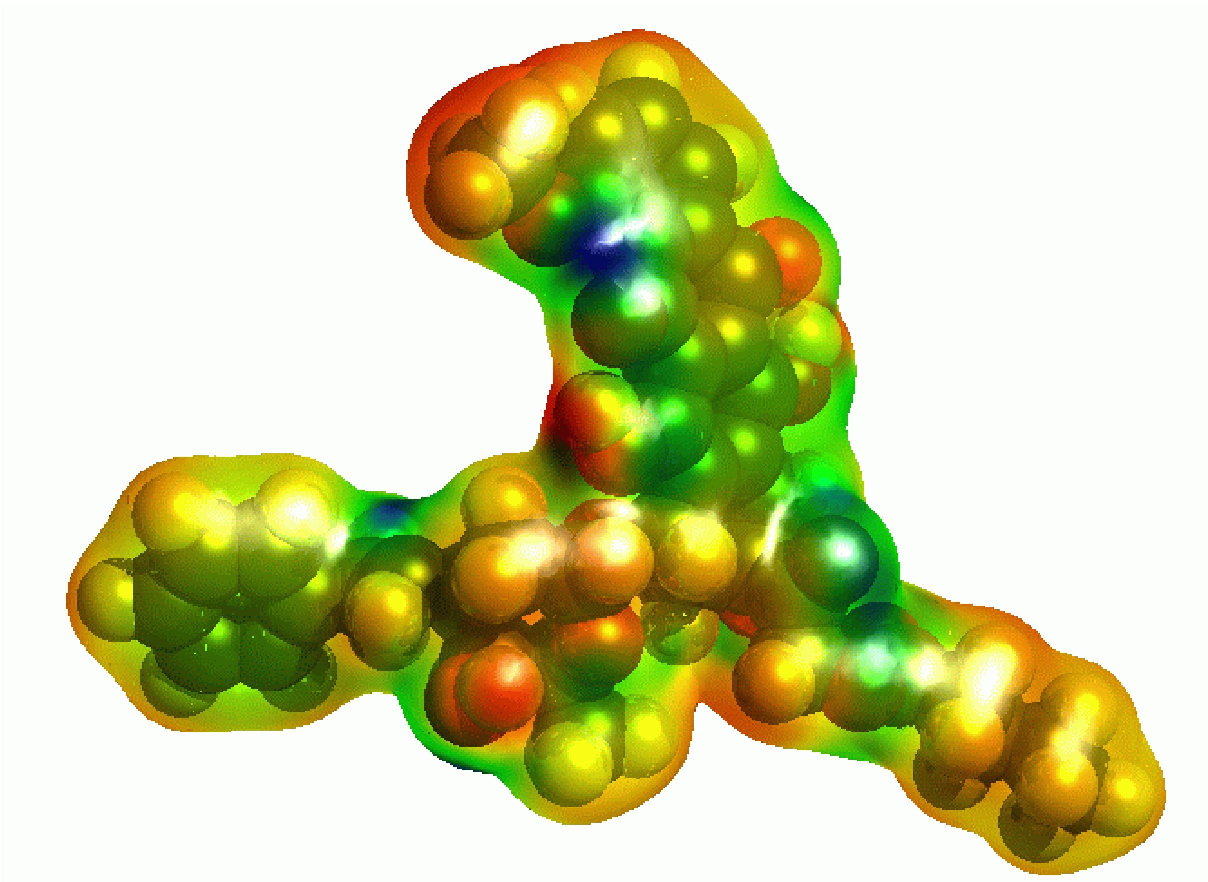


Figure 7: Marching-cube isodensity surface for the molecule shown in Figure 4. This surface is better suited for QSPR and surface-integral models

1.7 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:

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

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

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

where P_l^m *(cos* α*)* are associated Legendre functions and l and m are integers such that $-l<=m<=l$. In the above form, spherical harmonics are complex functions. Duncan and Olson [13] have used the real functions

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

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 1.8). In this way, a completely analytical description of the shape of the molecule and its intermolecular binding properties is obtained. [14] 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 (0.00002 $e^-Å^{-3}$) is lower than that (0.0003 $e^-Å^{-3}$) 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'09™ 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.

The choice of center for fitting to a spherical-harmonic expansion is critical. ParaSurf'09™ therefore goes through a multi-step procedure in order to find a suitable center. This procedure is retained for all molecules for which the ParaSurf'08™ found a suitable center. However, if the algorithms implemented in ParaSurf'08™ fail to find a suitable center, the additional technique implemented in ParaSurf'09 will probably work.

The problem with many molecules is that, for instance, the center of mass does not lie within the molecular volume. This can easily be the case for, for instance, U- or L-shaped molecules. The procedure implemented in ParaSurf'09 works as follows:

1. The program first calculates the center of mass and tests whether it lies within the volume of the molecule. If it does, it is used as the molecular center. If not, the program moves on to the next step.

2. ParaSurf™ calculates the principal moments of inertia of the molecule and derives a center from them by assuming that the molecule is U- or V-shaped. The procedure tries to place the center at the base center of the molecule. This procedure was implemented in ParaSurf'08™ as a fallback if the center of mass proved unsuitable. If it also fails to find a suitable center, ParaSurf'09™ moves on to a third option, which finds a center for all but the most difficult molecules.

3. The new procedure first searches for the largest plane in the molecule (i.e. the one that contains the most atoms). This search has some leeway, so that the atoms must not all lie exactly in the plane. As a second step, the second largest plane is sought. The molecular center is then placed in the hinge area between the two planes, as illustrated in Figure 8:

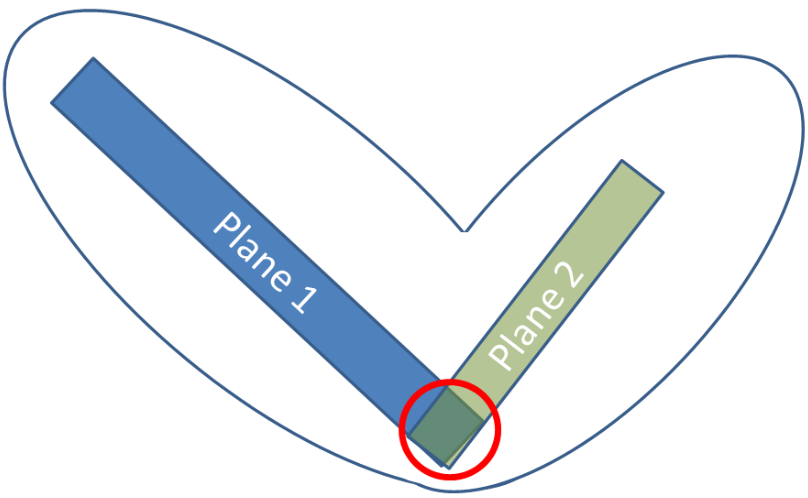


Figure 8: Schematic representation of the planes and hinge area used to determine the center for spherical-harmonic expansions.

1.8 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.

1.8.1 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.

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

The NAO-PC model [15, 16] 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, [17] AM1 [18] and PM3. [19] 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. [15] 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 [20] or AM1* [21] that use *d*-orbitals in the basis set.

1.8.1.2 The multipole model

The integrals needed to evaluate **equation (4)** in MNDO-type methods use a multipole approximation [17, 20] 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. [22] 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™.

1.8.2 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:

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

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]

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

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:

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

and the local electronegativity, χ_L:

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

1.8.3 Local polarizability

Within the NDDO, the molecular electronic polarizability is easily accessible using the parameterized version [23] of the variational technique introduced by Rivail, [24] which can also be partitioned into an additive polarizability scheme. [25] This allows us to define the local polarizability, α_L, at a point near the molecule as

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

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.

1.8.4 Field normal to the surface

The electrostatic field (the first derivative of the potential) normal to the molecular surface is closely related to the electrostatic solvation energy in implicit solvation models. [26, 27] This field also has the advantage that it is largely independent of the total molecular charge, so that charged molecules can be compared with neutral ones. If the molecular electrostatic potential is used for this purpose, the charge of ions leads a shift in the potential descriptors, so that molecules and ions with different charges cannot be compared directly. The direction of the normal field (inwards or outwards) also defines, for instance hydrogen-bond donors and acceptors specifically.

1.9 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	Symbol in CSV file
μ	Dipole moment		`dipole`
μ_D	Dipolar density	[25]	`dipden`
α	Molecular electronic polarizabilty	[28]	`polarizability`
MW	Molecular weight		`MWt`
G	Globularity	[29]	`globularity`
A	Molecular surface area		`totalarea`

Descriptor	Description	Formula/ Reference	Symbol in CSV file
VOL	Molecular volume		volume
V_{max}	Maximum (most positive) MEP	[30]	MEPmax
V_{min}	Minimum (most negative) MEP	[30]	MEPmin
$\overline{V}_{+}$	Mean of the positive MEP values	[30]	meanMEP+
$\overline{V}_{-}$	Mean of the negative MEP values	[30]	meanMEP-
$\overline{V}$	Mean of all MEP values	[30]	meanMEP
ΔV	MEP-range	[30]	MEP-range
σ_{+}^{2}	Total variance in the positive MEP values	[30]	MEPvar+
σ_{-}^{2}	Total variance in the negative MEP values	[30]	MEPvar-
σ_{tot}^{2}	Total variance in the MEP	[30]	MEPvartot
ν	MEP balance parameter	[30]	MEPbalance
$\sigma_{tot}^{2}\nu$	Product of the total variance in the MEP and the balance parameter	[30]	var*balance
γ_1^V	Skewness of the MEP-distribution	$\gamma_1^{\alpha_L} = \frac{\sum_{i=1}^{N}\left(\alpha_L^i - \overline{\alpha}_L\right)^3}{(N-1)\sigma^3}$	MEPskew
γ_2^V	Kurtosis of the MEP-distribution	$\gamma_2^V = \frac{\sum_{i=1}^{N}\left(V_i - \overline{V}\right)^4}{(N-1)\sigma^4} - 3$	MEPkurt
$\int_V$	Integrated MEP over the surface	$\int_V = \sum_{i=1}^{N} V_i a_i$	MEPint
$IE_L^{\max}$	Maximum value of the local ionization energy		IELmax
$IE_L^{\min}$	Minimum value of the local ionization energy		IELmin
$\overline{IE_L}$	Mean value of the local ionization energy	$\overline{IE_L} = \frac{1}{N}\sum_{i=1}^{N} IE_L^i$	IELbar
ΔIE_L	Range of the local ionization energy	$\Delta IE_L = IE_L^{\max} - IE_L^{\min}$	IELrange
σ_{IE}^2	Variance in the local ionization energy	$\sigma_{IE}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[IE_L^i - \overline{IE_L}\right]^2$	IELvar

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\gamma_1^{IE_L}$	Skewness of the local ionization energy distribution	$\gamma_1^{IE_L} = \frac{\sum_{i=1}^{N}\left(IE_L^i - \overline{IE}_L\right)^3}{(N-1)\sigma^3}$	**IELskew**
$\gamma_2^{IE_L}$	Kurtosis of the local ionization energy distribution	$\gamma_2^{IE_L} = \frac{\sum_{i=1}^{N}\left(IE_L^i - \overline{IE}_L\right)^4}{(N-1)\sigma^4} - 3$	**IELkurt**
$\int_{IE_L}$	Integrated local ionization energy over the surface	$\int_{IE_L} = \sum_{i=1}^{N} IE_L^i a_i$	**IELint**
$EA_L^{\max}$	Maximum of the local electron affinity		**EALmax**
$EA_L^{\min}$	Minimum of the local electron affinity		**EALmin**
$\overline{EA}_{L+}$	Mean of the positive values of the local electron affinity	$\overline{EA}_{L+} = \frac{1}{N^+}\sum_{i=1}^{N^+} EA_{L+}^i$	**EALbar+**
$\overline{EA}_{L-}$	Mean of the negative values of the local electron affinity	$\overline{EA}_{L-} = \frac{1}{N^-}\sum_{i=1}^{N^-} EA_{L-}^i$	**EALbar-**
$\overline{EA}_L$	Mean value of the local electron affinity	$\overline{EA}_L = \frac{1}{N}\sum_{i=1}^{N} EA_L^i$	**EALbar**
ΔEA_L	Range of the local electron affinity	$\Delta EA_L = EA_L^{\max} - EA_L^{\min}$	**EALrange**
σ_{EA+}^2	Variance in the local electron affinity for all positive values	$\sigma_{EA+}^2 = \frac{1}{m}\sum_{i=1}^{m}\left[EA_i^+ - \overline{EA^+}\right]^2$	**EALvar+**
σ_{EA-}^2	Variance in the local electron affinity for all negative values	$\sigma_{EA-}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[EA_i^- - \overline{EA^-}\right]^2$	**EALvar-**
σ_{EAtot}^2	Sum of the positive and negative variances in the local electron affinity	$\sigma_{EAtot}^2 = \sigma_{EA+}^2 + \sigma_{EA-}^2$	**EALvartot**
ν_{EA}	Local electron affinity balance parameter	$\nu_{EA} = \frac{\sigma_{EA+}^2 \cdot \sigma_{EA-}^2}{\left[\sigma_{EA}^2\right]^2}$	**EALbalance**
δA_{EA}^+	Fraction of the surface area with positive local electron affinity	$\delta A_{EA}^+ = \frac{A_{EA}^+}{A}$, A = total surface area	**EALfraction+**
A_{EA}^+	Surface area with positive local electron affinity		**EALarea+**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\gamma_1^{EA_L}$	Skewness of the local electron affinity distribution	$\gamma_1^{EA_L} = \frac{\sum_{i=1}^{N}\left(EA_L^i - \overline{EA_L}\right)^3}{(N-1)\sigma^3}$	**EALskew**
$\gamma_2^{EA_L}$	Kurtosis of the local electron affinity distribution	$\gamma_2^{EA_L} = \frac{\sum_{i=1}^{N}\left(EA_L^i - \overline{EA_L}\right)^4}{(N-1)\sigma^4} - 3$	**EALkurt**
$\int_{EA_L}$	Integrated local electron affinity over the surface	$\int_{IE_L} = \sum_{i=1}^{N} EA_L^i a_i$	**EALint**
$\alpha_L^{\max}$	Maximum value of the local polarizability		**POLmax**
$\alpha_L^{\min}$	Minimum value of the local polarizability		**POLmin**
$\overline{\alpha_L}$	Mean value of the local polarizability	$\overline{\alpha_L} = \frac{1}{N}\sum_{i=1}^{N}\alpha_L^i$	**POLbar**
$\Delta\alpha_L$	Range of the local polarizability	$\Delta\alpha_L = \alpha_L^{\max} - \alpha_L^{\min}$	**POLrange**
σ_α^2	Variance in the local polarizability	$\sigma_\alpha^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\alpha_L^i - \overline{\alpha_L}\right]^2$	**POLvar**
$\gamma_1^{\alpha_L}$	Skewness of the local polarizability distribution	$\gamma_1^{\alpha_L} = \frac{\sum_{i=1}^{N}\left(\alpha_L^i - \overline{\alpha}_L\right)^3}{(N-1)\sigma^3}$	**POLskew**
$\gamma_2^{\alpha_L}$	Kurtosis of the local polarizability distribution	$\gamma_2^{\alpha_L} = \frac{\sum_{i=1}^{N}\left(\alpha_L^i - \overline{\alpha}_L\right)^4}{(N-1)\sigma^4} - 3$	**POLkurt**
$\int_{\alpha_L}$	Integrated local polarizability over the surface	$\int_{\alpha_L} = \sum_{i=1}^{N}\alpha_L^i a_i$	**POLint**
$\chi_L^{\max}$	Maximum value of the local electronegativity		**ENEGmax**
$\chi_L^{\min}$	Minimum value of the local electronegativity		**ENEGmin**
$\overline{\chi_L}$	Mean value of the local electronegativity	$\overline{\chi_L} = \frac{1}{N}\sum_{i=1}^{N}\chi_L^i$	**ENEGbar**
$\Delta\chi_L$	Range of the local electron electronegativity	$\Delta\chi_L = \chi_L^{\max} - \chi_L^{\min}$	**ENEGrange**
σ_χ^2	Variance in the local electronegativity	$\sigma_\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\chi_L^i - \overline{\chi_L}\right]^2$	**ENEGvar**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\gamma_1^{\chi_L}$	Skewness of the local electronegativity distribution	$\gamma_1^{\chi_L} = \frac{\sum_{i=1}^{N}\left(\chi_L^i - \overline{\chi}_L\right)^3}{(N-1)\sigma^3}$	**ENEGskew**
$\gamma_2^{\chi_L}$	Kurtosis of the local electronegativity distribution	$\gamma_2^{\chi_L} = \frac{\sum_{i=1}^{N}\left(\chi_L^i - \overline{\chi}_L\right)^4}{(N-1)\sigma^4} - 3$	**ENEGkurt**
$\int_{\chi_L}$	Integrated local electronegativity over the surface	$\int_{\chi_L} = \sum_{i=1}^{N} \chi_L^i a_i$	**ENEGint**
η_L^{max}	Maximum value of the local hardness		**HARDmax**
η_L^{min}	Minimum value of the local hardness		**HARDmin**
$\overline{\eta_L}$	Mean value of the local hardness	$\overline{\eta_L} = \frac{1}{N}\sum_{i=1}^{N} \eta_L^i$	**HARDbar**
$\Delta\eta_L$	Range of the local electron hardness	$\Delta\eta_L = \eta_L^{max} - \eta_L^{min}$	**HARDrange**
σ_η^2	Variance in the local hardness	$\sigma_\eta^2 = \frac{1}{N}\sum_{i=1}^{N}\left[\eta_L^i - \overline{\eta_L}\right]^2$	**HARDvar**
$\gamma_1^{\eta_L}$	Skewness of the local hardness distribution	$\gamma_1^{\eta_L} = \frac{\sum_{i=1}^{N}\left(\eta_L^i - \overline{\eta}_L\right)^3}{(N-1)\sigma^3}$	**HARDskew**
$\gamma_2^{\eta_L}$	Kurtosis of the local hardness distribution	$\gamma_2^{\eta_L} = \frac{\sum_{i=1}^{N}\left(\eta_L^i - \overline{\eta}_L\right)^4}{(N-1)\sigma^4} - 3$	**HARDkurt**
$\int_{\eta_L}$	Integrated local hardness over the surface	$\int_{\eta_L} = \sum_{i=1}^{N} \eta_L^i a_i$	**HARDint**
Additionally if the Shannon Entropy is calculated			
H_{in}^{max}	Maximum value of the internal Shannon Entropy		**SHANImax**
H_{in}^{min}	Minimum value of the internal Shannon Entropy		**SHANImin**
$\overline{H}_{in}$	Mean value of the internal Shannon Entropy	$\overline{H}_{in} = \frac{1}{N}\sum_{i=1}^{N} H_{in}^i$	**SHANIbar**
$\sigma_{H_{in}}^2$	Variance in the internal Shannon Entropy	$\sigma_{H_{in}}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[H_{in}^i - \overline{H}_{in}\right]$	**SHANIvar**
$\int_{H_{in}}$	Integrated internal Shannon Entropy over the surface	$\int_{H_{in}} = \sum_{i=1}^{N} H_{in}^i a_i$	**SHANItot**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
And if the external Shannon Entropy is available			
$H_{ex}^{\max}$	Maximum value of the external Shannon Entropy		**SHANEmax**
$H_{ex}^{\min}$	Minimum value of the external Shannon Entropy		**SHANEmin**
$\overline{H_{ex}}$	Mean value of the external Shannon Entropy	$\overline{H_{ex}} = \frac{1}{N}\sum_{i=1}^{N} H_{ex}^{i}$	**SHANEbar**
$\sigma_{H_{ex}}^2$	Variance in the external Shannon Entropy	$\sigma_{H_{ex}}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[H_{ex}^{i} - \overline{H_{ex}}\right.$	**SHANEvar**
$\int_{H_{ex}}$	Integrated internal Shannon Entropy over the surface	$\int_{H_{ex}} = \sum_{i=1}^{N} H_{ex}^{i} a_i$	**SHANEtot**
$F_N^{\max}$	Maximum value of the electrostatic field normal to the surface		**FNmax**
$F_N^{\min}$	Minimum value of the field normal to the surface		**FNmin**
ΔF_N	Range of the field normal to the surface	$\Delta F_N = F_N^{\max} - F_N^{\min}$	**FNrange**
$\overline{F_N}$	Mean value of the field normal to the surface	$\overline{F_N} = \frac{1}{N}\sum_{i=1}^{N} \chi_L^i$	**FNmean**
σ_F^2	Variance in field normal to the surface	$\sigma_F^2 = \frac{1}{N}\sum_{i=1}^{N}\left[F_N^i - \overline{F_N}\right]^2$	**FNvartot**
σ_{F+}^2	Variance in the field normal to the surface for all positive values	$\sigma_{F+}^2 = \frac{1}{m}\sum_{i=1}^{m}\left[F_N^{i+} - \overline{F_N^{+}}\right]^2$	**FNvar+**
σ_{F-}^2	Variance in the field normal to the surface for all negative values	$\sigma_{F-}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[F_N^{i-} - \overline{F_N^{-}}\right]^2$	**FNvar-**
ν_F	Normal field balance parameter	$\nu_F = \frac{\sigma_{F+}^2 \cdot \sigma_{F-}^2}{\left[\sigma_F^2\right]^2}$	**FNbal**
$\gamma_1^{F_N}$	Skewness of the field normal to the surface	$\gamma_1^{F_N} = \frac{\sum_{i=1}^{N}\left(F_N^i - \overline{F_N}\right)^3}{(N-1)\sigma^3}$	**FNskew**
$\gamma_2^{F_N}$	Kurtosis of the field normal to the surface	$\gamma_2^{F_N} = \frac{\sum_{i=1}^{N}\left(F_N^i - \overline{F_N}\right)^4}{(N-1)\sigma^4} - 3$	**FNkurt**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\int_{F_N}$	Integrated field normal to the surface over the surface	$\int_{F_N} = \sum_{i=1}^{N} F_N^i a_i$	**FNint**
$\int_{F_N}^{+}$	Integrated field normal to the surface over the surface for all positive values	$\int_{F_N} = \sum_{i=1}^{N} F_N^i a_i \text{ if } F_N^i \geq 0$	**FN+**
$\int_{F_N}^{-}$	Integrated field normal to the surface over the surface for all negative values	$\int_{F_N} = \sum_{i=1}^{N} F_N^i a_i \text{ if } F_N^i < 0$	**FN-**
$\int_{\lvert F_N \rvert}$	Integrated absolute field normal to the surface over the surface	$\int_{F_N} = \sum_{i=1}^{N} \lvert F_N^i \rvert a_i$	**FNabs**

1.10 Surface-integral models

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

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

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.

1.11 Spherical harmonic "hybrids"

Once the molecular shape or a local property have been fitted to a spherical-harmonic expansion, [12] 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.

1.12 Descriptors and moments based on surface-integral models

ParaSurf™ uses local properties defined in a surface-integral model (SIM, see 1.10) 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.

1.13 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. [31]

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

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

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™ 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, 32] in their correct protonation states and at geometries obtained by optimizing with AM1 starting from the bound conformation. [30, 33]

Alternatively, the user can define a custom "external" statistics file using the ParaSurf™ module **binner** (available free of charge for ParaSurf™ users). 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. [33]

1.14 Surface autocorrelations

Gasteiger et al. [34] introduced the concept of surface autocorrelations as powerful descriptions of molecular binding properties for quantitative structure-activity relationships (QSARs). In ParaSurf™, autocorrelations *A(R)* are defined as:

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

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™. 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™ 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 9 shows the four autocorrelation functions for trimethoprim calculated with AM1.

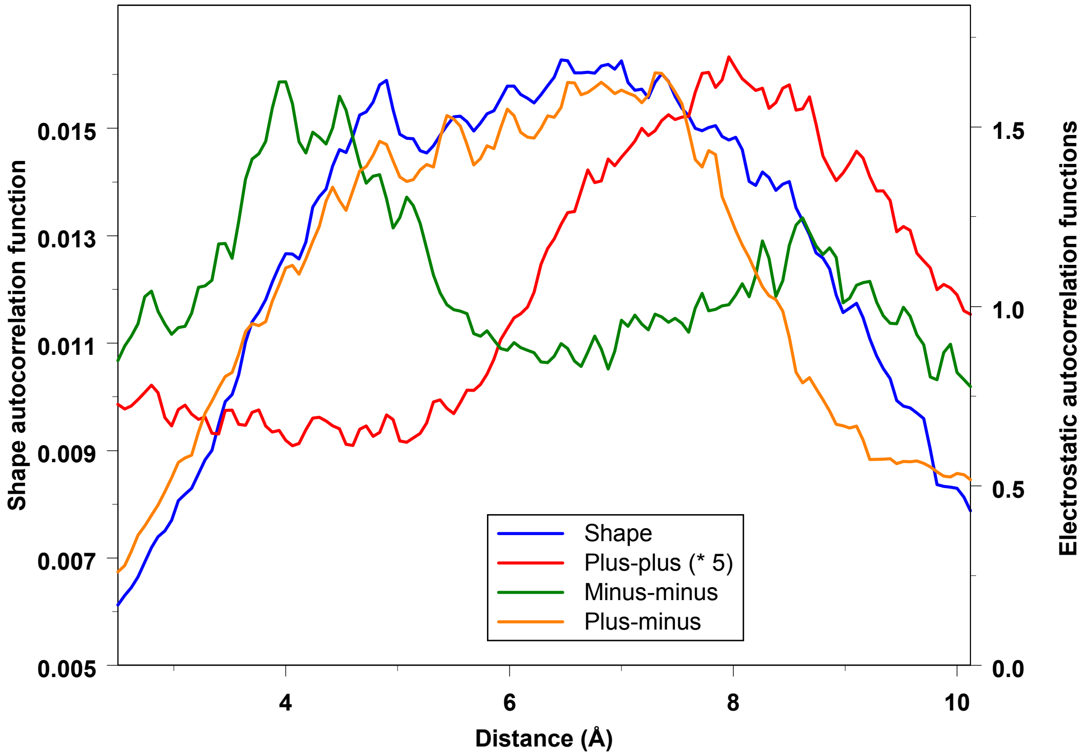


Figure 9: The four autocorrelation functions calculated using the AM1 Hamiltonian for trimethoprim.

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.sdf** output file. The similarities *S* are defined as:

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

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**).

1.15 Standard Rotationally Invariant Fingerprints (RIFs)

Mavridis et al. [**35**] introduced standard rotationally invariant fingerprints (RIFs) based on the spherical-harmonic hybridization coefficients defined above. These fingerprints provide a detailed description of the molecular shape, electrostatics, donor/acceptor properties and polarizability as a standard series of 54 floating point numbers.

1.16 Maxima and Minima of the Local Properties

Jakobi et al. [**36**] have described the calculation and use of the most significant maxima and minima of the local properties on the surface of the molecule. These points were used in the ParaFrag procedure to detect scaffold hops with high similarity and can be viewed as pharmacophore points.

1.17 CypScore descriptors and raw scores

Hennemann et al. [**11**] have used quantities calculated by ParaSurf™ as descriptors for five separate models for the lability of a given reactive position (atom) towards oxidation by a cytochrome P450 “superenzyme”. The descriptors and the raw (unscaled) scores for each atom of the appropriate element (but not necessarily atom type) are calculated by ParaSurf’09™ and output to the SDF file (but not the output file). The scaling factors needed to convert these raw scores to the unified CypScore scale were not published, but can be obtained by reverse engineering the data given in the paper. Scripts and workflows are available for this purpose.

2 PROGRAM OPTIONS

2.1 Command-line 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). **<name>** is not required if the first argument is **-version** (see below) Using this option, the input file is assumed to be **<name>_v.sdf** if a file with this name exists. Otherwise the file **<name>.sdf** 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 **<name>_m.sdf** The output files are **<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.
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** $e^{-}Å^{-3}$ (default for shrink-wrap surface = **0.00002**; default for marching-cube surface = **0.0003**; minimum possible value = **0.00001**)

		minimum possible value = 0.00001)
rsol=	*n.nn*	A solvent-probe radius of **n.nn** Å is used for calculating the solvent-excluded or solvent-accessible surface (default=1.0, allowed range is from 0.0 to 2.0 Å)
mesh=	*n.nn*	The mesh size used to triangulate the surface is set to **n.nn** Å (default value = 0.2 Å, allowed range is from 0.1 to 1.0 Å)
estat=	naopc	Use NAO-PC electrostatics
	multi	Use multipole electrostatics (default)
psf=	on	Write .psf surface file
	off	Do not write **.psf** surface file (default)
asd=	on	Write anonymous SD (**.asd**) file
	off	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
	anr	Write **.vmp** file for debugging. Map the number of the atom assigned to the surface element onto the surface
	fnm	Write **.vmp** file for debugging. Map F_N onto the surface
	sha	Write **.vmp** file for debugging. Map the Shannon entropy 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 3.8.1
	auto	ParaSurf™ calculates an automatic grid (see 3.8.2)
lattice=	*n.nn*	Sets the lattice spacing for the **grid=auto** option (see 3.8.2)
sim=	<filename>	One or more surface-integral models will be read from the file **<filename>.sim** in the ParaSurf™ root 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.
autocorr **=<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 **aclist=**).
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.
aclist= **<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.
riflist= **<filename>**	An ASCII table of the calculated a standard rotationally invariant fingerprint (RIF) 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.
translate ***=n.nn***	Requests that ParaSurf™ performs low-resolution spherical-harmonic fits using translated centers at (***+n.nn,0,0***) , (***-n.nn,0,0***), (***0,+n.nn,0***), (***0,-n.nn,0***), (***0,0,+n.nn***) and (***0,0,-n.nn***) relative to the original center. The default value of ***n.nn*** is 0.5 Å. This value is obtained if **translate** is used alone. The maximum value of ***n.nn*** allowed is 1.0 Å. The translate option will be needed for later versions of ParaFit™ that allow translation of the molecule when overlaying.
translate2 ***=n.nn***	Requests that ParaSurf™ performs a more detailed translation scan with low-resolution spherical-harmonic fits using translated centers at (***+n.nn,0,0***) , (***+2n.nn,0,0***), (***-n.nn,0,0***), (***-2n.nn,0,0***), (***0,+n.nn,0***), (***0,+2n.nn,0***), (***0,-n.nn,0***), (***0,-2n.nn,0***), (***0,0,+n.nn***), (***0,0,+2n.nn***), (***0,0,-n.nn***) and (***0,0,-2n.nn***) relative to the original center. The default value of ***n.nn*** is 0.25 Å. This value is obtained if **translate2** is used alone. The maximum value of ***n.nn*** allowed is 0.5 Å. The **translate2** option will be needed for later versions of ParaFit™ that allow translation of the molecule when overlaying.
-version	Must be the first argument. Requests that ParaSurf™ prints the version number to the standard output channel and then stops without performing a calculation.

Examples:

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

Use the input file **test_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.

2.2 Options defined in the input SDF-file

2.2.1 Defining the center for spherical-harmonic fits

The automatic determination of the molecular center for spherical-harmonic fitting can be overridden by adding a field to the Input (usually VAMP) SDF-file with the tag:

<SPHH_CENTER>

The center can be defined using Cartesian coordinates using an input line (immediately after the **SPHH_CENTER** tag) of the format:

```
Cartesian    x.xx    y.yy    z.zz
```

where ***x.xx***, ***y.yy*** and ***z.zz*** are the *x*, *y*, and *z*-coordinates, respectively. The capitalization of "**Cartesian**" is required.

Alternatively, a list of atoms can be given using the format

```
Atoms   n1   n2   n3   n4   n5   n6  ….
```

where ***n1*** etc. are the numbers of the atoms to be used to calculate the center of gravity. The capitalization of "**Atoms**" is required and the list of atoms is limited to one line.

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 (also 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.
RIF table file	`User defined`	An ascii, comma-separated file that contains a line of the standard rotationally invariant fingerprint (RIF [35]) 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, [37] 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 10.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 10: The headers and titles, atomic coordinates and bond definitions from a VAMP .sdf file. The format follows the MDL definition. [32]

<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	[17]	YES	YES	YES
AM1	[18]	YES	YES	YES
PM3	[19]	YES	YES	YES
MNDO/c	[37]	YES	YES	NO
MNDO/d	[20]	NO	YES	NO
AM1*	[21]	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-eigenvectors 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'™ 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™ 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™ 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

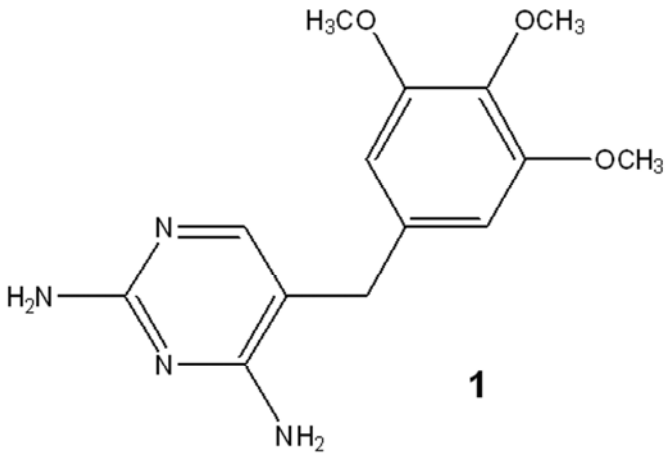
The file Vhamil.par is used by VAMP to define the available Hamiltonians and elements and supply the parameters. This file is also used by ParaSurf™ for the same purpose. A Vhamil.par file for standard Hamiltonians and elements is supplied with the ParaSurf™ program. In order to be sure that all Hamiltonians 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 11 shows the output for a calculation using the options **`surf=wrap fit=sphh translate`** for trimethoprim, 1.



1

```
<> ParaSurf'09, Revision A1
<> Copyright (c) 2006,2007,2008,2009 Friedrich-Alexander-Universitaet
                 Erlangen-Nuernberg and Cepos InSilico Ltd.
                 All rights reserved.

<> Input = trimethoprim.sdf


<<>> Molecule    1 of    1 <<>>



<> Program options :

   Using shrink-wrap isocontour surface
   Fitting surface to spherical harmonics
   Translations for spherical-harmonic fits: 1 step of   0.5000 Angstrom in each direction.
   Using an isodensity surface contour
   Isodensity value = 0.2000E-04 electrons/Angstrom**3
   Triangulation mesh =    0.20 Angstrom
   Using multipole electrostatics
```

Figure 11: ParaSurf® output for trimethoprim, **1**, using a spherical-harmonic surface.

```
<> AM1   calculation for  Trimethoprim

<> Translated spherical-harmonic fits:
     dx       dy       dz      rmsd
   0.0000  0.0000  0.0000  0.4034
   0.5000  0.0000  0.0000  0.5684
  -0.5000  0.0000  0.0000  0.5002
   0.0000  0.5000  0.0000  0.5611
   0.0000 -0.5000  0.0000  0.5193
   0.0000  0.0000  0.5000  0.5689
   0.0000  0.0000 -0.5000  0.4283


<> Fitting surface to spherical harmonics
<> Order(l)    RMSD

      0       1.92526847
      1       1.96114689
      2       1.55521125
      3       1.10609483
      4       0.93107676
      5       0.70605297
      6       0.63661488
      7       0.57077524
      8       0.52400109
      9       0.50050583
     10       0.47261493
     11       0.44416316
     12       0.41920775
     13       0.40320743
     14       0.39308983
     15       0.38341761

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

      0      11.06621848
      1      11.02831889
      2       8.63330698
      3       6.86247364
      4       5.49824707
      5       4.58527334
      6       4.17139337
      7       3.45052537
      8       3.12581239
      9       2.77798689
     10       2.36033975
     11       2.07232627
     12       1.90441930
     13       1.72381187
     14       1.64582625
     15       1.46855812
     16       1.27875373
     17       1.07480393
     18       0.93507876
     19       0.88299081
     20       0.82791747
```

Figure 11: continued

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

      0       56.96181478
      1       50.08877418
      2       45.40744632
      3       43.50297274
      4       40.06772200
      5       35.49615651
      6       32.77544861
      7       26.87818719
      8       23.10705894
      9       19.60935653
     10       17.98417050
     11       16.21352138
     12       15.12917785
     13       14.62643122
     14       13.87383640
     15       13.44294950
     16       13.23244532
     17       12.62943635
     18       12.27106930
     19       12.27106930
     20       12.27106930

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

      0       12.18668274
      1       11.86538321
      2       11.74571609
      3        9.50312285
      4        8.72650603
      5        7.28921579
      6        7.13957211
      7        6.79022442
      8        6.48006338
      9        6.02636118
     10        5.73169119
     11        5.46777405
     12        5.18598187
     13        4.51689005
     14        4.11336321
     15        3.92017066
     16        3.68134487
     17        3.60264328
     18        3.40103297
     19        3.23507723
     20        3.10515890
```

Figure 11: continued

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

      0       1.69554330
      1       1.67580613
      2       1.55177489
      3       1.39498502
      4       1.20933761
      5       1.11363208
      6       1.00041103
      7       0.83474040
      8       0.81023222
      9       0.77273063
     10       0.72281272
     11       0.63576134
     12       0.58709249
     13       0.51478216
     14       0.50567303
     15       0.50567303
     16       0.50567303
     17       0.50567303
     18       0.50567303
     19       0.50567303
     20       0.50567303

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

      0       0.02367100
      1       0.01665303
      2       0.01371808
      3       0.01112413
      4       0.00912405
      5       0.00817327
      6       0.00769192
      7       0.00722555
      8       0.00694967
      9       0.00643212
     10       0.00588304
     11       0.00574182
     12       0.00531887
     13       0.00531378
     14       0.00521159
     15       0.00514920
     16       0.00514920
     17       0.00514920
     18       0.00514920
     19       0.00514920
     20       0.00514920

<> Property ranges:
   Density    :   0.3567E-05 to   0.9969E-04
   IE(l)      :       391.05 to       671.20
   EA(l)      :      -108.56 to       -38.29
   MEP        :       -48.50 to        16.80
   Alpha(l)   :       0.2368 to       0.3374
   Field(N)   :       -10.95 to         2.43
```

Figure 11: continued

```
<> Descriptors :

    Dipole moment        :       1.2467 Debye
    Dipolar density      :   0.1933E-02 Debye.Angstrom**-3
    Molecular pol.       :     128.5408 Angstrom**3
    Molecular weight     :       290.32
    Globularity          :       0.7689
    Total surface area   :       469.51 Angstrom**2
    Molecular volume     :       644.94 Angstrom**3

    Most positive MEP    :        16.80 kcal/mol
    Most negative MEP    :       -48.50 kcal/mol
    Mean +ve MEP         :         5.59 kcal/mol
    Mean -ve MEP         :       -10.80 kcal/mol
    Mean MEP             :        -3.13 kcal/mol
    MEP range            :        65.30 kcal/mol
    MEP +ve Variance     :        10.80 (kcal/mol)**2
    MEP -ve Variance     :        94.38 (kcal/mol)**2
    MEP total variance   :       105.18 (kcal/mol)**2
    MEP balance parameter:       0.0921
    MEP balance*variance :       9.6898 kcal/mol
    MEP skewness         :      -1.1813
    MEP kurtosis         :       1.3859
    Integral MEP         : -1166.52     kcal.Angstrom**2/mol

    Maximum IE(l)        :       671.20 kcal/mol
    Minimum IE(l)        :       391.05 kcal/mol
    Mean IE(l)           :       475.70 kcal/mol
    IE(l) range          :       280.15 kcal/mol
    IE(l) variance       :      3233.28 (kcal/mol)**2
    IE(l) skewness       :       0.6770
    IE(l) kurtosis       :      -0.2281
    Integral IE(l)       :  9650.55     eV.Angstrom**2

    Maximum EA(l)        :       -38.29 kcal/mol
    Minimum EA(l)        :      -108.56 kcal/mol
    Mean +ve EA(l)       :         0.00 kcal/mol
    Mean -ve EA(l)       :       -93.87 kcal/mol
    Mean EA(l)           :       -93.87 kcal/mol
    EA(l) range          :        70.27 kcal/mol
    EA(l) +ve variance   :         0.00 (kcal/mol)**2
    EA(l) -ve variance   :       142.48 (kcal/mol)**2
    EA(l) total variance :       142.48 (kcal/mol)**2
    EA(l) skewness       :       1.7822
    EA(l) kurtosis       :       4.1719
    Integral EA(l)       : -1913.53     eV.Angstrom**2
    EA(l) balance param. :       0.0000
    Fraction pos. EA(l)  :       1.0000 ( =  469.51 Angstrom**2)

    Max. local Eneg.     :       299.60 kcal/mol
    Min. local Eneg.     :       143.17 kcal/mol
    Mean local Eneg.     :       190.92 kcal/mol
    Local Eneg. range    :       156.43 kcal/mol
    Local Eneg. variance :       958.81 (kcal/mol)**2
    Local Eneg. skewness :         0.82
    Local Eneg. kurtosis :         0.02
    Integral local Eneg. :  3868.51     eV.Angstrom**2

    Max. local hardness  :       371.59 kcal/mol
    Min. local hardness  :       247.44 kcal/mol
    Mean local hardness  :       284.79 kcal/mol
    Local hard. range    :       124.15 kcal/mol
    Local hard. variance :       729.07 (kcal/mol)**2
    Local hard. skewness :         0.58
    Local hard. kurtosis :        -0.48
    Integral local Hard. :  5782.04     eV.Angstrom**2
```

Figure 11: continued

```
   Maximum field normal :     2.43 kcal/mol.Angstrom
   Minimum field normal :   -10.95 kcal/mol.Angstrom
   Mean field           :    -0.63 kcal/mol.Angstrom
   Field range          :    13.38 kcal/mol.Angstrom
   Total field variance :     2.83 (kcal/mol.Angstrom)**2
   +ve field variance   :     2.00 (kcal/mol.Angstrom)**2
   -ve field variance   :     3.50 (kcal/mol.Angstrom)**2
   Field balance param. :     0.23
   Field skew           :     2.04
   Field kurtosis       :   2.702
   Integral F(N)        :  -232.8     kcal.Angstrom/mol
   Integral F(N +ve)    :   151.9     kcal.Angstrom/mol
   Integral F(N -ve)    :  -384.8     kcal.Angstrom/mol
   Integral |F(N)|      :   536.7     kcal.Angstrom/mol

<> Spherical-Harmonic Hybridization:

   Shape hybrids        :
  17.575999    1.110912    3.450834    2.848845    1.410614    1.601925
   0.752083    0.688161    0.462308    0.389757    0.387720    0.318746
   0.254804    0.212811    0.209025    0.200487

   MEP hybrids          :
  13.221169    4.733832   26.182832   18.490714   13.995110   10.726190
   7.222067    8.122172    4.622533    4.919965    4.395602    3.581424
   2.541402    2.738680    1.628924    2.033210    1.920261    1.871571
   1.474479    1.036117    1.018238

   IE(l) hybrids        :
  1698.2772     90.2239     81.3709     62.4721     59.3232     54.5834
    53.5801     54.9031     40.0301     44.6893     25.5340     24.8825
    19.8329     15.9870     18.4079     15.7295     14.4314     16.4220
    14.7815      0.0000      0.0000

   EA(l) hybrids        :
   324.0505      6.1912     12.8642     25.9653     13.9372     16.5122
     8.5512      9.0229      8.0845      7.5242      6.2268      5.4009
     5.1364      7.5775      5.6359      4.4817      4.8393      3.7347
     3.7375      3.6142      3.4498

   Alpha(l) hybrids     :
 1.01354601  0.05241076  0.03341442  0.02993772  0.02191758  0.01497612
 0.01186878  0.00889862  0.00763997  0.00888553  0.00766420  0.00636663
 0.00683106  0.00625184  0.00851872  0.00656541  0.00000000  0.00000000
 0.00000000  0.00000000  0.00000000

   Field(N) hybrids     :
     2.9787      2.0853      2.9904      3.2154      3.0158      1.7511
     1.7779      2.1063      1.2183      1.0505      0.8405      1.0237
     0.7182      0.9211      0.6071      0.0000      0.0000      0.0000
     0.0000      0.0000      0.0000

<> Standard rotationally invariant fingerprint:(L. Mavridis, B. D. Hudson
   and D. W. Ritchie, J. Chem. Inf. Model., 2007, 47, 1787-1796.)

     4.19237     1.05400     1.85764     1.68785     1.18769
     1.26567    0.867227     3.63609     2.17574     5.11692
     4.30008     3.74100     3.27509     2.68739     2.84994
     2.15001     2.21810     2.09657     1.89247     1.59418
     41.2102     9.49863     9.02058     7.90393     7.70216
     7.38806     7.31984     18.0014     2.48822     3.58666
     5.09562     3.73325     4.06352     2.92425     1.00675
     0.228934    0.182796    0.173025    0.148046    0.122377
     0.108944    1.72589     1.44405     1.72928     1.79316
     1.73661     1.32328     1.33337     1.45132     1.10378
     1.02495     0.916811    1.01180     0.847483
```

Figure 11: continued

<> Atomic surface properties:

Atom		Area	MEP max	MEP min	IE(l) max	IE(l) min	EA(l) max	EA(l) min	mean pol.	Field(N) max	Field(N) min
C	1	0.000									
O	2	0.073	-38.63	-41.62	535.49	532.95	-77.41	-79.32	0.260	-6.89	-7.51
C	3	3.380	-6.38	-46.92	578.33	474.12	-41.12	-89.61	0.307	1.65	-7.54
C	4	1.259	-5.17	-15.60	573.38	498.18	-67.85	-90.38	0.322	-1.32	-2.72
C	5	0.699	-8.33	-13.90	573.55	531.88	-84.05	-92.54	0.320	-1.85	-3.05
C	6	0.000									
C	7	0.803	-9.81	-15.04	559.27	516.68	-71.66	-91.24	0.319	-0.88	-2.72
C	8	4.155	-1.37	-21.94	585.31	484.68	-48.35	-95.74	0.295	2.43	-1.73
N	9	4.183	-14.72	-32.41	535.55	452.61	-79.85	-104.91	0.279	1.59	-6.57
C	10	10.328	-1.84	-27.18	633.68	532.90	-38.86	-86.78	0.284	1.52	-2.98
N	11	0.000									
N	12	1.441	-14.69	-33.29	538.04	472.64	-74.72	-98.64	0.268	0.31	-5.93
C	13	6.263	-8.51	-26.51	637.06	512.83	-38.29	-82.55	0.286	2.16	-7.40
N	14	0.000									
C	15	2.104	-7.00	-15.39	589.73	496.41	-64.28	-92.21	0.316	-1.22	-10.76
C	16	3.888	-11.81	-43.40	569.55	479.96	-40.72	-85.82	0.310	-0.57	-8.44
O	17	0.000									
C	18	0.000									
C	19	6.191	-17.33	-48.50	583.22	465.65	-46.33	-86.05	0.315	-0.75	-7.77
O	20	2.039	-31.11	-44.39	532.30	445.05	-75.31	-93.15	0.247	-4.08	-7.49
C	21	0.000									
H	22	32.636	13.06	-39.48	560.92	405.94	-82.21	-99.66	0.297	1.30	-6.93
H	23	21.586	14.07	-20.13	561.93	408.35	-83.75	-96.08	0.294	1.53	-3.51
H	24	24.162	14.04	-30.19	567.47	407.17	-68.79	-95.29	0.292	1.32	-4.39
H	25	6.870	11.16	-5.32	527.12	425.97	-83.14	-95.64	0.288	1.11	-2.07
H	26	18.505	7.91	-7.43	535.39	400.22	-87.77	-100.66	0.299	1.23	-2.63
H	27	17.884	7.35	-20.62	543.64	399.67	-68.59	-100.81	0.303	1.04	-5.35
H	28	26.817	8.86	-27.87	536.84	413.57	-60.23	-103.43	0.284	1.94	-4.90
H	29	33.032	16.80	-28.67	671.20	474.61	-71.99	-107.85	0.248	2.20	-6.24
H	30	32.852	16.44	-28.20	669.03	476.76	-72.53	-107.92	0.241	2.18	-5.50
H	31	33.925	11.43	-28.95	657.35	471.32	-67.58	-108.56	0.244	2.17	-8.10
H	32	9.935	11.30	-27.11	642.05	472.51	-73.10	-99.96	0.259	2.29	-10.95
H	33	5.296	9.88	-13.60	504.92	434.61	-73.74	-96.58	0.295	1.34	-8.06
H	34	29.289	11.33	-40.37	565.02	406.91	-80.26	-99.68	0.291	1.32	-8.75
H	35	23.555	11.45	-33.33	567.66	407.69	-72.90	-95.65	0.292	1.33	-5.91
H	36	23.014	11.40	-13.11	563.89	407.56	-81.58	-99.03	0.295	1.28	-3.93
H	37	33.964	4.59	-38.61	554.62	394.98	-87.16	-108.55	0.294	0.57	-6.68
H	38	18.880	4.19	-40.99	558.72	396.48	-84.55	-108.23	0.298	0.50	-7.12
H	39	26.446	3.97	-34.03	550.48	391.05	-86.20	-107.61	0.296	0.55	-5.39
Total		465.455									

Figure 11: continued

```
<> Stationary points on the molecular surface (A. Jakobi, H. Mauser
    and T. Clark, J. Mol. Model., 2008, 14, 547-558)

          x           y           z      value

 <>       5 MEP Maxima      :
      4.7936      2.2899     -2.2084     11.45
      2.4965      4.0712      5.7445     12.67
      1.0770      5.8131      5.4602     16.80
     -2.2439     -2.1297      2.8572     11.43
     -5.5622     -3.4081     -0.8626     14.07
 <>       3 MEP Minima      :
     -0.3945     -4.1333     -2.6456    -48.50
      1.9197     -2.8411     -3.4453    -42.54
      2.5161     -2.1045     -3.3008    -43.40

 <>       3 IEL Maxima      :
     -1.4144      5.4218      3.9311     671.2
     -3.4005      1.4346      2.5848     657.3
      2.5158      3.2319      3.2339     642.2
 <>      13 IEL Minima      :
      2.1418      1.5561     -4.5855     407.7
      0.4940     -4.6999      1.7295     391.5
     -4.0895     -2.9712      1.8500     408.4
      3.6525      1.2975      1.2233     407.6
      0.1300     -4.7857      1.5109     391.0
     -2.8688     -6.9328     -1.6690     405.9
      0.1327      5.3805     -2.2317     413.6
     -3.7288     -1.8541     -3.8279     407.2
      2.7277     -6.5570     -2.9447     395.0
     -4.8206      1.4292      0.2199     399.7
      4.7699     -2.4304      0.0000     396.5
     -1.4313      2.8849     -3.5503     400.2
      6.1623     -1.4832     -2.6349     406.9

 <>       4 EAL Maxima      :
     -2.9928      2.6947      2.3379    -38.29
      0.4747     -0.8240     -3.7107    -40.72
     -1.6927      4.7425      3.2127    -38.86
     -0.8097     -1.8229     -3.5122    -41.12
 <>      10 EAL Minima      :
      0.2649     -5.2698     -2.4638    -106.9
      1.8583      6.9358      2.6243    -107.9
      3.2295     -3.3153     -3.2557    -108.5
     -1.0575      0.1415      5.3927    -108.6
      0.7838     -3.6877      1.7626    -104.7
      0.1522     -5.6007     -1.7682    -108.0
      0.9842     -4.8653     -3.4860    -106.7
      0.4813      2.5816      6.0814    -107.9
      0.9630     -3.4415      1.7646    -104.7
      0.4219     -4.0142      1.6785    -104.7

 <>       4 Alpha(l) Maxima :
      0.0000      0.0000     -3.4565    0.3237
     -1.3158     -0.7255     -3.2393    0.3286
     -2.9170      3.4094      0.9959    0.3259
      0.4062     -2.4010     -3.8398    0.3374
 <>       5 Alpha(l) Minima :
      0.7758     -4.0667     -3.4739    0.2387
      0.0553     -4.3637     -2.7928    0.2397
      0.6923      3.3892      6.7598    0.2372
     -1.5392     -0.4236      5.3172    0.2411
      0.4406     -4.1863     -3.2471    0.2368

 <>       0 F(N) Maxima      :
 <>       0 F(N) Minima      :

 <> ParaSurf used        16.04 seconds CPU time
```

Figure 11: continued

After printing the program options, ParaSurf'09™ prints the shift in coordinates of the center and the RMSD fits for the surface requested by the `translate` option. For speed, these fits use a lower number of surface points than the full fits that follow and are only calculated up to order six. The translated spherical-harmonic coefficients are printed in the output SDF file for use by ParaFit™. ParaSurf'09™ then moves on to fit the calculated shrink-wrap surface at full resolution for each of the local properties. It 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 five 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 five local properties. The coefficients are listed by increasing l starting from zero. The standard rotationally invariant fingerprint (RIF) [**35**] is printed. Note that the individual RIF-values correspond to the square roots of the hybridization coefficients from the tables above and that the RIF definition has been expanded to include hybridization coefficients of the field normal to the surface (the last 13 elements).

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, EA_L and F_N 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 ones 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.

The maxima and minima of the local properties selected according to the criteria outlined in reference **34** are then listed. These points are defined by their Cartesian coordinates and the corresponding values of the local property. In this example, no significant maxima and minima were found for the field normal to the surface. Generally, more maxima and minima are found for isodensity surfaces than for spherical-harmonic ones.

3.4.2 For a marching-cube surface

Figure 12 shows the output for a calculation using the options **`surf=cube`** for trimethoprim.

```
<> ParaSurf'09, Revision A1
<> Copyright (c) 2006,2007,2008,2009 Friedrich-Alexander-Universitaet
                 Erlangen-Nuernberg and Cepos InSilico Ltd.
                 All rights reserved.

<> Input = trimethoprim.sdf


<<>> Molecule    1 of    1 <<>>



<> Program options :

   Using marching-cube isodensity surface
   Surface fitting turned off
   Using an isodensity surface contour
   Isodensity value = 0.3000E-03 electrons/Angstrom**3
   Triangulation mesh =    0.20 Angstrom
   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
   Field(N)   :       -29.18 to        18.88
```

Figure 12: 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)**2
    MEP -ve Variance      :        239.92 (kcal/mol)**2
    MEP total variance    :        271.53 (kcal/mol)**2
    MEP balance parameter:         0.1028
    MEP balance*variance :       27.9261 kcal/mol
    MEP skewness          :       -1.0234
    MEP kurtosis          :        0.6111
    Integral MEP          : -1674.26     kcal.Angstrom**2/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)**2
    IE(l) skewness        :        0.4205
    IE(l) kurtosis        :       -0.7616
    Integral IE(l)        :  7764.76     eV.Angstrom**2

    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)**2
    EA(l) -ve variance    :        276.47 (kcal/mol)**2
    EA(l) total variance :        276.47 (kcal/mol)**2
    EA(l) skewness        :        1.4621
    EA(l) kurtosis        :        1.5753
    Integral EA(l)        : -1438.92     eV.Angstrom**2
    EA(l) balance param. :        0.0000
    Fraction pos. EA(l)   :        1.0000 ( =  369.79 Angstrom**2)

    Max. local Eneg.      :        290.14 kcal/mol
    Min. local Eneg.      :        143.75 kcal/mol
    Mean local Eneg.      :        198.61 kcal/mol
    Local Eneg. range     :        146.39 kcal/mol
    Local Eneg. variance :       1205.84 (kcal/mol)**2
    Local Eneg. skewness :          0.52
    Local Eneg. kurtosis :         -0.78
    Integral local Eneg. :  3162.92     eV.Angstrom**2
```

Figure 12: continued

```
   Max. local hardness  :       371.30 kcal/mol
   Min. local hardness  :       247.91 kcal/mol
   Mean local hardness  :       287.69 kcal/mol
   Local hard. range    :       123.39 kcal/mol
   Local hard. variance :       724.88 (kcal/mol)**2
   Local hard. skewness :         0.45
   Local hard. kurtosis :        -0.66
   Integral local Hard. :  4601.84     eV.Angstrom**2

   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**6
   Alpha(l) skewness    :      -0.8040
   Alpha(l) kurtosis    :      -0.3752
   Integral Alpha(l)    :  104.483     Angstrom**5

   Maximum field normal :    18.88 kcal/mol.Angstrom
   Minimum field normal :   -29.18 kcal/mol.Angstrom
   Mean field           :    -0.85 kcal/mol.Angstrom
   Field range          :    48.05 kcal/mol.Angstrom
   Total field variance :    17.27 (kcal/mol.Angstrom)**2
   +ve field variance   :    11.03 (kcal/mol.Angstrom)**2
   -ve field variance   :    23.22 (kcal/mol.Angstrom)**2
   Field balance param. :     0.22
   Field skew           :     2.80
   Field kurtosis       :   7.752
   Integral F(N)        :  -298.0     kcal.Angstrom/mol
   Integral F(N +ve)    :   312.3     kcal.Angstrom/mol
   Integral F(N -ve)    :  -610.3     kcal.Angstrom/mol
   Integral |F(N)|      :   922.6     kcal.Angstrom/mol

<> Atomic surface properties:

  Atom    Area       MEP            IE(l)           EA(l)        mean      Field(N)
                  max     min     max     min     max     min     pol.    max     min
C     1  0.257  -23.83  -46.81  569.04  546.04  -81.50  -92.65   0.268   -5.92  -12.94
O     2  3.658  -15.50  -69.70  594.61  456.71  -63.66  -81.53   0.269    9.41  -16.81
C     3  6.490   -7.36  -64.78  643.17  499.72  -30.09  -99.43   0.304   12.85   -9.93
C     4  2.166   -3.31  -19.76  632.00  493.47  -39.95 -100.75   0.316   -0.27   -5.07
C     5  1.600   -3.37  -18.28  633.56  547.00  -53.42 -100.30   0.313   -0.43   -3.87
C     6  0.000
C     7  2.042   -4.01  -22.33  605.70  512.44  -49.80  -91.06   0.317    6.32   -8.09
C     8  5.665    4.11  -28.25  638.27  488.34  -35.80  -88.30   0.288    6.91   -4.32
N     9  6.693  -19.84  -58.79  571.34  417.76  -54.71 -103.24   0.260    7.37  -19.27
C    10  9.411   -0.86  -46.23  654.76  543.32  -41.17  -81.09   0.279    5.81   -7.91
N    11  0.537  -46.53  -53.00  615.72  593.86  -60.28  -78.98   0.276   -2.51  -11.48
N    12  6.122  -16.63  -55.66  571.30  417.23  -51.01  -98.93   0.247   14.20  -19.99
C    13  7.570  -10.11  -44.07  644.87  527.82  -37.47  -82.80   0.284   12.28  -16.70
N    14  0.713  -41.89  -57.27  618.93  590.90  -64.88  -81.21   0.287  -11.29  -29.18
C    15  4.127   -8.07  -22.33  640.32  494.13  -30.91 -100.89   0.314   18.88   -6.00
C    16  5.886  -15.50  -60.73  641.06  507.48  -29.09  -94.41   0.307   14.49  -15.05
O    17  1.261  -18.70  -69.88  567.04  464.95  -64.27  -87.64   0.252   -4.23  -25.48
C    18  0.289  -16.74  -56.71  573.39  531.54  -74.34  -94.89   0.267   -5.84  -19.88
C    19  5.580  -15.52  -60.03  617.79  492.33  -39.25  -96.42   0.314    4.02   -5.35
O    20  3.960  -31.04  -63.94  579.08  438.53  -69.08  -94.47   0.265   -1.78  -13.88
C    21  0.543  -26.28  -54.82  563.68  530.83  -90.92 -106.12   0.269   -4.37  -10.55
H    22 20.848   22.24  -43.97  561.11  407.48  -83.45  -97.55   0.297    2.66  -13.21
H    23 16.018   22.21  -47.65  566.01  408.34  -70.47  -97.22   0.294    6.30   -5.56
H    24 16.235   22.21  -45.75  567.60  407.90  -66.77  -96.72   0.290    2.62   -9.87
H    25  7.143   16.35   -8.05  537.67  429.84  -70.41  -97.59   0.288    1.88   -4.03
H    26 13.545   13.09   -5.83  579.22  401.23  -85.87 -100.66   0.299    2.24   -3.30
H    27 13.114   11.65  -38.33  611.02  400.48  -74.22 -100.87   0.301    1.87  -13.51
H    28 17.462   13.34  -29.83  533.62  415.38  -56.04 -100.30   0.282    2.54   -5.68
H    29 20.093   24.36  -44.05  639.92  488.09  -72.40 -107.69   0.247    3.99  -15.00
H    30 20.380   24.82  -48.21  644.36  488.15  -70.42 -107.77   0.241    4.31  -17.04
H    31 20.025   22.73  -51.85  642.47  483.89  -67.45 -108.35   0.245    9.76  -27.09
H    32 10.792   22.31  -49.42  644.46  478.52  -79.21 -102.13   0.259    7.86  -26.66
H    33  7.935   15.01  -21.60  523.98  429.34  -65.58  -98.60   0.294    8.18   -7.77
H    34 20.353   17.51  -31.87  560.07  408.27  -87.93  -99.07   0.290    2.42  -12.17
H    35 16.221   17.81  -47.09  565.56  408.31  -66.99  -96.49   0.291    2.42  -12.12
H    36 16.250   17.77  -37.52  557.41  408.14  -69.22  -96.81   0.294   12.54   -6.31
H    37 20.708    8.06  -41.68  545.33  396.15  -95.89 -109.82   0.294    0.99   -9.80
H    38 16.217    8.06  -54.42  595.55  394.15  -82.57 -109.54   0.296    9.51  -11.49
H    39 18.651    7.90  -40.13  586.15  392.35  -74.89 -109.41   0.295    2.99   -4.47

Total   366.558
```

Figure 12: continued

```
<> Stationary points on the molecular surface (A. Jakobi, H. Mauser
    and T. Clark, J. Mol. Model., 2008, 14, 547-558)

         x           y           z       value

 <>      11 MEP Maxima      :
     -3.2288      1.8355     -3.6285     7.168
      3.6569      1.8897     -3.0618     17.81
     -1.7098      3.4897     -3.2285     12.76
     -0.8648      3.3397     -3.2785     13.34
     -5.4931     -3.9386     -2.2285     22.24
     -4.2431      0.2897     -2.4285     9.660
      3.0069     -5.0936     -0.8285     8.063
     -3.7931     -0.7603      0.1132     6.919
     -1.9431     -1.8603      1.1798     19.01
     -2.3764     -1.7103      1.7715     22.73
      1.0569      5.1397      4.3215     24.82
 <>      12 MEP Minima      :
      1.4569     -3.0103     -3.9285    -63.17
      2.0184     -1.7936     -4.0285    -69.88
     -1.4630     -4.0603     -3.4285    -69.70
     -1.6931     -4.0603     -0.8285    -55.73
     -0.7431      5.5397      0.1165    -53.25
     -0.3431      5.5897      0.1498    -57.47
      0.6569      5.2397      0.2882    -54.40
      0.0569      5.5397      0.2082    -58.79
     -3.7431      0.5680      1.3882    -55.54
     -3.6931      0.7730      1.3265    -57.27
     -1.5431      4.9397      2.9548    -53.00
     -1.5431      2.1564      3.4715    -55.66

 <>      10 IEL Maxima      :
     -0.9431     -0.9603     -3.8285     641.1
     -0.3098     -1.4603     -0.4785     643.2
      0.5569      2.1397      0.6548     654.8
     -3.4431      1.4314      0.9715     625.4
     -2.2931      3.5980      0.9515     650.5
     -0.2431     -0.0103      1.0765     644.9
     -1.1431      5.3397      1.6998     634.2
     -3.0431      1.5397      2.1715     635.0
      1.4569      3.2397      2.3798     644.4
     -1.7264      3.7397      3.1965     634.5
 <>      17 IEL Minima      :
      1.2569      0.8230     -4.9285     411.0
      1.6501      1.3397     -4.9785     408.3
     -4.2098     -2.2603     -4.4285     407.9
     -1.9431      2.5980     -3.9785     401.2
      2.3236     -6.2603     -3.6285     396.2
      5.2069     -1.0936     -3.6285     408.3
     -0.7431      3.9847     -3.2285     418.8
     -3.7098     -6.3603     -2.6785     407.5
     -0.3931      4.7397     -2.8285     415.4
      3.6569     -2.5603     -1.2285     394.2
     -4.7598      1.1397     -1.0525     400.5
      2.9019      0.9397     -0.2452     408.2
      3.0402      1.1397     -0.3285     408.1
      0.0569     -4.8436      0.2715     392.3
      0.3069      5.4397      0.1632     417.8
     -4.2014     -3.0603      0.4215     408.3
     -0.9681      1.6564      3.4715     417.2
```

Figure 12: continued

```
<>      5 EAL Maxima       :
     -0.0931     -1.2270     -4.2285   -29.09
     -1.5231     -1.7186     -3.9785   -30.74
     -1.4098     -1.8603     -4.0285   -30.09
     -2.7431      3.7564     -0.4835   -35.80
      0.7986     -1.0603     -0.4785   -29.74
 <>     15 EAL Minima       :
      0.8569     -4.9186     -3.9785   -106.5
      1.1736     -4.6853     -4.1285   -106.5
      1.6569     -4.3820     -4.1785   -106.7
      1.8319     -4.1770     -4.1285   -106.9
      2.0569     -3.9603     -4.0285   -107.7
      2.4819     -3.6270     -3.7785   -109.8
     -0.0431     -5.5103     -2.4178   -109.4
      3.9569     -3.9770     -2.0285   -106.9
      2.4569     -2.6270     -0.2785   -106.7
      0.0569      5.3397     -0.4285   -103.2
      0.7736     -5.7603     -0.0285   -104.7
      0.6286     -4.4603      0.5715   -104.6
      1.2768      6.2897      1.7715   -107.7
     -1.8024     -0.2436      3.7548   -108.4
      0.0569      2.9147      4.9548   -107.8

 <>      9 Alpha(l) Maxima :
     -0.2598     -2.4603     -4.3285   0.3300
      0.0569     -2.3488     -4.3452   0.3301
     -1.9431     -1.2153     -3.8285   0.3246
     -0.3431     -0.0603     -4.0285   0.3211
     -1.9431     -1.0853     -3.8118   0.3257
      0.1286     -2.2603     -0.5285   0.3240
      0.3569     -0.2936     -0.2285   0.3192
     -3.1431      2.9397     -0.1785   0.3292
      0.5069      1.6814     -0.0285   0.3187
 <>      9 Alpha(l) Minima :
      2.2569     -1.7603     -4.0285   0.2343
     -1.7523     -4.0603     -3.5285   0.2365
     -1.6931     -4.0603     -0.8285   0.2373
      2.1569      5.2990      2.5715   0.2441
      1.5402      6.3397      2.4475   0.2441
      1.0569      2.5397      3.1715   0.2409
     -2.0848     -0.4603      3.7215   0.2411
     -0.9681      1.6564      3.4715   0.2288
      0.0569      3.3147      5.1548   0.2372

 <>      4 F(N) Maxima       :
      2.0819     -1.2803     -1.0452    13.47
     -1.5098     -3.7103     -0.7285    12.15
      0.2619     -0.0603     -0.1785    18.11
      0.2069      1.1397      2.2048    13.48
 <>     10 F(N) Minima       :
      1.3069     -3.5820     -4.1285   -12.39
      2.1786     -1.8603     -3.9785   -24.61
     -0.1431     -4.6103     -3.4527   -13.19
     -1.7306     -4.2853     -3.4118   -16.57
     -1.7413     -4.6603     -1.2285   -10.34
     -0.5598     -0.9053      0.3882   -12.28
     -0.0848      5.5897      0.3715   -18.92
     -3.6931      0.7730      1.3265   -28.42
     -1.3431      5.2397      3.1439   -11.61
     -1.5431      2.4680      3.4715   -19.67

 <> ParaSurf used        11.58 seconds CPU time
```

Figure 12: 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 13 and 14 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 12 except that an additional Shannon entropy block is printed after the descriptors, as shown in Figure 13:

```
                                    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
```

Figure 13: 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 14:

Shannon-entropy analysis :

Atom		Area	Shannon Entropy Internal				Shannon Entropy External			
			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

Figure 14: 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        +/+        -/-         +/-       IE(L)       EA(L)

Total       :     0.8924     0.5535     0.6968     0.6675     0.3740     0.8364

1. Quartal :      0.9039     0.3785     0.6334     0.7132     0.3319     0.8695
2. Quartal :      0.8861     0.3713     0.6761     0.5637     0.3086     0.8366
3. Quartal :      0.9348     0.7294     0.8152     0.6597     0.3651     0.9109
4. Quartal :      0.8450     0.7349     0.6623     0.7334     0.4902     0.7286


Entropies    :      Shape        +/+        -/-         +/-       IE(L)       EA(L)

Total       :     0.0962     0.3210     0.3078     0.3348     0.3314     0.2225

1. Quartal :      0.0533     0.3390     0.3516     0.3261     0.3203     0.2007
2. Quartal :      0.1802     0.3338     0.3391     0.3632     0.3102     0.2354
3. Quartal :      0.1112     0.3010     0.2296     0.3382     0.3329     0.1431
4. Quartal :      0.0399     0.3103     0.3108     0.3118     0.3621     0.3107


Corr.Coeff.  :
(R)          :      0.7914    -0.7041      0.5585     0.9201     0.9184     0.5880


Field Similarities :           Raw Data                                  Scaled
+/+         -/-         +/-             +/+         -/-         +/-

Total       :     0.7497     0.6651     0.5998        0.7418     0.7024     0.5768

1. Quartal :      0.6836     0.5321     0.4983        0.7843     0.8004     0.6583
2. Quartal :      0.5508     0.6104     0.5652        0.9313     0.9307     0.9169
3. Quartal :      0.8520     0.7296     0.7738        0.5967     0.8438     0.5746
4. Quartal :      0.9125     0.7882     0.5617        0.6522     0.2349     0.1574


Field Entropies    :           Raw Data                                  Scaled
+/+         -/-         +/-             +/+         -/-         +/-

Total       :     0.2592     0.3223     0.3300        0.2458     0.1861     0.1858

1. Quartal :      0.3377     0.3606     0.3559        0.2370     0.2517     0.1957
2. Quartal :      0.3636     0.3573     0.3637        0.1105     0.1161     0.1326
3. Quartal :      0.1947     0.3109     0.2440        0.3357     0.2179     0.2426
4. Quartal :      0.1407     0.2606     0.3563        0.3000     0.1586     0.1722
```

Figure 15: 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. Additionally, ParaSurf'09™ calculates the "entropies" and the correlation coefficients between the two molecules for each of the autocorrelations. The "entropies" S are defined as

$$S = \frac{\sum_{i=1}^{N} p_i \log(p_i)}{N} \quad \text{where } p = \frac{a_2}{a_1} \tag{20}$$

where a_1 is the larger of the two autocorrelation values and a_2 is the smaller. *N* is the number of autocorrelation points considered. The “entropy” is zero for identical autocorrelations and has a maximum value of one.

A new feature in ParaSurf’09™ is that autocorrelations are also calculated for the electroststic field normal to the molecular surface. This field generally gives more highly resolved autocorreltions than the electrostatic potential and is less sensitive to the total charge of the molecule. The field autocorrelations are compared for +/+, +/- and -/- combinations of F_N, analogously to the potential. These comparisons are made both for the raw autocorrelations and for one that are shifted and scaled to occupy a range between zero and one.

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 CUBE	Shrink-wrap surface Marching-cube surface
<fit>	NONE ISO SPHH	No fitting, unsmoothed marching-cube surface Marching-cube surface corrected to ± 2% of the preset isodensity value Spherical-harmonic surface fit
<electrostatic model>	NAOPC MULTI	NAO-PC electrostatics Multipole electrostatics
<isodensity level>	*n.nn*	The target isodensity value in $e^{-}Å^{-3}$
<solvent probe radius>		The radius of the solvent probe used to calculate the SES or SAS
<triangulation mesh>		The mesh size used to triangulate the Surface

<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 spherical-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 six 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}

<SPHERICAL_HARMONIC_FIELD(N)> The FN values at the spherical-harmonic surface (l = 20) in kcal mol-1 Å-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, $\sigma_{tot}^2\nu$, γ_1^V, γ_2^V, $\int_V$ ("MEP ",7f10.2/10x, f10.2,5f10.4,2x,g12.6)
IE(l):	IE_L^{max}, IE_L^{min}, $\overline{IE}_L$, ΔIE_L, σ_{IE}^2, γ_1^{IE}, γ_2^{IE}, $\int_{IE}$ ("IE(l) ",5f10.2,2f10.4/12x,g12.6)

EA(l):	EA_L^{max}, EA_L^{min}, $\overline{EA_{L+}}$, $\overline{EA_{L-}}$, $\overline{EA_L}$, ΔEA_L, σ^2_{EA+}, σ^2_{EA-}, σ^2_{EA}, ν_{EA}, δA^+_{EA}, A^+_{EA}, γ_1^{EA}, γ_2^{EA}, $\int_{EA}$ ("EA(l) ",7f10.2/2f10.2,2f10.4,f10.2,2f10.4/12x,g12.6)
Eneg(l):	χ_L^{max}, χ_L^{min}, $\overline{\chi_L}$, $\Delta\chi_L$, σ^2_χ, γ_1^χ, γ_2^χ, $\int_\chi$ ("Eneg(l) ",5f10.2,2f10.4/12x,g12.6)
Hard(l):	η_L^{max}, η_L^{min}, $\overline{\eta_L}$, $\Delta\eta_L$, σ^2_η, γ_1^η, γ_2^η, $\int_\eta$ ("Hard(l) ",5f10.2,2f10.4/12x,g12.6)
Alpha(l):	α_L^{max}, α_L^{min}, α_L, $\Delta\alpha_L$, σ^2_α, γ_1^α, γ_2^α, $\int_\alpha$ ("Alpha(l) ",5f10.2,2f10.4/12x,g12.6)
F_N	F_N^{max}, F_N^{min}, ΔF_N, $\overline{F_N}$, σ^2_F, σ^2_{F+}, σ^2_{F-}, ν_F, $\gamma_1^{F_N}$, $\gamma_2^{F_N}$, $\int_{F_N}$, $\int^+_{F_N}$, $\int^-_{F_N}$ ("Field desc",7f10.4/" ",6f10.4)

Jobs that include Shannon entropy give two extra sets of descriptors:

Shannon(i):	H_{in}^{max}, H_{in}^{min}, $\overline{H_{in}}$, $\sigma^2_{H_{in}}$, $\int_{H_{in}}$ ("Shannon(i) ",4f10.4,f10.2,f10.4)
Shannon(e):	H_{ex}^{max}, H_{ex}^{min}, $\overline{H_{ex}}$, $\sigma^2_{H_{ex}}$, $\int_{H_{ex}}$ ("Shannon(e) ",4f10.4,f10.2,f10.4)

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)
<FIELD(N) HYBRIDS>	(20 coefficients, 6f12.4)

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)
<ATOMIC SURFACE FIELD(N) MAXIMA>	FN maxima	(10f8.2)
<ATOMIC SURFACE FIELD(N) MINIMA>	FN minima	(10f8.2)

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

<PROPERTY MAXIMA and MINIMA>

The ParaSurf™ block for the maxima and minima of the local properties is defined as follows for each property:

Header line (maxima)	Number of maxima for the property: N_{max}, property {MEP, IEL, EAL or Alpha(L)}	(I3,a," Maxima")
Nmax maxima lines	$x, y, z,$ property value	(3f12.4,3x,g10.4)
Header line (minima)	Number of minima for the property: N_{max}, property {MEP, IEL, EAL or Alpha(L)}	(I3,a," Minima")
N_{min} minima lines	$x, y, z,$ property value	(3f12.4,3x,g10.4)

<STANDARD RIF>

The rotationally invariant fingerprint [35] is printed as a list of 54 floating point numbers (5g12.6). The first 41 are those defined in reference [35] and the last 13 are the square roots of the hybridization coefficients for the normal field from *l*=0-12.

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 then contains a table that gives all the autocorrelations as a table with the following headings:

Table 5: Column headings and definitions for autocorrelation tables.

Column heading	Contents
R	Reference distance (R in equation (18)
SHAPE	Shape autocorrelation
VPP	MEP +/+ autocorrelation
VPM	MEP +/- autocorrelation
VMM	MEP -/- autocorrelation
IEL	IEL autocorrelation
EAL	EAL autocorrelation
FPP	Normal field +/+ correlation (raw)
FPM	Normal field +/- correlation (raw)
FMM	Normal field -/- correlation (raw)
FPP_s	Normal field +/+ correlation (scaled)
FPM_s	Normal field +/- correlation (scaled)
FMM_s	Normal field -/- correlation (scaled)

The format of the columns is `(f10.3,4(1x,f10.6),2(1x,f10.3),3(1x,f8.1),3f8.3)`

Calculations with spherical-harmonic fits that use the **`TRANSLATE`** or **`TRANSLATE2`** options, an additional block with the header

`<TRANSLATED SPHERICAL HARMONIC FITS>`

is printed. This block consists of nine sets of results (the original center plus eight translated ones) for **`TRANSLATE`** and 16 for **`TRANSLATE2`**. The original center is denoted by the header

```
Origin    <shiftx> <shifty> <shiftz> <RMSD>
("Origin    :",3f12.4,f12.6)')
```

followed by the fitted coefficients (7f12.6). The shifted points are defined in the same way, but are denoted “**`Point N`**”

```
("Point   ",i2,":",3f12.4,f12.6)
```

3.6 The surface (.psf) file

The .psf file can be used to derive properties and descriptors from the ParaSurf™ results. Note that the format of the .psf file has changed relative to that used in ParaSurf'08™. 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, $atom_{tri}$,normal field	(3i6,f10.5,i6,g12.4)

(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 description 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_FIELD(N)>
<SPHERICAL_HARMONIC_ALPHA(l)>
<SHAPE HYBRIDS>
<MEP HYBRIDS>
<IE(L) HYBRIDS>
<EA(L) HYBRIDS>
<FIELD(N) HYBRIDS>
<ALPHA(L) HYBRIDS>
<STANDARD RIF>

<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, 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^2_{Tot}, ν, $\sigma^2_{tot}\nu$, γ_1^V, γ_2^V, $\int_V$ `("MEP        ",7f10.2/10x, f10.2,5f10.4,2x,g12.6)`
IE(l)	$IE_L^{\max}$, $IE_L^{\min}$, IE_L, ΔIE_L, σ^2_{IE}, γ_1^{IE}, γ_2^{IE}, $\int_{IE}$ `("IE(l)      ",5f10.2,2f10.4/12x,g12.6)`
EA(l)	$EA_L^{\max}$, $EA_L^{\min}$, $\overline{EA}_{L+}$, $\overline{EA}_{L-}$, $\overline{EA}_L$, ΔEA_L, σ^2_{EA+}, σ^2_{EA-}, σ^2_{EA}, ν_{EA}, δA^+_{EA}, A^+_{EA}, γ_1^{EA}, γ_2^{EA}, $\int_{EA}$ `("EA(l)      ",7f10.2/2f10.2,2f10.4,f10.2,2f10.4/12x,g12.6)`
Eneg(l)	$\chi_L^{\max}$, $\chi_L^{\min}$, $\overline{\chi}_L$, $\Delta\chi_L$, σ^2_χ, γ_1^χ, γ_2^χ, $\int_\chi$ `("Eneg(l)    ",5f10.2,2f10.4/12x,g12.6)`
Hard(l)	$\eta_L^{\max}$, $\eta_L^{\min}$, $\overline{\eta}_L$, $\Delta\eta_L$, σ^2_η, γ_1^η, γ_2^η, $\int_\eta$ `("Hard(l)    ",5f10.2,2f10.4/12x,g12.6)`
Alpha(l)	$\alpha_L^{\max}$, $\alpha_L^{\min}$, α_L, $\Delta\alpha_L$, σ^2_α, γ_1^α, γ_2^α, $\int_\alpha$ `("Alpha(l)   ",5f10.2,2f10.4/12x,g12.6)`
F_N	$F_N^{\max}$, $F_N^{\min}$, ΔF_N, $\overline{F}_N$, σ^2_F, σ^2_{F+}, σ^2_{F-}, ν_F, $\gamma_1^{F_N}$, $\gamma_2^{F_N}$, $\int_{F_N}$, $\int^+_{F_N}$, $\int^-_{F_N}$ `("Field desc",7f10.4/"          ",6f10.4)`

Jobs that include Shannon entropy give two extra sets of descriptors:

Shannon(i)	$H_{in}^{\max}$, $H_{in}^{\min}$, $\overline{H}_{in}$, $\sigma^2_{H_{in}}$, $\int_{H_{in}}$ `("Shannon(i)  ",4f10.4,f10.2,f10.4)`
Shannon(e)	$H_{ex}^{\max}$, $H_{ex}^{\min}$, $\overline{H}_{ex}$, $\sigma^2_{H_{ex}}$, $\int_{H_{ex}}$ `("Shannon(e)  ",4f10.4,f10.2,f10.4)`

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 then contains a table that gives all the autocorrelations as a table with the following headings:

Table 6: Column headings and definitions for the autocorrelation table in the output SDF file.

Column heading	Contents
R	Reference distance (R in equation (18)
SHAPE	Shape autocorrelation
VPP	MEP +/+ autocorrelation
VPM	MEP +/- autocorrelation
VMM	MEP -/- autocorrelation
IEL	IE_L autocorrelation
EAL	EA_L autocorrelation
FPP	Normal field +/+ correlation (raw)
FPM	Normal field +/- correlation (raw)
FMM	Normal field -/- correlation (raw)
FPP_s	Normal field +/+ correlation (scaled)
FPM_s	Normal field +/- correlation (scaled)
FMM_s	Normal field -/- correlation (scaled)

The format of the columns is `(f10.3,4(1x,f10.6),2(1x,f10.3),3(1x,f8.1),3f8.3)`

3.8 Grid calculations with ParaSurf™

3.8.1 User-specified Grid

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, IEL, EAL, αL). 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 16: Sample grid file

Gives the output shown in Figure 17:

```
<> ParaSurf´09 : 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 17: 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, IEL, and EAL in kcal mol^{-1}, αL in Ångstrom^3.

3.8.2 Automatic grids

ParaSurf™ can generate grids automatically for lead compounds in ComFA®-like procedures. The **`grid=auto`** option generates a grid around the molecule (with a 4 Å margin around the positions of the atoms in each direction) and includes all points for which the electron density is lower than 10^{-2} (i.e. for points outside the molecule). The spacing of the grid is set to a default value of 1.0 Å, but can be set to any value up to a maximum of 2.0 Å by the command-line argument **`lattice=n.n`**, which sets the lattice spacing to *n.n* Å. The grid thus generated is output (with the values of the local properties analogously to a calculation that uses an predefined grid and can be used for other molecules that have been aligned with the lead.

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 18:

```
> <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 18: 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 7: Definitions and order of the descriptors printed to the descriptor table if requested.

Column Header	Symbol[a]	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

Column Header	Symbol[a]	Descriptor
meanMEP-	$\bar{V}_-$	Mean of the negative MEP values
meanMEP	$\bar{V}$	Mean of all MEP values
MEPrange	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
MEPskew	γ_1^V	Skewness of the distribution of the MEP
MEPkurt	γ_2^V	Kurtosis of the distribution of the MEP
MEPint	$\int_V$	Integral of the MEP*area over the surface
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
IELrange	ΔIE_L	Range of the local ionization energy
IELvar	σ_{IE}^2	Variance in the local ionization energy
IELskew	γ_1^{IE}	Skewness of the distribution of IE(L)
IELkurt	γ_2^{IE}	Kurtosis of the distribution of IE(L)
IELint	$\int_{IE}$	Integral of the IE(L)*area over the surface
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
EALrange	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+	δA_{EA}^+	Fraction of the surface area with positive local electron affinity
EALarea+	A_{EA}^+	Surface area with positive local electron affinity
EALskew	γ_1^{EA}	Skewness of the distribution of the MEP

Column Header	Symbol[a]	Descriptor
EALkurt	γ_2^{EA}	Kurtosis of the distribution of the MEP
EALint	$\int_{EA}$	Integral of the MEP*area over the surface
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
POLrange	$\Delta\alpha_L$	Range of the local polarizability
POLvar	σ_α^2	Variance in the local polarizability
POLskew	γ_1^α	Skewness of the distribution of the local polarizability
POLkurt	γ_2^α	Kurtosis of the distribution of the local polarizability
POLint	$\int_\alpha$	Integral of the α(L)*area over the surface
ENEGmax	$\chi_L^{\max}$	Maximum of the local electronegativity
ENEGmin	$\chi_L^{\min}$	Minimum of the local electronegativity
ENEGbar	$\overline{\chi_L}$	Mean value of the local electronegativity
ENEGrange	$\Delta\chi_L$	Range of the local electronegativity
ENEGvar	σ_χ^2	Variance in the local electronegativity
ENEGskew	γ_1^χ	Skewness of the distribution of the local electronegativity
ENEGkurt	γ_2^χ	Kurtosis of the distribution of the local electronegativity
ENEGint	$\int_\chi$	Integral of the χ(L)*area over the surface
HARDmax	$\eta_L^{\max}$	Maximum of the local electronegativity
HARDmin	$\eta_L^{\min}$	Minimum of the local electronegativity
HARDbar	$\overline{\eta_L}$	Mean value of the local electronegativity
HARDrange	$\Delta\eta_L$	Range of the local electronegativity
HARDvar	σ_η^2	Variance in the local electronegativity
HARDskew	γ_1^η	Skewness of the distribution of the local electronegativity
HARDkurt	γ_2^η	Kurtosis of the distribution of the local electronegativity
HARDint	$\int_\eta$	Integral of the χ(L)*area over the surface
FNmax	$F_N^{\max}$	Maximum value of the field normal to the surface
FNmin	$F_N^{\min}$	Minimum value of the field normal to the surface
FNrange	ΔF_N	Range of the field normal to the surface
FNmean	$\overline{F_N}$	Mean value of the field normal to the surface
FNvartot	σ_F^2	Variance in field normal to the surface
FNvar+	σ_{F+}^2	Variance in the field normal to the surface for all positive values

Column Header	Symbol[a]	Descriptor
FNvar-	σ^2_{F-}	Variance in the field normal to the surface for all negative values
FNbal	ν_F	Normal field balance parameter
FNskew	$\gamma_1^{F_N}$	Skewness of the field normal to the surface
FNkurt	$\gamma_2^{F_N}$	Kurtosis of the field normal to the surface
FNint	$\int_{F_N}$	Integrated field normal to the surface over the surface
FN+	$\int^+_{F_N}$	Integrated field normal to the surface over the surface for all positive values
FN-	$\int^-_{F_N}$	Integrated field normal to the surface over the surface for all negative values
FNabs	$\int_{\lvert F_N \rvert}$	Integrated absolute field normal to the surface over the surface

[a]Symbols as used in section 1.9.

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	$\sigma^2_{H(in)}$	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	$\sigma^2_{H(ex)}$	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,IE,IEQ1,IEQ2,IEQ3, IEQ4, EAi, EAQ1, EAQ2, EAQ3, EAQ4,Fpp, FppQ1,FppQ2, FppQ3, FppQ4, Fmm, FmmQ1, FmmQ2, FmmQ3, FmmQ4, Fpm, FpmQ1, FpmQ2, FpmQ3, FpmQ4, FsPP, FsppQ1, FsppQ2, FsppQ3, FsppQ4, Fsmm, FsmmQ1, FsmmQ2, FsmmQ3, FsmmQ4, Fspm, FspmQ1, FspmQ2, FspmQ3, FspmQ4, Hshape, HshapeQ1, HshapeQ2, HshapeQ3, HshapeQ4, HVpp, HVppQ1, HVppQ2, HVppQ3, HVppQ4,HVmm, HVmmQ1, HVmmQ2, HvMMQ3, HVmmQ4, HVpm, HVpmQ1, HVpmQ2, HVpmQ3, HVpmQ4, HIE, HIEQ1, HIEQ2, HIEQ3, HIEQ4,HEAi, HEAiQ1, HEAiQ2, HEAiQ3, HEAiQ4, HFpp, HFppQ1, HFppQ2, HFppQ3, HFppQ4, HFmm, HFmmQ1, HFmmQ2, HFmmQ3, HFmmQ4, HFpm, HFpmQ1, HFpmQ2, HFpmQ3, HFpmQ4, HFSpp, HFSppQ1, HFSppQ2, HFSppQ3, HFSppQ4, HFSmm, HFSmmQ1, HFSmmQ2, HFSmmQ3, HFSmmQ4, HFSpm, HFSpmQ1, HFSpmQ2, HFSpmQ3, HFSpmQ4, Rr, Rpp, Rmm, Ri, Re, RFpp, RFmm, RFpm, RFSpp, RFSmm, RFSpm

The MolID column contains the name of the molecule as given in its SDF-file and similarities for each type of autocorrelation as follows:

Table 8: Definitions of the elements of the Autocorrelation similarity block in the output SDF file.

Autocorrelation	Similarity					Entropy					Correl. Coeff.
	Total	Quartal				Total	Quartal				
		1	2	3	4		1	2	3	4	
Shape	shape	shapeQ1	shapeQ2	shapeQ3	shapeQ4	Hshape	HshapeQ1	HshapeQ2	HshapeQ3	HshapeQ4	Rr
MEP +/+	Vpp	VppQ1	VppQ2	VppQ3	VppQ4	HVpp	HVppQ1	HVppQ2	HVppQ3	HVppQ4	Rpp
MEP -/-	Vpm	VpmQ1	VpmQ2	VpmQ3	VpmQ4	HVpm	HVpmQ1	HVpmQ2	HVpmQ3	HVpmQ4	Rpm
MEP +/-	Vmm	VmmQ1	VmmQ2	VmmQ3	VmmQ4	HVmm	HVmmQ1	HVmmQ2	HVmmQ3	HVmmQ4	Rmm
IE_L	IE	IEQ1	IEQ2	IEQ3	IEQ4	HIE	HIEQ1	HIEQ2	HIEQ3	HIEQ4	Ri
EA_L	EA	EAQ1	EAQ2	EAQ3	EAQ4	HEA	HEAQ1	HEAQ2	HEAQ3	HEAQ4	Re
F_N +/+	Fpp	FppQ1	FppQ2	FppQ3	FppQ4	HFpp	HFppQ1	HFppQ2	HFppQ3	HFppQ4	RFpp
F_N -/-	Fpm	FpmQ1	FpmQ2	FpmQ3	FpmQ4	HFpm	HFpmQ1	HFpmQ2	HFpmQ3	HFpmQ4	RFpm
F_N +/-	Fmm	FmmQ1	FmmQ2	FmmQ3	FmmQ4	HFmm	HFmmQ1	HFmmQ2	HFmmQ3	HFmmQ4	RFmm
F_N +/+ (scaled)	FSpp	FSppQ1	FSppQ2	FSppQ3	FSppQ4	HFSpp	HFSppQ1	HFSppQ2	HFSppQ3	HFSppQ4	RFSpp
F_N -/- (scaled)	FSpm	FSpmQ1	FSpmQ2	FSpmQ3	FSpmQ4	HFSpm	HFSpmQ1	HFSpmQ2	HFSpmQ3	HFSpmQ4	RFSpm
F_N +/- (scaled)	FSmm	FSmmQ1	FSmmQ2	FSmmQ3	FSmmQ4	HFSmm	HFSmmQ1	HFSmmQ2	HFSmmQ3	HFSmmQ4	RFSmm

The last five entries are the correlation coefficients for the eleven autocorrelations.

3.12 Shared files

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

4 TIPS FOR USING PARASURF´09™

4.1 Choice of surface

ParaSurf™ was originally written to use isodensity surfaces. However, calculations that use either a solvent-excluded or solvent-accessible surface are very much faster than their equivalents with isodensity surfaces and will usually give comparable results. Surface-integral models may benefit from using a solvent-accessible surface with a solvent radius of 0.5-1.0 Å as this appears to be the most relevant surface for many physical properties. Surfaces fitted to spherical-harmonic expansions require more CPU-time than marching-cube surfaces but are essential for fast numerical applications such as ParaFit™. Again, solvent-excluded shrink-wrap surfaces are faster to calculate than their isodensity equivalents.

4.2 ParaSurf™ and ParaFit™

ParaFit™ is Cepos InSilico's very fast shape-matching program that is based on spherical-harmonic expansions generated by ParaSurf™. ParaFit™ can be used to overlay molecules with a common scaffold by defining the center to be used for generating the spherical-harmonic fit in ParaSurf™ in the input SDF-file (see **1.1.4** and **2.2**)

4.3 QSAR using grids

As outlined in **3.8.2**, ParaSurf™ can generate a grid for the lead molecule automatically that can then be used for a set of aligned (e.g. with ParaFit™) molecules for grid-based QSAR. This procedure has proven to be especially effective for test datasets, especially if the molecules are aligned to a common scaffold, as outlined in **4.2**.

5 SUPPORT

5.1 Contact

Questions regarding ParaSurf™ should be sent directly to:

support@ceposinsilico.com

5.2 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.

5.3 Cepos Insilico Ltd.

Computer-Chemie-Centrum (CCC)
Nägelsbachstr. 25
91052 Erlangen
Germany

support@ceposinsilico.com
Tel. +49-9131-9704910
Fax. +49-9131-9704911

www.ceposinsilico.com/contact

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/products/datasheets/vamp.pdf**).

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 J. H. Van Drie, *"Shrink-wrap" surfaces: A new method for incorporating shape into pharmacophoric 3D database searching*, *J. Chem. Inf. and Comp. Sci.*, **1997**, *37*, 38-41; J. H. Van Drie und R. A. Nugent, *Addressing the challenges of combinatorial chemistry: 3D databases, pharmacophore recognition and beyond*, *SAR and QSAR in Env. Res.*, **1998**, *9*, 1-21; J. Erickson, D. J. Neidhart, J. Van Drie, D. J. Kempf, X. C. Wang, D. W. Norbeck, J. J. Plattner, J. W. Rittenhouse, M. Turon, N. Wideburg, et al., *Design, activity, and 2.8 A crystal structure of a C2 symmetric inhibitor complexed to HIV-1 protease*, *Science*, **1990**, *249*, 527-533.

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 *CypScore: Quantitative Prediction of Reactivity toward Cytochromes P450 Based on Semiempirical Molecular Orbital Theory*, M. Hennemann, A. Friedl, M. Lobell, J. Keldenich, A. Hillisch, T. Clark and A. H. Göller, *ChemMedChem*, **2009**, *4*, 657-669.

12 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.

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

14 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.

15 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.

16 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.

17 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.

18 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.

19 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.

20 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.

21 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.

22 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.

23 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.

24 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.

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

26 R. J. Abraham, B. D. Hudson, M. W. Kermode and J. R. Mines, *J. Chem. Soc., Faraday Trans. I*, **1988**, *84*, 1911-1917.

27 Tomasi, B. Mennucci and R. Cammi, *Quantum Mechanical Continuum Solvation Models, Chem. Rev.,* **2005,** 105, 2999.

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

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

30 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.

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

32 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/**

33 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.

34 **http://www.mdl.com/solutions/white_papers/ctfile_formats.jsp**

35 L. Mavridis, B. D. Hudson and D. W. Ritchie, *Toward high throughput virtual screening using spherical harmonic surface representations*, *J. Chem. Inf. Model.,* **2007**, *47*, 1787-1796.

36 A. Jakobi, H. Mauser and T. Clark, *ParaFrag—an approach for surface-based similarity comparison of molecular fragments, J. Mol. Model.,* **2008**, *14*, 547-558.

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