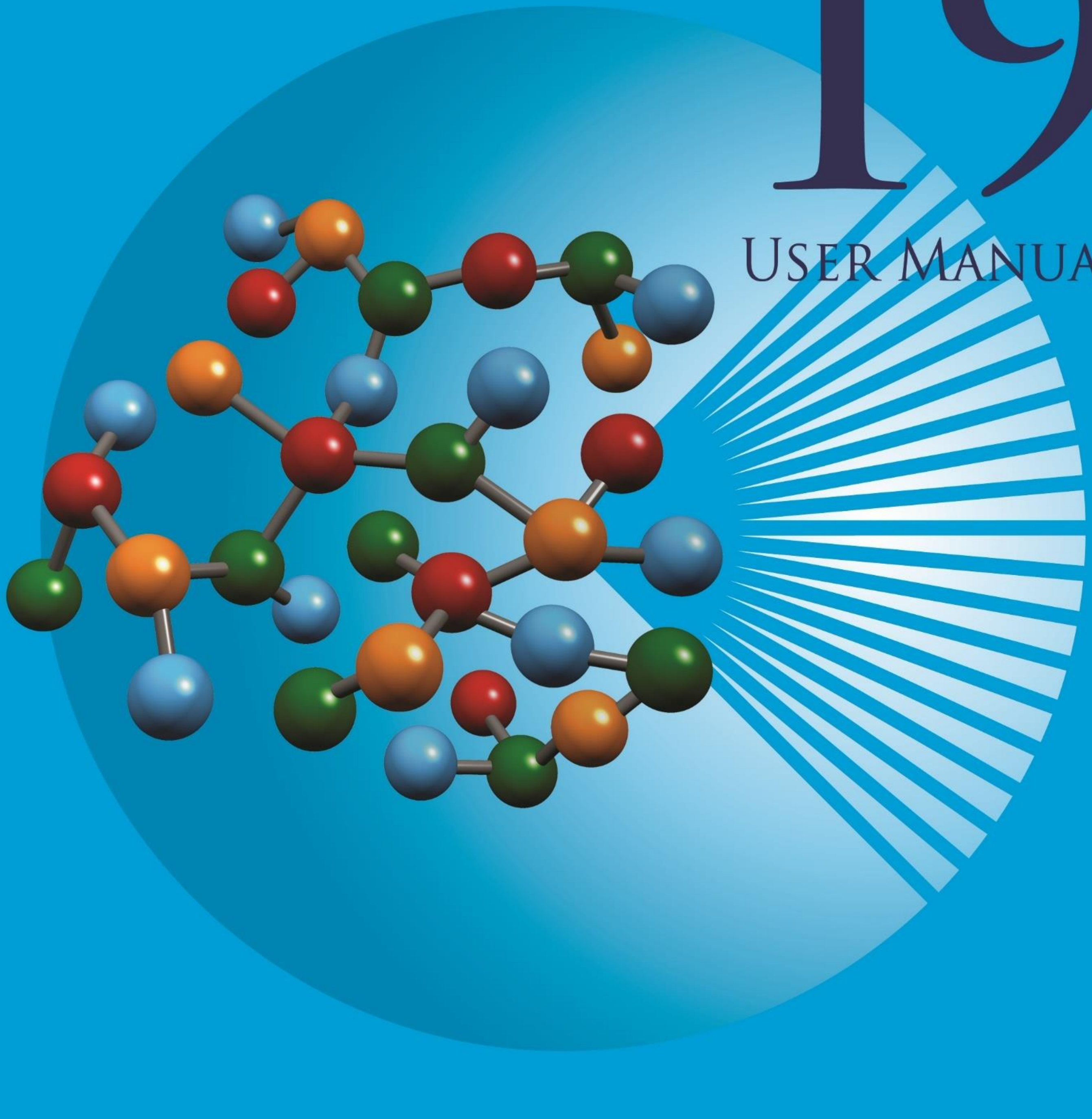
CEPOS INSILICO
SCIENCE FOR SCIENTISTS



19

USER MANUAL

PARASURF™

Impressum

Waldstraße 15
90587 Obermichelbach
Germany
www.ceposinsilico.com

Manual Timothy Clark

Layout www.eh-bitartist.de

TABLE OF CONTENTS

PROGRAM HISTORY

Release Date	Version	Platforms
1st July 2005	ParaSurf´05™ initial release (Revision A1)	32-bit Windows 32-bit Linux Irix
1st January 2006	ParaSurf´05™ Revision B1 (customer-feedback release)	
1st July 2006	ParaSurf´06™ Revision A1	32-bit Windows 32-bit Linux 64-bit Linux Irix
1st July 2007	ParaSurf´07™ Revision A1	
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	
1st September 2009	New Vhamil.par file including PM6 and first-row transition metals in AM1*	
1st February 2010	ParaSurf´09™ Revision B1 (additional atom-centred descriptors)	
1st July 2010	ParaSurf´10™ Revision A1	
1st July 2011	ParaSurf´11™ Revision A1	
1st September 2013	ParaSurf´12™ Revision A1	
1st November 2019	ParaSurf19™ Revision A1	64-bit Windows 64-bit Linux
1st March 2020	ParaSurf19™ Revision A2 (Vhamil.par replaced by EMPIRE <Hamiltonian>.par file)	64-bit Windows 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.

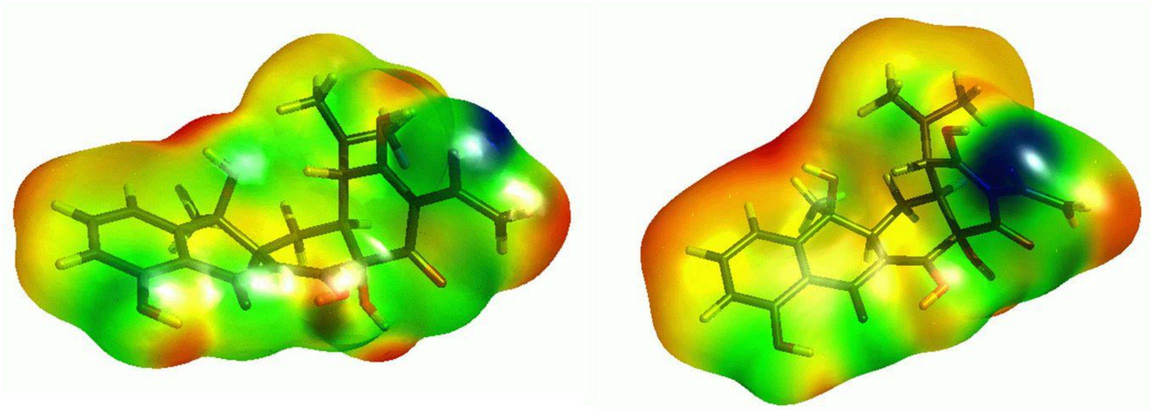


Figure 1 Marching-cube (left) and shrink-wrap (right, fitted to a spherical-harmonic approximation) isodensity surfaces calculated with ParaSurf™ using the default settings

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 81 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´12™

The functionality of ParaSurf19™ has been extended to allow surface-based substructure searches. In detail, the changes relative to ParaSurf'12™ are:

- ParaSurf19 now outputs separate .psf and .sdf files for fragments. These allow fragments to be matched to complete molecules (or other fragments) using CImatch™
- ParaSurf19 includes performance enhancements and parallelization of the surface-determination and the autocorrelation calculation
- Revision A2 now uses the EMPIRE <Hamiltonian>.par file, rather then the Vhamil.par file, with SDF inputs

1.1.1 EMPIRE™

ParaSurf19 is compatible with CeposInSilico's EMPIRE19™ program for performing semiempirical molecular orbital calculations and communicates with EMPIRE using the .sdf file format.

1.2 Isodensity surfaces

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

1.3 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.

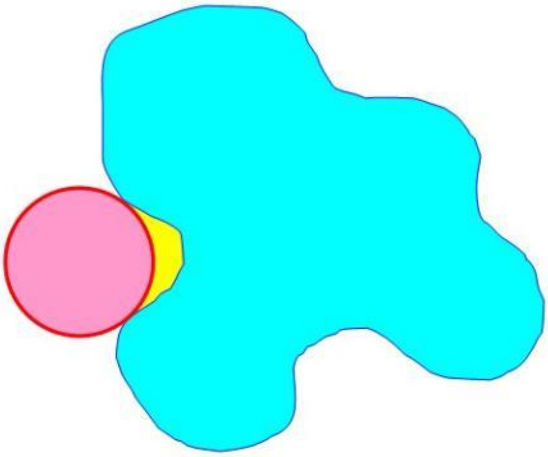


Figure 2 2D-representation of a solvent-excluded surface.

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**.

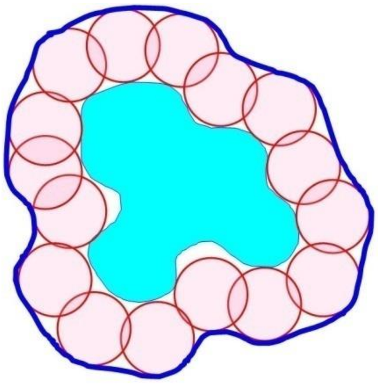


Figure 3 The solvent-accessible surface is obtained by rolling a spherical "solvent molecule".

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 centre 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**):

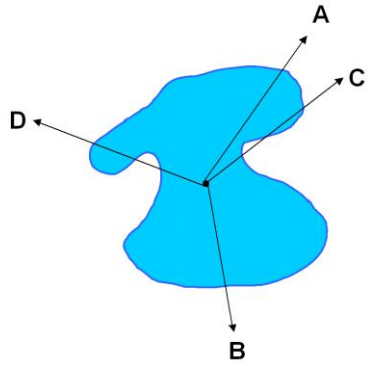


Figure 4 2D-representation of a molecular surface with single-valued (A and B) and multiply valued (C and D) radial vectors from the centre.

Single-valued surfaces are necessary for spherical-harmonic fitting (see Section **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 centre in ParaSurf™ is the centre of gravity (CoG). In special cases in which the CoG lies outside the molecule, another centre may be chosen.

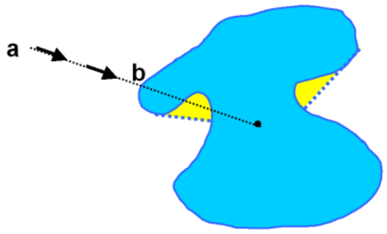


Figure 5 2D-representation of the shrink-wrap algorithm. The algorithms scans along the vector from point a towards the centre of the molecule until the electron density reaches the preset value (point b). The algorithm results in enclosures (marked yellow) for multi-valued radial vectors.

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

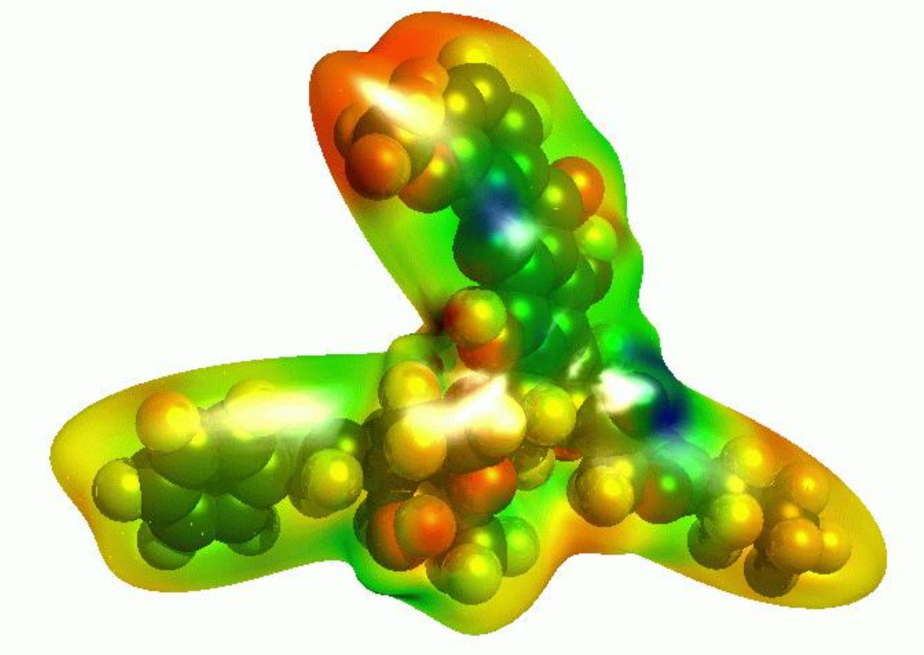


Figure 6 Spherical-harmonic approximation of a shrink-wrap isodensitiy surface. Note the areas where the surface does not follow the indentations of the molecule.

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.

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 centre along the radial vector defined by the angles α and β as:

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

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

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

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

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

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

ParaSurf™ not only fits the surface itself (i.e. the radial distances) to spherical harmonic expansions, but also the four local properties (see Section **1.8**). In this way, a completely analytical description of the shape of the molecule and its intermolecular binding properties is obtained.[13] This description can be truncated at different orders l depending on the application and the precision needed. Thus, a simple description of the molecular properties (shape, MEP, IE_L, EA_L and α_L) to order 2 consists of only five sets of nine coefficients each, or 45 coefficients. These coefficients can be rotated, overlaps calculated etc. [5] to give fast scanning of large numbers of compounds.

Note that, because of the approximate nature of the spherical-harmonic fits, the default isodensity level for the shrink-wrapped surface (0.0005 $e^-Å^{-3}$) is lower than that (0.007 $e^-Å^{-3}$) 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. ParaSurf19™ 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 centre for fitting to a spherical-harmonic expansion is critical. ParaSurf19™ therefore goes through a multi-step procedure in order to find a suitable centre. This procedure is retained for all molecules for which the ParaSurf'08™ found a suitable centre. However, if the algorithms implemented in ParaSurf'08™ fail to find a suitable centre, the additional technique first implemented in ParaSurf'12™ will probably work.

The problem with many molecules is that, for instance, the centre 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 ParaSurf19™ works as follows:

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

2. ParaSurf™ calculates the principal moments of inertia of the molecule and derives a centre from them by assuming that the molecule is U- or V-shaped. The procedure tries to place the centre at the base centre of the molecule. This procedure was implemented in ParaSurf'08™ as a fallback if the centre of mass proved unsuitable. If it also fails to find a suitable centre, ParaSurf19™ moves on to a third option, which finds a centre 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 centre is then placed in the hinge area between the two planes, as illustrated in **Figure 7**:

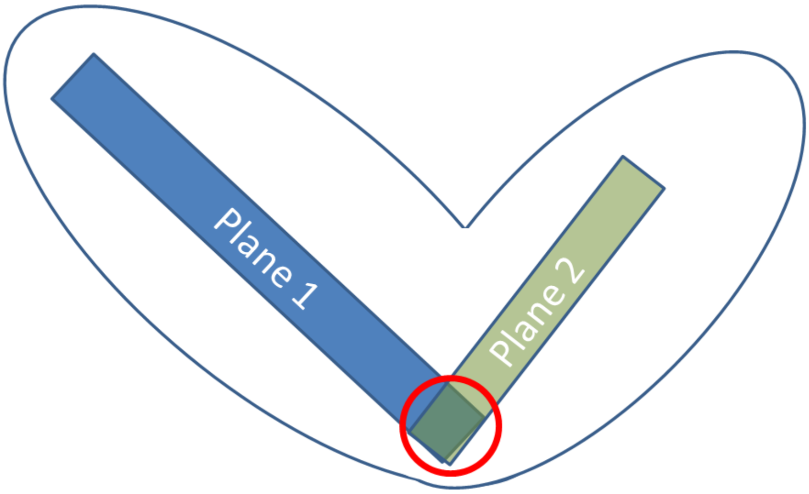


Figure 7 Schematic representation of the planes and hinge area used to determine the centre for spherical-harmonic expansions.

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 [**14**] 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,[**15**] AM1 [**16**] and PM3.[**17**] The negative charges are located at the charge centres of each lobe of the natural atomic orbitals, which are obtained by diagonalizing the one-atom blocks of the density matrix.[**18**] The NAO-PC charges are calculated by VAMP and output in the .sdf file for use in ParaSurf™. The NAO-PC model is therefore only available when using ParaSurf™ with VAMP .sdf input. NAO-PC charges are also not available for semiempirical Hamiltonians such as MNDO/d[**19**] or AM1*[**18**] 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 [**20**] that extends to quadrupoles. We can therefore also use this approximation to calculate atom-centred monopoles, dipoles and quadrupoles for each atom in the molecule.[**21**] This multipole model is applicable to all methods, including those with d-orbitals, and can be used with MOPAC output files as input to ParaSurf™.

1.8.2 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] An intensity-filtering technique [20b] was introduced in ParaSurf'10™ to allow the local electron affinity to be calculated for Hamiltonians such as AM1* and MNDO/d that use polarisation d-functions.

Equation (6) requires that the occupied and virtual orbitals be approximately equivalent to each other. This is not the case for semiempirical Hamiltonians (such as AM1*) that include d-orbitals as polarisation functions or for extensive basis sets in Hartree-Fock ab initio or in Density-Functional theory (DFT) calculations. A new technique has therefore been defined [11] to exclude pure polarisation functions from the summation in **Equation (6)**. This technique is now default in ParaSurf19™ and gives reliable results. For continuity, a new command-line option (EAL09) has been introduced to request that the calculation of the local electron affinity be performed exactly as in ParaSurf'09™ and earlier versions.

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

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

and the local electronegativity, χ_L:

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

1.8.3 Local polarizability

Within the NDDO, the molecular electronic polarizability is easily accessible using the parameterized version [**22**] of the variational technique introduced by Rivail,[**23**] which can also be partitioned into an additive polarizability scheme.[**20a**] Versions of ParaSurf™ up to ParaSurf'11 used an isotropic definition of the the local polarizability, α_L, at a point near the molecule:

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

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.

ParaSurf19 makes use of the fact that the atomic polarizability tensors produced by the procedure described in reference [**20a**] are anisotropic. It uses this atomic anisotropy to calculate a more highly resolved local polarizability that is now standard in ParaSurf19. The keyword "parasurf11" ensures backwards compatibility with the isotropic local polarizability used in earlier versions.

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.[**24**] 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 81 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	[20a]	`dipden`
α	Molecular electronic polarisabilty	[25]	`polarizability`
MW	Molecular weight		`MWt`
G	Globularity	[26]	`globularity`
A	Molecular surface area		`totalarea`
VOL	Molecular volume		`volume`
V_{max}	Maximum (most positive) MEP	[27]	`MEPmax`
V_{min}	Minimum (most negative) MEP	[27]	`MEPmin`
$\overline{V}_+$	Mean of the positive MEP values	[27]	`meanMEP+`
$\overline{V}_-$	Mean of the negative MEP values	[27]	`meanMEP-`
$\overline{V}$	Mean of all MEP values	[27]	`meanMEP`
ΔV	MEP-range	[27]	`MEP-range`
σ_+^2	Total variance in the positive MEP values	[27]	`MEPvar+`
σ_-^2	Total variance in the negative MEP values	[27]	`MEPvar-`
σ_{tot}^2	Total variance in the MEP	[27]	`MEPvartot`
ν	MEP balance parameter	[27]	`MEPbalance`
$\sigma_{tot}^2\nu$	Product of the total variance in the MEP and the balance parameter	[27]	`var*balance`
γ_1^V	Skewness of the MEP-distribution	$\gamma_1^{\alpha_L} = \frac{\sum_{i=1}^{N}\left(\alpha_L^i - \bar{\alpha}_L\right)^3}{(N-1)\sigma^3}$	`MEPskew`

Descriptor	Description	Formula/ Reference	Symbol in CSV file
γ_2^V	Kurtosis of the MEP-distribution	$\gamma_2^V = \frac{\sum_{i=1}^{N}\left(V_i - \bar{V}\right)^4}{(N-1)\sigma^4} - 3$	`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`
$\gamma_1^{IE_L}$	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`

Descriptor	Description	Formula/ Reference	Symbol in CSV file
σ^2_{EA+}	Variance in the local electron affinity for all positive values	$\sigma^2_{EA+} = \frac{1}{m}\sum_{i=1}^{m}\left[EA_i^+ - \overline{EA^+}\right]^2$	**`EALvar+`**
σ^2_{EA-}	Variance in the local electron affinity for all negative values	$\sigma^2_{EA-} = \frac{1}{n}\sum_{i=1}^{n}\left[EA_i^- - \overline{EA^-}\right]^2$	**`EALvar-`**
σ^2_{EAtot}	Sum of the positive and negative variances in the local electron affinity	$\sigma^2_{EAtot} = \sigma^2_{EA+} + \sigma^2_{EA-}$	**`EALvartot`**
ν_{EA}	Local electron affinity balance parameter	$\nu_{EA} = \frac{\sigma^2_{EA+} \cdot \sigma^2_{EA-}}{\left[\sigma^2_{EA}\right]^2}$	**`EALbalance`**
$\delta \mathrm{A}^+_{EA}$	Fraction of the surface area with positive local electron affinity	$\delta \mathrm{A}^+_{EA} = \frac{\mathrm{A}^+_{EA}}{\mathrm{A}}$, A = total surface area	**`EALfraction+`**
A^+_{EA}	Surface area with positive local electron affinity		**`EALarea+`**
$\gamma_1^{EA_L}$	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^2_α	Variance in the local polarizability	$\sigma^2_\alpha = \frac{1}{N}\sum_{i=1}^{N}\left[\alpha_L^i - \overline{\alpha_L}\right]^2$	**`POLvar`**
$\gamma_1^{\alpha_L}$	Skewness of the local polarizability distribution	$\gamma_1^{\alpha_L} = \frac{\sum_{i=1}^{N}\left(\alpha_L^i - \bar{\alpha}_L\right)^3}{(N-1)\sigma^3}$	**`POLskew`**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\gamma_2^{\alpha_L}$	Kurtosis of the local polarizability distribution	$\gamma_2^{\alpha_L} = \frac{\sum_{i=1}^{N}\left(\alpha_L^i - \bar{\alpha}_L\right)^4}{(N-1)\sigma^4} - 3$	**`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`**
$\gamma_1^{\chi_L}$	Skewness of the local electronegativity distribution	$\gamma_1^{\chi_L} = \frac{\sum_{i=1}^{N}\left(\chi_L^i - \bar{\chi}_L\right)^3}{(N-1)\sigma^3}$	**`ENEGskew`**
$\gamma_2^{\chi_L}$	Kurtosis of the local electronegativity distribution	$\gamma_2^{\chi_L} = \frac{\sum_{i=1}^{N}\left(\chi_L^i - \bar{\chi}_L\right)^4}{(N-1)\sigma^4} - 3$	**`ENEGkurt`**
$\int_{\chi_L}$	Integrated local electronegativity over the surface	$\int_{\chi_L} = \sum_{i=1}^{N} \chi_L^i a_i$	**`ENEGint`**
$\eta_L^{\max}$	Maximum value of the local hardness		**`HARDmax`**
$\eta_L^{\min}$	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	$\gamma_1^{\eta_L} = \frac{\sum_{i=1}^{N}\left(\eta_L^i - \bar{\eta}_L\right)^3}{(N-1)\sigma^3}$	**`HARDskew`**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\gamma_2^{\eta_L}$	Kurtosis of the local hardness distribution	$\gamma_2^{\eta_L} = \frac{\sum_{i=1}^{N}\left(\eta_L^i - \bar{\eta}_L\right)^4}{(N-1)\sigma^4} - 3$	**HARDkurt**
$\int_{\eta_L}$	Integrated local hardness over the surface	$\int_{\eta_L} = \sum_{i=1}^{N} \eta_L^i a_i$	**HARDint**
$F_N^{\max}$	Maximum value of the electrostatic field normal to the surface		**FNmax**
$F_N^{\min}$	Minimum value of the field normal to the surface		**FNmin**
$\overline{F_N}$	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	$\gamma_1^{F_N} = \frac{\sum_{i=1}^{N}\left(F_N^i - \bar{F}_N\right)^3}{(N-1)\sigma^3}$	**FNskew**
$\gamma_2^{F_N}$	Kurtosis of the field normal to the surface	$\gamma_2^{F_N} = \frac{\sum_{i=1}^{N}\left(F_N^i - \bar{F}_N\right)^4}{(N-1)\sigma^4} - 3$	**FNkurt**
$\int_{F_N}$	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-**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\int_{\lvert F_N \rvert}$	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`**
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`**
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`**

1.10 Surface-integral models (polynomial version)

The polynomial 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 polynomial 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 Section **3.10**. 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 Binned surface-integral models

A more recent type of SIM model, binned SIM models, [13] is now implemented in ParaSurf19™ for the negative logarithm of the water-octanol partition constant, logP$_{OW}$. These models divide the surface into bins according to the values of the local properties and use the total surface area assigned to each bin as descriptors for multiple linear regression models. These models have been implemented for marching cube surfaces using either the isodensity or solvent-excluded surfaces and for the AM1, AM1*, MNDO, MNDO/d, PM3 and PM6 Hamiltonians. In contrast to polynomial SIM models, they are encoded in the program and are output under the heading “ParaSurf™ ADMET Profiler”. These logP$_{OW}$ models are available for the MNDO, AM1, PM3, MNDO/d, AM1* and PM6 Hamiltonians. The models use “conformationally averaged” structures within a standard calculational protocol in which the initial 3D structure is produced by CORINA [28] as the starting geometry for the semiempirical geometry optimization and uses only this one conformation to predict logP$_{OW}$ for the compound. These models were trained with all verified values contained in the LogKOW dataset [29] and are those denoted “single conformation” trained with the “full” dataset in the original literature.

Table 2 The 28 local properties and products thereof used to construct binned area descriptors.

	IE_L	EA_L	α_L [a]	F_N	η_L	χ_L
MEP	$MEP \times IE_L$	$MEP \times EA_L$	$MEP \times \alpha_L$	$MEP \times F_N$	$MEP \times \eta_L$	$MEP \times \chi_L$
IE_L		$IE_L \times EA_L$	$IE_L \times \alpha_L$	$IE_L \times F_N$	$IE_L \times \eta_L$	$IE_L \times \chi_L$
EA_L			$EA_L \times \alpha_L$	$EA_L \times F_N$	$EA_L \times \eta_L$	$EA_L \times \chi_L$
α_L [a]				$\alpha_L \times F_N$	$\alpha_L \times \eta_L$	$\alpha_L \times \chi_L$
F_N					$F_N \times \eta_L$	$F_N \times \chi_L$
η_L						$\eta_L \times \chi_L$

[a] Not used for MNDO/d, AM1* or PM6

Local hydrophobicities and logP$_{OW}$ models are available for the following combinations of Hamiltonians, surfaces and contours. The three letter model code is used to write the local hydrophobicity to the output .vmp and .psf files or to specify that the descriptors for the model are written out.

Table 3 Local hydrophobicity models and their model codes (all models use the single CORINA-derived conformations and are trained with the “full” dataset

Hamiltonian	Model code	
	ParaSurf’11™	ParaSurf’19™
AM1	LP1	OW1
AM1*	LP2	OW2
PM3	LP3	OW3
MNDO	LP4	OW4
MNDO/d	LP5	OW5
PM6	LP6	OW6

1.12 Spherical harmonic “hybrids”

Once the molecular shape or a local property have been fitted to a spherical-harmonic expansion, [**16**] 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.

1.13 Descriptors and moments based on polynomial surface-integral models

ParaSurf™ uses local properties defined in a surface-integral model (SIM, see Section **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 $\mathbf{r_i}$.

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 modelled by the SIM. Note that no units are given in the output because they depend on the property modelled by the SIM.

1.14 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.[**30**]

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 (13)** becomes

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

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

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

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

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

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

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

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

ParaSurf™ 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 [**31**] in their correct protonation states and at geometries obtained by optimizing with AM1 starting from the bound conformation.[**27**]

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.

1.15 Surface autocorrelations

Gasteiger et al. [27] 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 now defined as:

$$A(R) = \sum_{i=1}^{\text{npoints-1}} \sum_{\text{j=i+1}}^{\text{npoints}} \omega_{\text{ij}} \delta\left(R, r_{ij}\right) \tag{18}$$

where r_{ij} is the distance between surface points i and j and ω_{ij} is a function of one or more local properties at the points i and j and δ_{ij} is 1.0 if r_{ij} is inside the bin centred on R and zero otherwise. Note that this is a different definition of the autocorrelation function to that used in earlier versions of ParaSurf™. Also, because the new algorithm for calculating the autocorrelations is very fast, all surface points are used, rather than sampling 10% as earlier.

Seven different autocorrelation functions are calculated by ParaSurf™. These are:

Shape autocorrelation	$\omega_{ij} = 1.0$	
MEP autocorrelation	$\omega_{ij} = V_i \times V_j$	
Plus-plus MEP autocorrelation	$\omega_{ij} = V_i \times V_j$ $\omega_{ij} = 0.0$	$(V_i > 0$ and $V_j > 0)$ $(V_i < 0$ or $V_j < 0)$
Minus-minus MEP autocorrelation	$\omega_{ij} = V_i \times V_j$	$(V_i < 0$ and $V_j < 0)$
Plus-minus MEP autocorrelation	$\omega_{ij} = -V_i \times V_j$ $\omega_{ij} = 0.0$	$(V_i \times V_j < 0)$ $(V_i \times V_j > 0)$
Local ionization energy autocorrelation	$\omega_{ij} = IE_L^i \times IE_L^j$	
Local electron affinity autocorrelation	$\omega_{ij} = EA_L^i \times EA_L^j$	
Local polarizability autocorrelation	$\omega_{ij} = \alpha_L^i \times \alpha_L^j$	

Generally, the shape autocorrelation and that based on the local polarizability correlate strongly with each other. The MEP correlation is the sum of its three components (plus-plus, plus-minus and minus-minus). However, the three components enable us to distinguish between ++ and – pairs of surface points, which both give a positive contribution to the autocorrelation function.

ParaSurf™ calculates autocorrelations as vectors of A(R) values 64 elements long starting at an R-value of 0.0 Å and increasing in bins of width 0.2 Å up to a maximum value of 12.8 Å). **Figure 8** shows the eight autocorrelation functions for trimethoprim calculated with AM1.

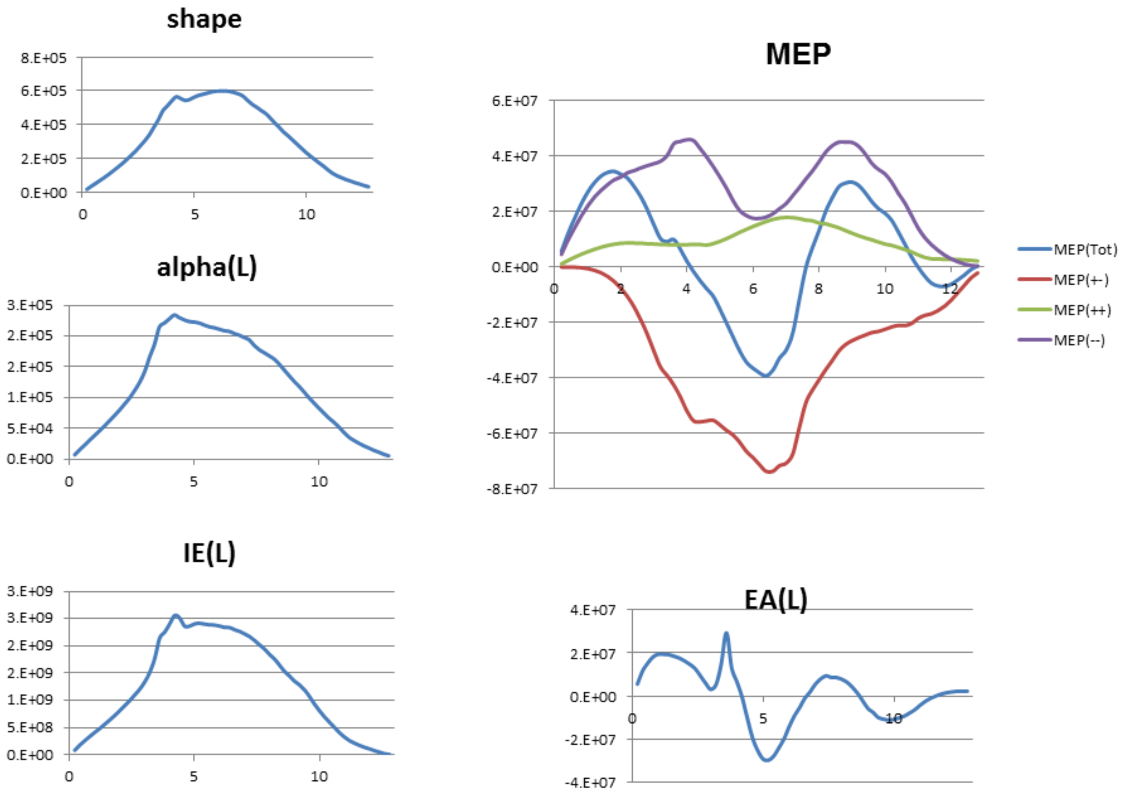


Figure 8 The eight autocorrelation functions calculated using the AM1 Hamiltonian for trimethoprim.

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, are calculated and written out. The floating-point Tanimoto similarities S are defined as:

$$S = \frac{\sum_{i=1}^{nbins} A_A^i A_B^i}{\sum_{i=1}^{nbins} \left(A_A^i\right)^2 + \sum_{i=1}^{nbins} \left(A_A^i\right)^2 - \sum_{i=1}^{nbins} A_A^i A_B^i} \qquad (19)$$

where A_A^i is the value of the autocorrelation function for molecule A and bin i etc. Note that although the normal range for a Tanimoto coefficient is from zero to one, marginally negative values may occur for the local ionisation energy or electron affinity.

These similarities are calculated for the entire range of each of the eight autocorrelation functions. These individual similarities can be written to a table file (see **0**) and are printed in the output file (see **3.5.4**).

1.16 Standard rotationally invariant fingerprints (RIFs)

Mavridis et al. [32] 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.

1.17 Maxima and minima of the local properties

Jakobi et al. [33] 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.

1.18 Atom-centred descriptors

Hennemann et al. [34] have used atom-centred quantities calculated by ParaSurf™ as descriptors in order to calculate the strengths of hydrogen bonds [34a] and for chemical reactivity models [34b]. These descriptors (based on conventional solvent-accessible surface areas [35] using Bondi van der Waals radii [36] and a default solvent radius of 1.4 Å), C-H bond orders for hydrogen atoms, the constitution of the localized lone-pair orbitals on nitrogen atoms and the π-charges of carbon atoms in conjugated π-systems. These descriptors are now output by ParaSurf19™.

1.19 Fragment analysis

ParaSurf19™ can divide the input molecule into fragments (which must be defined in the input SDF file) and perform a full surface analysis for each fragment. This option and its output will be described in detail below. ParaSurf19™ now outputs .psf and .sdf files for each fragment for use in CImatch19™ for substructure similarity.

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 4 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>.vwf` if a file with this name exists.
		
		Otherwise an SDF file `<name>_v.sdf` `<name>.sdf` will be used as input in the order given.
		
		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 = `0.0005`; default for marching-cube surface = `0.007`; 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 (gives ParaSurf'11 electrostatics with the "parasurf11" keyword, otherwise ParaSurf19)
	`newmp`	Use ParaSurf'12 or 19 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
`vmpfrag=`	`on`	Equivalent to `vmp=`, but writes separate `.vmp` files for each fragment with only its atoms and the MEP projected onto the fragment surface. The files are named `<filename><fragmentname>.vmp`, where `<fragmentname>` is the name assigned to the fragment in the input SDF file.
	`off`	No fragment .vmp files will be written.
	`all`	As for `on`, except that the atoms for the entire molecule are written to the .vmp files with the surface for the fragment only.

`grid=`	`<filename>`	Read the Cartesian coordinates at which to calculate a grid of $(\log_{10}(\rho))$, MEP, IE_L, EA_L, α_L, χ_L, η_L and their first derivatives in x, y and z-directions). See Section **3.9.1**
	`auto`	ParaSurf™ calculates an automatic grid that excludes areas closer than 0.5 Å to the nuclei (see Section **3.9.2**)
	`vdw`	ParaSurf™ calculates an automatic grid that excludes areas closer than the corresponding van der Waals radius to the nuclei
	`box`	ParaSurf™ calculates an automatic grid including all points regardless of their proximity to nuclei
	`surf`	The properties of the surface points are written to the .psf file
`lattice=`	`n.nn`	Sets the lattice spacing for the `grid=auto`, `vdw` **or** `box` option (see Section **3.9.2**)
`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	`on`	The atomic and surface coordinates in the `.psf` output file will be centred 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.
`centre=`	`off`	The atomic and surface coordinates in the `.psf` output file will not be centred 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`		Requests that the surface autocorrelation functions be calculated and written to the output `.sdf` file.
	`=<filename>`	`<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 autocorrelations will be written to the file `<filename>`. A total of 448 variables (7 properties in 64 bins each) are written for each molecule.

`aslist=`	`<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 centres 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 centre. 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 centres 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 centre. 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.
`fragments`		Perform a fragment analysis. The fragments must be defined in the input SDF file
`desfile=`	`<filename>`	Write the binned SIM descriptors 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. The descriptors are written as a comma-separated table with headers. Note that `desmodel` must also be defined.
`desmodel=`	`<code>`	The bin definitions for the model denoted by `<code>` will be used to calculate the descriptors for the table of binned SIM descriptors. The possible values of `<code>` and their definitions are given in **Table 2**.
`-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.

eal09		Do not use the selection procedure for virtual orbitals [**11**] when calculating the local electron affinity. This option provides continuity with earlier versions of ParaSurf™
parasurf11		Backwards compatibility option: electron densities, local properties and electrostatic potential and field are calculated using the algorithms from ParaSurf'11
precise		More precise output of the local properties in grid calculations
locpol=	**aniso**	Use the local polarizability calculated from anisotropic atomic polarizability tensors (default)
	old	Use isotropic atomic polarizabilities to calculate the local polarizability (implied by the "parasurf11" option)
no_derivatives		Do not calculate the first derivatives of the local properties (default is to calculate the derivatives)

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 centre for spherical-harmonic fits

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

```
<SPHH_CENTER>
```

The centre 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 centre of gravity. The capitalization of "`Atoms`" is required and the list of atoms is limited to one line.

2.2.2 Defining fragments

Molecular fragments can be defined in the input SDF file and fragments calculations requested using the fragments options, for instance

`parasurf test surf=cube fragments`

Figure 9 shows a sample <fragment> block from an SDF input file.

```
> <FRAGMENTS>
Start fragment
phenyl
   3   4   5  15  16  19  25  33
End fragment

Start fragment
methoxy1
   1   2  22  23  24
End fragment

Start fragment
methoxy2
  17  18  34  35  36
End fragment

Start fragment
methoxy3
  20  21  37  38  39
End fragment

Start fragment
methylene
   6  26  27
End fragment

Start fragment
thymine
   7   8   9  10  11  12  13  14  28  29  30  31  32
End fragment

Start fragment
everybody
   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20
  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39
End fragment

> <END_FRAGMENTS>
```

Figure 9 A sample <FRAGMENTS> input block.

The first line after each “**`Start fragment`**” line (note the upper and lower case, which are necessary) defines the name given to the fragment. This is followed by the numbers of the atoms that make up the fragment (20i4, fixed format). Note that the fragments need not be mutually exclusive. The fragment “**`everybody`**” in the above example, for instance is the entire molecule. The fragment-definition block begins with

`> <FRAGMENTS>`

and ends with

`> <END_FRAGMENTS>`

tags.

Figure 10 shows the input molecule and the fragments.

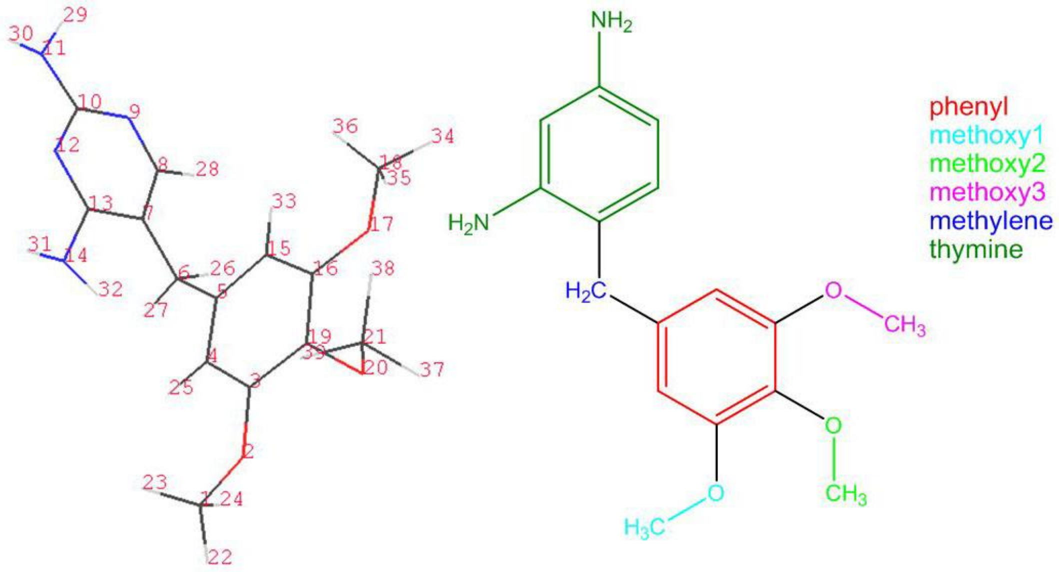


Figure 10 The fragments defined in the SDF input example.

In a FRAGMENTS run, ParaSurf™ first performs a calculation for the entire molecule and then analyses the molecular surface according to the standard ParaSurf™ technique used to assign surface triangles to individual atoms. The output for the phenyl fragment is shown in **Figure 11**. A similar output section is printed for each fragment. The results and the descriptors for each fragment are taken from the surface for the whole molecule and therefore refer to the fragment (both its electronic properties and the area of its surface) within the context of the molecule itself. The coordinates given for the maxima and minima of the local properties refer to the input geometry of the entire molecule.

```
<> Results for fragment number  1 : phenyl
Surface area    :        47.21 Angstrom**2
Fragment charge :        -0.01

               MEP     IEL      EAL     HARD     ENEG      F(N)     POL
Mean    :     -6.0    402.5    -67.6   235.1    167.5     -4.0      0.76
Mean +ve:     10.9               0.0                       8.6
Mean -ve:    -13.0             -67.6                      -8.7
Maximum :     18.9    525.3    -19.9   303.2    222.2     16.9       1.4
Minimum :    -53.1    337.6   -103.9   191.6    142.3    -42.4      0.33
Variance:    200.3    919.4    562.5   594.8    146.2     85.0      0.54E-01
Var. +ve:     29.8               0.0                      21.9
Var. -ve:    103.2             562.5                      28.5
Balance :    0.174             0.000                     0.246
Skew    :     -0.4      0.7      0.2    -0.1      1.1      1.6     -0.41
Kurtosis:      0.4      0.6     -1.3    -1.0      1.7     -0.2     -0.72

MEP Maxima for this fragment
     Number      x         y         z        MEP
      4 :     1.6569    1.8397   -2.5118   15.9683
      6 :    -4.3431   -1.6853   -1.1618   18.8870

IEL Maxima for this fragment
     Number      x         y         z        IEL
      1 :     1.6861   -2.6103   -3.4142  525.3250

EAL Maxima for this fragment
     Number      x         y         z        EAL
      1 :    -1.4714   -2.0603   -3.9285  -22.4779
      2 :     0.0069   -1.2520   -4.1285  -19.8861
      3 :    -2.5431   -0.8270   -3.5285  -27.9933
      4 :     0.8569   -1.1103   -0.6285  -25.1242
      6 :    -0.7431   -2.4270   -0.4785  -28.5599

POL Maxima for this fragment
     Number      x         y         z        POL
      1 :    -0.7431   -3.7603   -3.4785    1.3387
      2 :    -0.8648   -3.8603   -3.2785    1.3698

POL Minima for this fragment
     Number      x         y         z        POL
     23 :    -4.1431   -1.0103   -2.5285    0.3418
     33 :     1.6569    1.1897   -1.0285    0.3529

FN Maxima for this fragment
     Number      x         y         z        FN
     12 :     1.2569    2.0897   -1.7902   14.2146
     15 :    -4.2931   -1.0603   -1.0285   15.7029
```

Figure 11 ParaSurf19™ output for the phenyl fragment defined above.

The individual surfaces of the fragments are shown in **Figure 12**.

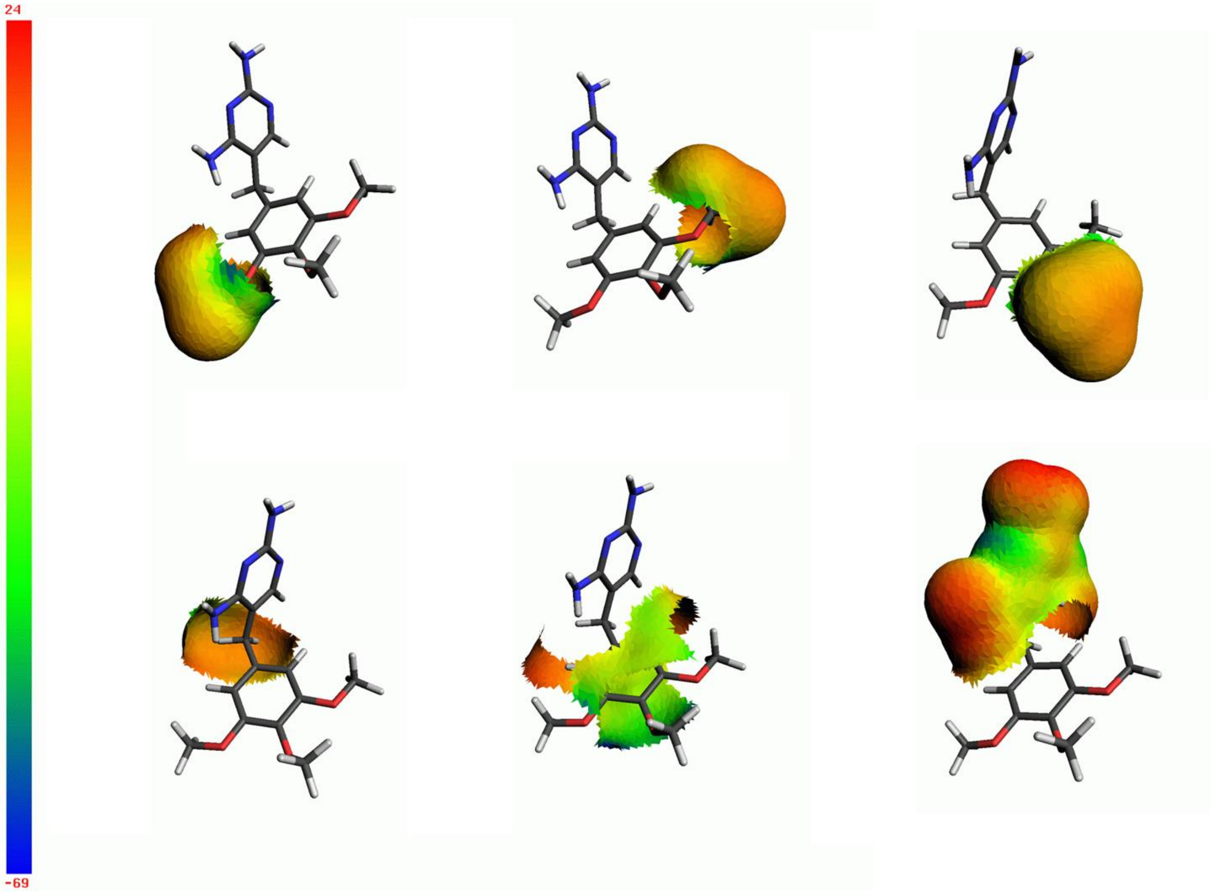


Figure 12 Surfaces calculated for the individual fragments, colour coded according to the MEP in kcal mol^{-1}. The fragments (clockwise from the top right) are methoxy1, methoxy2, methoxy3, thymine, phenyl and methylene.

ParaSurf19™ writes both .psf and abbreviated (only atoms and bonds) .sdf output files for each fragment. These files are named **<molecule>_<fragment>_e.psf** and **<molecule>_<fragment>_e.sdf**, where **<fragment>** is the fragment name defined in the input .sdf file.

These two files are needed for substructure matching using CImatch™.

3 INPUT AND OUTPUT FILES

ParaSurf™ uses the following files for input and output:

Table 5 ParaSurf™ input and output files

File	Name	Description

Input	`<filename>.vwf`
	(if available) or
	`<filename>_v.sdf`
	or
	`<filename>.sdf`
	or
	`<filename>_m.sdf`
		EMPIRE™ `.vwf` file
		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 CeposMopac 6 `.sdf` file. It is strongly recommended to use the `EF` option for geometry optimizations in Mopac.

Hamiltonian	`<Hamiltonian>.par`	The EMPIRE parameters file (found in the EMPIRE etc directory). The environment variable `EMPIRE_ROOT` must be set to point to this directory. The name `<Hamiltonian>` will be taken from the input SDF file. Calculations using the hpCADD Hamiltonian must use the `.vwf` file as input because atom types are not defined in SDF files.
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.
Binned SIM descriptor file	`User defined`	An ascii, comma-separated file that contains a line of the descriptors generated for the bin definitions used in the model defined by `<code>` in the `desmodel=` command-line option. A header defining the descriptors is printed as the first line.
Autocorrelation fingerprint file	`User defined`	An ascii, comma-separated file that contains the molecule's ID and 448 binned autocorrelation values. The file will be overwritten if it exists
Autocorrelation similarity file	`User defined`	An ascii, fixed format file that contains a line of seven 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 [36]) 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 13**. 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 13 The headers and titles, atomic coordinates and bond definitions from a VAMP .sdf file. The format follows the MDL definition. [26].

<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 6 Hamiltonians and the available electrostatic and polarizability models.

Hamiltonian	Reference	Electrostatics		Local Polarizability
		NAO-PC	Multipole	
MNDO	[20b]	YES	YES	YES
AM1	[22]	YES	YES	YES
PM3	[23]	YES	YES	YES
MNDO/c	[38]	YES	YES	NO
MNDO/d	[20a]	NO	YES	NO
AM1*	[24a]	NO	YES	NO
RM1	[39]	NO	YES	NO
PM6	[40]	NO	YES	NO
hpCADD	[41]	NO	YES	NO
MNDO-F	[42]	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:

- Polarisabilities 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 <Hamiltonian>.par file

The file <Hamiltonian>.par is used by EMPIRE to define the named Hamiltonian and elements and supply the parameters. This file is also used by ParaSurf™ for the same purpose. The <Hamiltonian>.par file is not necessary if an EMPIRE™ .vwf file is used as input.

3.4 The EMPIRE™ binary wavefunction file (.vwf)

The binary wavefunction file contains all the information necessary to process the results of the EMPIRE™ calculation further. It is the new primary input file format for ParaSurf and does not require a Vhamil.par file to be present. Its contents are:

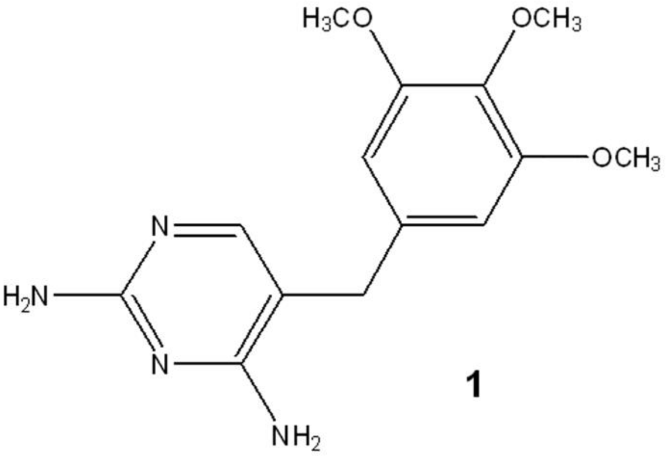
Definition		Type
First comment line from the input		character(len=80)
Title line from the input		character(len=80)
Hamiltonian		character(len=6) "AM1 ", "AM1* ", "MNDO ","MNDO/c","MNDO/d", "PM3 " or "PM6 "
Formalism		character(len=3) "RHF"
Number of atoms		integer
Number of orbitals (Norbs)		integer
Number of doubly occupied orbitals		integer
Number of singly occupied orbitals		Integer = 0
Charge on the molecule		integer
Heat of formation in kcal mol^{-1}		double precision
Energy of the HOMO (eV)		double precision
Energy of the LUMO (eV)		double precision
x, y and z-components of the dipole moment (Debye)		double precision(1:3)
For each atom:		
	Atomic number	integer
	Cartesian coordinates	double precision(1:3)
	Number of atomic orbitals	integer
	Principal quantum number (s and p)	integer
	Principal quantum number (d)	integer
	Slater exponents (ζ_s, ζ_p, ζ_d)	double precision(1:3)
	Overlap integrals (β_s, β_p, β_d)	double precision(1:3)
	Multipole parameters (dd and qq)	double precision(1:2)
End atoms		
Eigenvalues (eV)		Double precision(1:Norbs)
Eigenvectors		Double precision(1:Norbs^2)

3.5 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.

3.5.1 For a spherical-harmonic surface

Figure 14 shows the output for a calculation using the options **`surf=wrap fit=sphh translate`** for trimethoprim, **1**.



1

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
<> ParaSurf'19
 <> Copyright (c) 2006-2019 Friedrich-Alexander-Universitaet Erlangen-Nuernberg
                   and Cepos InSilico GmbH All rights reserved.

 <> Input = trimethiprim.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 =  0.5000E-03 electrons/Angstrom**3
    Triangulation mesh =    0.20 Angstrom
    Using multipole electrostatics


 <> AM1   calculation for  Trimethoprim

 <> Translated spherical-harmonic fits:
      dx      dy      dz     rmsd
    0.0000  0.0000  0.0000  0.3827
    0.5000  0.0000  0.0000  0.5516
   -0.5000  0.0000  0.0000  0.4939
```

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
     0.0000  0.5000  0.0000  0.5333
     0.0000 -0.5000  0.0000  0.5173
     0.0000  0.0000  0.5000  0.5406
     0.0000  0.0000 -0.5000  0.4147


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

        0        2.01177719
        1        2.06264149
        2        1.57429428
        3        1.14144633
        4        0.96586545
        5        0.69034677
        6        0.60076280
        7        0.51821086
        8        0.48396719
        9        0.45955001
       10        0.42519621
       11        0.39310379
       12        0.37725481
       13        0.36600493
       14        0.34829905
       15        0.32860993

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

        0       12.40803637
        1       12.34784414
        2        9.25211030
        3        8.29692710
        4        6.82440697
        5        5.74002976
        6        4.82285463
        7        4.20321063
        8        3.80848913
        9        3.53212800
       10        3.08074911
       11        2.62541255
       12        2.37675753
       13        2.06595225
       14        1.98683370
       15        1.84188725
       16        1.70418851
       17        1.54855928
       18        1.34322707
       19        1.21349689
       20        1.08035065

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

        0       44.57878887
        1       39.49221723
        2       37.79258047
        3       36.00326805
        4       32.26724224
        5       29.12695030
        6       26.30443933
        7       25.27443867
        8       23.78306670
        9       22.00750560
       10       21.22324848
       11       20.31770921
       12       19.17042684
       13       17.99969469
       14       17.10345128
       15       16.83692340
       16       16.18125074
```

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
      17      14.74040808
      18      13.59009122
      19      12.96531563
      20      12.96531563

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

       0      14.76660122
       1      14.57921630
       2      14.82522427
       3      11.54502778
       4      10.82050211
       5       9.79342370
       6       9.76520958
       7       9.37069356
       8       8.72479764
       9       8.01299396
      10       7.52757624
      11       7.18259970
      12       6.90528697
      13       5.87616823
      14       5.39515470
      15       4.92869560
      16       4.68556059
      17       4.48184805
      18       4.20875146
      19       4.02983055
      20       3.88447733

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

       0      10.23375986
       1      10.22721474
       2       9.38433421
       3       9.05353285
       4       8.17889165
       5       7.54963618
       6       6.97197468
       7       6.65848164
       8       6.39618968
       9       6.09545254
      10       5.46552427
      11       4.86094641
      12       4.46112161
      13       4.12120028
      14       4.03929276
      15       3.83138764
      16       3.55514598
      17       3.21320455
      18       2.77835041
      19       2.50117320
      20       2.28148569

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

       0       0.24597908
       1       0.24673395
       2       0.23405408
       3       0.20695856
       4       0.18909827
       5       0.17468336
       6       0.16936390
       7       0.15804017
       8       0.14856557
       9       0.13972827
      10       0.12529394
      11       0.11480278
      12       0.10565425
```

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
     13       0.09908593
     14       0.09602514
     15       0.09019338
     16       0.08240006
     17       0.07510580
     18       0.06885870
     19       0.06580610
     20       0.06287666

<> Property ranges:
   Density   :   0.1004E-03 to   0.1836E-02
   IE(l)     :       325.65 to       554.64
   EA(l)     :      -112.61 to       -29.37
   MEP       :       -50.22 to        19.02
   Alpha(l)  :       0.2167 to       1.5089
   Field(N)  :       -52.89 to        19.81

<> Descriptors :

   Dipole moment        :       1.2467 Debye
   Dipolar density      :   0.2445E-02 Debye.Angstrom**-3
   Molecular pol.       :     128.5408 Angstrom**3
   Molecular weight     :       290.32
   Globularity          :       0.7559
   Total surface area   :       408.28 Angstrom**2
   Molecular volume     :       509.81 Angstrom**3

   Most positive MEP    :        19.02 kcal/mol
   Most negative MEP    :       -50.22 kcal/mol
   Mean +ve MEP         :         6.28 kcal/mol
   Mean -ve MEP         :       -11.89 kcal/mol
   Mean MEP             :        -2.37 kcal/mol
   MEP range            :        69.24 kcal/mol
   MEP +ve Variance     :        15.50 (kcal/mol)**2
   MEP -ve Variance     :       133.21 (kcal/mol)**2
   MEP total variance   :       148.71 (kcal/mol)**2
   MEP balance parameter:       0.0934
   MEP balance*variance :      13.8883 kcal/mol
   MEP skewness         :      -1.3157
   MEP kurtosis         :       1.5929
   Integral MEP         :  -823.474     kcal.Angstrom**2/mol

   Maximum IE(l)        :       554.64 kcal/mol
   Minimum IE(l)        :       325.65 kcal/mol
   Mean IE(l)           :       422.44 kcal/mol
   IE(l) range          :       228.99 kcal/mol
   IE(l) variance       :      1947.25 (kcal/mol)**2
   IE(l) skewness       :       0.6846
   IE(l) kurtosis       :      -0.3181
   Integral IE(l)       :   7450.37     eV.Angstrom**2

   Maximum EA(l)        :       -29.37 kcal/mol
   Minimum EA(l)        :      -112.61 kcal/mol
   Mean +ve EA(l)       :         0.00 kcal/mol
   Mean -ve EA(l)       :       -93.26 kcal/mol
   Mean EA(l)           :       -93.26 kcal/mol
   EA(l) range          :        83.24 kcal/mol
   EA(l) +ve variance   :         0.00 (kcal/mol)**2
   EA(l) -ve variance   :       203.49 (kcal/mol)**2
   EA(l) total variance :       203.49 (kcal/mol)**2
   EA(l) skewness       :       1.5615
   EA(l) kurtosis       :       3.0484
   Integral EA(l)       :  -1647.42     eV.Angstrom**2
   EA(l) balance param. :       0.0000
   Fraction pos. EA(l)  :       0.0000 ( =    0.00 Angstrom**2)

   Max. local Eneg.     :       234.22 kcal/mol
   Min. local Eneg.     :       113.00 kcal/mol
   Mean local Eneg.     :       164.59 kcal/mol
   Local Eneg. range    :       121.21 kcal/mol
   Local Eneg. variance :       453.76 (kcal/mol)**2
```

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
    Local Eneg. skewness :         0.66
    Local Eneg. kurtosis :         0.01
    Integral local Eneg. :   2901.47     eV.Angstrom**2

    Max. local hardness  :       321.14 kcal/mol
    Min. local hardness  :       187.77 kcal/mol
    Mean local hardness  :       257.85 kcal/mol
    Local hard. range    :       133.37 kcal/mol
    Local hard. variance :       621.61 (kcal/mol)**2
    Local hard. skewness :         0.39
    Local hard. kurtosis :        -0.39
    Integral local Hard. :   4548.90     eV.Angstrom**2

    Maximum alpha(l)     :    1.509     Angstrom**3
    Minimum alpha(l)     :   0.2167     Angstrom**3
    Mean alpha(l)        :   0.4780     Angstrom**3
    Alpha(l) range       :    1.292     Angstrom**3
    Variance in alpha(l) :   0.5893E-01 Angstrom**6
    Alpha(l) skewness    :       1.5346
    Alpha(l) kurtosis    :       1.5720
    Integral Alpha(l)    :   194.797     Angstrom**5

    Maximum field normal :    19.81 kcal/mol.Angstrom
    Minimum field normal :   -52.89 kcal/mol.Angstrom
    Mean field           :    -0.27 kcal/mol.Angstrom
    Field range          :    72.70 kcal/mol.Angstrom
    Total field variance :   104.50 (kcal/mol.Angstrom)**2
    +ve field variance   :    11.74 (kcal/mol.Angstrom)**2
    -ve field variance   :   116.79 (kcal/mol.Angstrom)**2
    Field balance param. :     0.08
    Field skew           :     2.46
    Field kurtosis       :   5.021
    Integral F(N)        :   44.29     kcal.Angstrom/mol
    Integral F(N +ve)    :   1456.     kcal.Angstrom/mol
    Integral F(N -ve)    :  -1411.     kcal.Angstrom/mol
    Integral |F(N)|      :   2867.     kcal.Angstrom/mol

 <> Spherical-Harmonic Hybridization:

    Shape hybrids        :
   16.009628    1.373167    3.546672    2.694125    1.350668    1.644624
    0.762716    0.658342    0.375899    0.356134    0.403193    0.301343
    0.190579    0.199740    0.230850    0.204322

    MEP hybrids          :
    7.934381    4.585824   29.916318   11.316554   14.111406   11.572654
   10.218143    8.395453    5.647876    4.106887    4.808912    4.310994
    3.286470    3.516688    1.782163    2.046608    1.478736    1.563289
    1.807359    1.409069    1.370528

    IE(l) hybrids        :
   1475.0734     69.4741     57.6041     47.8482     44.9670     44.0421
     43.8721     28.8369     37.2336     34.3592     20.0558     16.1335
     20.3418     18.4038     18.1736     13.1804     13.3178     17.9975
     17.9273     16.0483      0.0000

    EA(l) hybrids        :
    317.0698      7.7090     18.9862     33.3973     15.7480     17.4261
     11.2668      9.8258     11.5616     10.4208      9.5707      8.2159
      6.5635      9.7503      6.2945      7.1239      4.4548      4.3692
      4.8386      4.2292      3.5039

    Alpha(l) hybrids     :
  1.83506449  0.07660952  0.42325808  0.37116637  0.26168092  0.25387096
  0.19306697  0.22512754  0.17314967  0.16214546  0.15522714  0.12040044
  0.12344010  0.10628415  0.08308407  0.08170594  0.07634731  0.07232668
  0.07259004  0.05268140  0.04875321

    Field(N) hybrids     :
      2.6299      2.0991     15.8335      8.7229     12.1467     10.8157
      9.8725      9.3441      7.0443      5.6697      7.0721      6.5543
```

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
      6.2485      6.0541      3.3904      3.6946      3.2560      3.7721
      4.2037      3.4602      2.9100

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

       4.00120       1.17182       1.88326       1.64138       1.16218
       1.28243      0.873336       2.81680       2.14145       5.46958
       3.36401       3.75652       3.40186       3.19658       2.89749
       2.37653       2.02655       2.19292       2.07629       1.81286
       38.4067       8.33511       7.58974       6.91724       6.70575
       6.63643       6.62360       17.8065       2.77652       4.35732
       5.77904       3.96837       4.17446       3.35661       1.35465
      0.276784      0.650583      0.609234      0.511548      0.503856
      0.439394       1.62171       1.44884       3.97914       2.95346
       3.48521       3.28872       3.14205       3.05681       2.65411
       2.38110       2.65934       2.56013       2.49969

 <> Atomic surface properties:


 Atom     Area         MEP             IE(l)           EA(l)      mean      Field(N)         Eneg(L)         Hard(L)
                      max     min     max     min     max     min    pol.     max     min     max     min     max     min
C    1   1.190   -4.99  -24.60  474.40  440.53  -84.14  -97.49   0.695   -0.83  -19.62  195.13  171.52  279.27  268.74
O    2   3.298  -15.21  -48.75  483.72  395.45  -68.96  -88.24   1.013  -10.63  -35.96  204.79  153.60  278.92  238.20
C    3   5.583    2.76  -42.26  473.24  343.23  -31.92 -102.63   0.845    3.14  -23.16  199.07  137.84  274.57  192.42
C    4   2.301   -1.09  -11.01  420.74  358.07  -47.79  -88.79   0.765    0.11   -7.46  167.86  147.78  252.89  202.93
C    5   1.315   -2.83  -11.04  443.62  381.62  -69.72  -93.99   0.780   -2.54   -6.76  174.81  152.40  268.81  226.23
C    6   0.000
C    7   1.851   -1.56  -11.81  420.70  353.35  -50.94  -93.17   0.866   -0.53   -7.17  169.93  141.70  250.77  207.08
C    8   6.600    0.72  -22.08  435.21  362.23  -35.07  -99.15   0.779    2.08  -15.47  176.51  150.00  258.95  202.94
N    9   7.657   -9.91  -40.75  443.85  328.33  -82.12 -109.10   1.186  -10.05  -52.89  175.00  113.00  269.71  214.57
C   10  10.568    5.91  -23.31  493.46  388.30  -37.01  -96.70   0.791    4.70  -15.87  208.30  154.49  288.55  213.45
N   11   1.649  -14.33  -23.08  511.29  446.71  -63.02  -82.87   0.793  -16.61  -22.27  216.19  190.10  295.11  254.86
N   12   6.592   -9.43  -39.67  444.68  325.65  -80.18 -106.59   1.031  -10.86  -51.95  172.11  114.98  273.94  206.65
C   13   7.498    4.98  -25.74  511.74  366.79  -34.15  -96.77   0.813    6.63  -21.17  215.43  152.07  296.31  201.46
N   14   1.811  -14.21  -22.59  489.48  443.66  -66.69  -87.80   0.801  -18.53  -29.45  206.02  184.52  285.56  256.30
C   15   4.876   -0.83  -11.93  434.24  358.03  -52.14 -100.50   0.747    2.70   -9.39  173.03  145.57  261.21  205.09
C   16   5.487   -2.05  -37.60  476.55  346.16  -29.37  -99.54   0.821    2.98  -22.97  200.45  139.16  276.10  187.77
O   17   0.967  -23.73  -38.37  475.37  409.71  -71.36  -88.47   0.959  -11.16  -33.69  198.69  165.37  276.68  241.01
C   18   0.868    9.69  -22.92  505.17  434.80  -83.99  -98.88   0.690    5.54  -12.51  208.01  167.96  297.16  266.84
C   19   5.387   -1.43  -46.05  485.13  325.82  -36.16 -103.21   1.007   -5.13  -29.69  201.97  133.24  286.16  189.70
O   20   4.532  -34.23  -50.22  477.62  372.36  -73.80  -95.94   1.086  -18.29  -43.77  201.91  143.17  279.48  228.85
C   21   1.026  -20.40  -37.71  465.20  425.64  -93.36 -109.78   0.771   -4.77  -21.36  184.55  157.93  280.68  267.71
H   22  24.370    9.65  -29.35  508.71  383.58  -87.48  -97.43   0.402    6.63  -16.30  209.44  144.98  299.27  238.60
H   23  17.654   12.04  -11.12  503.85  382.54  -76.95  -97.30   0.370    7.51   -8.90  207.29  145.64  296.55  236.85
H   24  18.559   11.71  -14.40  502.99  382.80  -77.32  -96.21   0.366    6.31   -3.91  206.62  145.37  296.38  237.44
H   25   6.426   12.26    0.38  450.48  398.83  -84.46  -99.09   0.375    9.18   -4.08  178.56  152.24  271.92  243.04
H   26  14.980    8.87   -4.47  463.19  374.80  -88.84 -102.54   0.381    6.24   -3.25  182.47  139.03  281.34  235.77
H   27  14.448    8.05  -15.74  466.07  374.02  -85.09 -102.86   0.384    5.61   -6.12  190.49  137.76  279.96  236.26
H   28  18.509    9.52  -25.29  438.41  392.07  -70.47 -102.76   0.407    8.09  -22.53  174.36  149.49  270.08  233.30
H   29  23.099   18.93  -25.76  550.91  428.54  -78.35 -111.77   0.334   19.12  -24.36  232.83  159.04  321.14  269.51
H   30  22.784   19.02  -25.71  549.21  416.66  -78.19 -111.79   0.316   19.81  -23.60  234.22  154.05  320.21  262.61
H   31  22.910   15.18  -25.53  554.64  420.49  -77.17 -112.61   0.314   13.89  -26.30  234.06  155.95  320.58  264.54
H   32   8.916   14.53  -11.92  552.74  397.85  -78.64 -104.83   0.357   14.60  -19.29  232.60  149.52  320.13  248.33
H   33   6.289   10.08   -4.74  441.04  384.41  -73.12  -98.75   0.443    8.53   -6.17  172.27  147.83  268.77  228.76
H   34  23.001    9.47  -23.20  498.69  384.37  -87.93  -99.94   0.374    6.36  -15.55  204.62  145.09  294.07  238.62
H   35  18.462    9.75  -15.61  500.50  380.61  -73.19  -96.24   0.380    6.52   -6.87  205.69  145.35  294.81  231.10
H   36  18.392    9.64   -6.07  502.41  382.26  -80.84  -97.20   0.382    8.17  -12.02  206.76  145.20  295.65  237.06
H   37  24.577    3.25  -31.68  487.62  370.69  -99.58 -109.64   0.394    3.95  -22.89  192.74  133.11  294.88  237.57
H   38  17.255    2.55  -34.53  495.67  367.07  -87.27 -109.17   0.387    2.36  -21.41  196.52  130.24  299.15  236.77
H   39  21.193    2.66  -35.53  490.81  359.03  -87.36 -108.60   0.376    2.69  -25.26  194.23  130.19  296.58  226.39

Total   402.881

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

          x            y            z       value

 <>       9 MEP Maxima         :
     -2.0943      -1.9103       1.9958      14.48
      1.9308       2.6575      -1.9070      10.08
      4.2004       2.2889      -1.9835      9.694
      1.4638       5.7589       5.4770      12.75
     -1.4292      -1.7902       3.3075      15.18
     -0.1036       3.5080      -2.4658      9.524
```

ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface

```
    -4.5244     -1.3404     -0.6240     12.26
     1.9409      3.3233      5.3060     17.29
     2.3347      3.9929      4.8681     19.02
<>      5 MEP Minima      :
     0.6146      5.8476      1.0368    -40.75
     0.0000     -4.3478     -2.5240    -50.22
    -1.4714      2.3729      4.3876    -39.67
     0.8011     -3.6508     -3.4297    -45.17
     2.6850     -2.2208     -2.9412    -38.37

<>      7 IEL Maxima      :
    -3.0425     -1.2046      2.7457     554.6
     0.2652     -0.8162      2.3579     538.2
    -0.1428     -1.3587      2.3663     539.8
    -1.9644     -1.8860      2.7447     546.7
    -0.9455     -1.8557      2.5130     543.5
     2.1930      5.5001      5.4578     550.9
     1.0590      5.7161      5.3691     550.9
<>      4 IEL Minima      :
    -1.3070      2.2637      4.5275     325.7
     0.2228     -2.1196     -3.6915     325.8
    -0.2634      1.6256      4.6963     332.5
     0.9079      5.7325      1.0276     328.3

<>      4 EAL Maxima      :
    -2.9083      2.6186      2.2719    -34.15
    -2.2332      4.3872      0.4313    -35.07
    -1.0955     -2.0487     -3.3586    -31.92
     0.7486     -1.0304     -3.6793    -29.37
<>      8 EAL Minima      :
     0.2957     -5.5123     -1.4812    -109.8
     1.1122     -5.2323     -3.1054    -107.1
    -0.9624     -0.1287      4.9079    -112.6
     2.6858     -2.4183      0.9711    -106.7
     3.1809     -3.6014     -3.0616    -109.3
     4.1893     -2.5658      0.2149    -106.6
     0.5536      2.7836      5.8582    -111.8
     1.8617      6.2852      2.3982    -111.8

<>      1 Alpha(l) Maxima :
     0.9079      5.7325      1.0276     1.509
<>     43 Alpha(l) Minima :
     0.1941     -0.5973      3.5616    0.2183
    -2.6409     -5.9315      0.0000    0.3151
     0.0000     -5.4569      0.0000    0.3507
    -3.8091      2.1992     -1.6009    0.3088
    -3.2197      2.3393     -2.2977    0.3049
     4.7632     -0.6017     -4.0286    0.3010
     2.0757      1.6767      0.7160    0.3543
    -1.7867      4.6716     -1.3420    0.2991
    -3.8319      1.4763      1.1018    0.3538
    -3.3602     -0.9081      1.6253    0.2685
     0.0000     -1.1114      3.2106    0.2167
    -1.8684     -5.7503     -2.2128    0.3172
     2.5578     -5.0048      1.5081    0.3072
     1.3439      3.4788     -1.7414    0.3159
     1.3551      2.3352     -2.7118    0.3274
    -4.8888     -2.1766      0.0000    0.3256
     3.1679      0.2910     -4.5679    0.2972
     2.2360      1.3732      0.8281    0.3515
    -2.4257      0.7858      4.4899    0.2262
    -2.5222     -3.6775      1.4074    0.3008
     0.1013     -3.7287      1.1772    0.3636
    -1.7602     -5.9235     -1.3715    0.3161
    -5.0212     -3.6481      0.5445    0.3369
    -2.4804     -5.5710     -2.8512    0.3166
     4.0476     -0.2121      0.7176    0.2939
     2.4151      5.0624      5.1803    0.2208
     3.0686      2.6189     -0.8974    0.3557
    -0.4681      4.4656      5.9016    0.2285
     4.6449     -3.9692      0.2672    0.3134
```

ParaSurf™ output for trimethoprim, 1, using a spherical-harmonic surface

```
       4.4728     -3.4303     -1.7844     0.3240
      -4.8828     -1.8743     -1.9115     0.3157
      -3.9258     -0.7208     -2.0799     0.3162
      -3.3468     -3.7159     -3.8577     0.3006
       3.5219     -6.1001     -0.6180     0.3331
       3.0798     -4.7429     -3.2785     0.3043
       5.4843     -1.0145     -0.7349     0.2994
       5.7718      1.0663     -2.4400     0.3606
       1.2318     -3.0701      1.4574     0.3178
       2.0933      5.6781      2.0628     0.2651
       2.5348      4.8420      2.6986     0.2362
       2.0840      3.3480      4.9571     0.2216
       2.6333      4.8195      3.3405     0.2301
       0.3362      6.5480      2.7231     0.2358

 <>      13 F(N) Maxima      :
      -0.5388     -1.6584      2.0781      14.28
       1.1425      5.9891      5.1567      12.89
      -1.6503     -1.7170      3.4429      13.58
       0.3958      5.0472      6.0897      10.52
       2.1539      3.5125      4.9561      17.87
      -0.2643      0.4295      4.5979      11.42
      -1.4216     -0.3913      4.9111      11.92
      -0.0484     -1.0597      3.5333      13.59
       1.4978      2.9955      5.6177      16.98
       1.9409      3.3233      5.3060      18.58
       2.5161      4.5343      4.5739      14.90
       2.6834      5.1538      3.8993      18.03
       2.6859      5.2414      3.5885      18.51
 <>      13 F(N) Minima      :
      -0.5249      1.6155      4.6668     -47.69
      -1.3070      2.2637      4.5275     -45.62
       2.8614     -2.0789     -2.9678     -31.33
      -3.4167      1.0124      2.5040     -26.32
      -0.7780     -4.8993     -0.4346     -30.95
      -1.2131      5.3508      4.2164     -21.65
      -0.6484     -4.9585     -1.1099     -31.95
       0.9079      5.7325      1.0276     -50.35
       0.0536     -4.2250     -2.7041     -41.20
      -0.0608      1.3326      4.4429     -48.57
       3.4583     -1.4318      0.1637     -12.76
      -0.6375     -4.8517     -1.5491     -33.01
       1.0471     -3.6236     -3.4752     -41.55

 <> ParaSurf used         7.26 seconds CPU time
```

Figure 14 ParaSurf™ output for trimethoprim, **1**, using a spherical-harmonic surface.

After printing the program options, ParaSurf19™ prints the shift in coordinates of the centre 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.

ParaSurf19™ 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) [**36**] 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 [**32**] 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.

3.5.2 For a marching-cube surface

Figure 15 shows the output for a calculation using the options `surf=cube` for trimethoprim.

ParaSurf™ output for trimethoprim using a marching-cube surface.

```
<> ParaSurf'19
 <> Copyright (c) 2006-2019 Friedrich-Alexander-Universitaet Erlangen-Nuernberg
                   and Cepos InSilico GmbH All rights reserved.

 <> Input = trimethiprim.sdf


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



 <> Program options :

    Using marching-cube isodensity surface
    Surface fitting turned off
    Using an isodensity surface contour
    Isodensity value =  0.7000E-02 electrons/Angstrom**3
    Triangulation mesh =    0.20 Angstrom
    Using multipole electrostatics


 <> AM1   calculation for  Trimethoprim
 <> Number of triangles =  12956
 <> Number of unique points :   6483

 <> Property ranges:
    Density    :   0.6700E-02 to   0.7334E-02
    IE(l)      :       312.99 to       539.60
```

ParaSurf™ output for trimethoprim using a marching-cube surface.

```
   EA(l)     :      -112.06 to       -19.89
   MEP       :       -64.18 to        28.35
   Alpha(l)  :       0.2319 to       1.5629
   Field(N)  :       -96.37 to        62.75

<> Descriptors :

   Dipole moment        :       1.2467 Debye
   Dipolar density      :   0.4108E-02 Debye.Angstrom**-3
   Molecular pol.       :     128.5408 Angstrom**3
   Molecular weight     :       290.32
   Globularity          :       0.6828
   Total surface area   :       319.86 Angstrom**2
   Molecular volume     :       303.48 Angstrom**3

   Most positive MEP    :        28.35 kcal/mol
   Most negative MEP    :       -64.18 kcal/mol
   Mean +ve MEP         :        10.17 kcal/mol
   Mean -ve MEP         :       -19.06 kcal/mol
   Mean MEP             :        -3.76 kcal/mol
   MEP range            :        92.53 kcal/mol
   MEP +ve Variance     :        41.19 (kcal/mol)**2
   MEP -ve Variance     :       267.51 (kcal/mol)**2
   MEP total variance   :       308.70 (kcal/mol)**2
   MEP balance parameter:       0.1156
   MEP balance*variance :      35.6945 kcal/mol
   MEP skewness         :      -1.0392
   MEP kurtosis         :       0.6468
   Integral MEP         :  -1121.97     kcal.Angstrom**2/mol

   Maximum IE(l)        :       539.60 kcal/mol
   Minimum IE(l)        :       312.99 kcal/mol
   Mean IE(l)           :       419.37 kcal/mol
   IE(l) range          :       226.62 kcal/mol
   IE(l) variance       :      1693.06 (kcal/mol)**2
   IE(l) skewness       :       0.5370
   IE(l) kurtosis       :      -0.2654
   Integral IE(l)       :   5814.56     eV.Angstrom**2

   Maximum EA(l)        :       -19.89 kcal/mol
   Minimum EA(l)        :      -112.06 kcal/mol
   Mean +ve EA(l)       :         0.00 kcal/mol
   Mean -ve EA(l)       :       -86.90 kcal/mol
   Mean EA(l)           :       -86.90 kcal/mol
   EA(l) range          :        92.17 kcal/mol
   EA(l) +ve variance   :         0.00 (kcal/mol)**2
   EA(l) -ve variance   :       363.34 (kcal/mol)**2
   EA(l) total variance :       363.34 (kcal/mol)**2
   EA(l) skewness       :       1.3463
   EA(l) kurtosis       :       1.2524
   Integral EA(l)       :  -1215.56     eV.Angstrom**2
   EA(l) balance param. :       0.0000
   Fraction pos. EA(l)  :       0.0000 ( =    0.00 Angstrom**2)

   Max. local Eneg.     :       231.55 kcal/mol
   Min. local Eneg.     :       108.12 kcal/mol
   Mean local Eneg.     :       166.23 kcal/mol
   Local Eneg. range    :       123.43 kcal/mol
   Local Eneg. variance :       417.35 (kcal/mol)**2
   Local Eneg. skewness :         0.33
   Local Eneg. kurtosis :        -0.11
   Integral local Eneg. :   2299.50     eV.Angstrom**2

   Max. local hardness  :       312.13 kcal/mol
   Min. local hardness  :       191.56 kcal/mol
   Mean local hardness  :       253.13 kcal/mol
   Local hard. range    :       120.57 kcal/mol
   Local hard. variance :       610.84 (kcal/mol)**2
   Local hard. skewness :         0.10
   Local hard. kurtosis :        -0.25
   Integral local Hard. :   3515.06     eV.Angstrom**2
```

ParaSurf™ output for trimethoprim using a marching-cube surface.

```
     Maximum alpha(l)      :    1.563     Angstrom**3
     Minimum alpha(l)      :   0.2319     Angstrom**3
     Mean alpha(l)         :   0.6118     Angstrom**3
     Alpha(l) range        :    1.331     Angstrom**3
     Variance in alpha(l)  :   0.8362E-01 Angstrom**6
     Alpha(l) skewness     :       0.7110
     Alpha(l) kurtosis     :      -0.6335
     Integral Alpha(l)     :   192.581     Angstrom**5

     Maximum field normal  :    62.75 kcal/mol.Angstrom
     Minimum field normal  :   -96.37 kcal/mol.Angstrom
     Mean field            :    -0.24 kcal/mol.Angstrom
     Field range           :   159.12 kcal/mol.Angstrom
     Total field variance  :   526.01 (kcal/mol.Angstrom)**2
     +ve field variance    :    70.61 (kcal/mol.Angstrom)**2
     -ve field variance    :   604.96 (kcal/mol.Angstrom)**2
     Field balance param.  :     0.09
     Field skew            :     2.30
     Field kurtosis        :   3.507
     Integral F(N)         :   14.38      kcal.Angstrom/mol
     Integral F(N +ve)     :   2555.      kcal.Angstrom/mol
     Integral F(N -ve)     :  -2541.      kcal.Angstrom/mol
     Integral |F(N)|       :   5096.      kcal.Angstrom/mol

  <> Atomic surface properties:


   Atom   Area       MEP           IE(l)          EA(l)        mean     Field(N)       Eneg(L)         Hard(L)
                  max    min     max    min     max    min    pol.    max    min     max    min     max    min
 C    1  3.633  14.36 -27.83  492.02 425.67  -81.81  -98.38  0.729  11.25 -21.64  201.75 164.72  290.27 260.96
 O    2  5.872 -16.60 -61.42  517.24 376.50  -65.77  -85.49  1.083  -1.05 -73.60  224.26 154.07  292.99 222.43
 C    3  6.547   2.50 -44.56  505.88 359.14  -22.48  -87.04  0.908   9.75 -28.36  213.01 152.78  292.88 192.76
 C    4  5.042  11.25 -15.57  482.09 357.89  -27.99 -103.87  0.839  -1.35 -18.32  197.72 149.42  286.32 193.04
 C    5  2.637   0.36 -15.56  478.00 380.11  -39.08  -99.66  0.854   0.21 -18.11  196.09 156.43  281.92 209.82
 C    6  0.736   2.62  -6.47  484.34 448.62  -90.49 -102.29  0.787   1.54 -11.13  194.67 174.65  289.95 273.97
 C    7  3.866  -1.32 -16.72  471.01 352.79  -29.26  -97.15  0.914   8.05 -20.00  190.22 145.74  281.42 197.86
 C    8  7.747   8.21 -30.02  485.19 370.16  -27.37 -100.39  0.838   9.32 -27.15  195.38 153.82  289.81 201.38
 N    9  8.821  -6.66 -61.59  513.72 315.67  -68.02 -106.03  1.220  -1.63 -93.58  211.59 108.12  302.12 205.46
 C   10  9.463  10.11 -34.05  533.20 394.25  -35.99  -99.27  0.895  16.99 -30.32  223.14 168.66  310.05 215.12
 N   11  3.906  18.92 -40.19  525.07 373.54  -56.93  -87.16  0.924  12.75 -60.56  227.36 149.70  297.71 223.84
 N   12  8.304  -7.20 -58.48  498.61 312.99  -67.59 -105.70  1.114  -3.98 -96.37  204.03 109.56  294.58 199.79
 C   13  7.528   8.25 -31.03  528.52 376.72  -29.28 -100.87  0.883  15.91 -29.86  224.59 163.15  306.16 204.32
 N   14  3.839  16.28 -39.24  517.43 374.79  -56.71  -89.93  0.939  14.21 -65.25  223.29 150.78  294.14 223.72
 C   15  6.243   2.69 -15.27  471.78 355.95  -27.14  -91.73  0.831   3.88 -19.40  190.59 158.13  281.19 192.04
 C   16  6.375  -1.57 -42.44  496.09 359.34  -19.89  -90.31  0.891   7.12 -26.86  208.41 158.55  288.34 191.56
 O   17  4.315 -12.61 -52.21  517.66 379.79  -66.11  -94.14  0.982  -3.07 -65.91  222.62 148.11  295.03 223.59
 C   18  3.756  13.44 -26.67  492.59 427.45  -76.64  -97.98  0.764  10.83 -30.90  205.62 164.89  290.68 261.83
 C   19  5.797  -4.74 -53.09  525.33 337.63  -23.77  -96.21  0.962   0.65 -42.42  222.15 142.33  303.17 193.97
 O   20  5.835 -37.88 -64.18  524.48 358.58  -72.15  -95.74  1.098   9.43 -73.93  222.32 136.94  302.16 221.62
 C   21  4.118   4.81 -44.55  495.26 412.37  -90.33 -111.11  0.769  12.48 -34.16  202.46 151.69  292.79 259.79
 H   22 13.548  13.64 -26.76  467.39 383.82  -84.51  -97.89  0.437  17.14  -5.25  189.48 145.14  277.91 238.67
 H   23 11.587  18.01  -5.94  480.28 382.02  -81.52  -95.71  0.409  10.75  -7.60  199.38 145.39  281.10 236.63
 H   24 11.815  16.82 -11.32  468.75 382.79  -75.15  -95.47  0.409  14.32  -3.06  193.24 145.38  278.42 237.41
 H   25  6.767  18.89   0.75  451.49 404.00  -79.12 -101.07  0.393  15.88  -5.11  179.17 153.86  274.47 241.64
 H   26 10.613  14.79  -4.33  469.02 375.28  -92.05 -102.60  0.418  11.05 -10.08  186.54 139.27  282.48 236.00
 H   27  9.786  13.28 -16.03  476.30 373.94  -90.75 -102.75  0.424  16.09 -10.21  192.20 137.74  284.11 236.20
 H   28 11.191  15.84 -24.19  453.71 393.19  -60.41 -100.67  0.431  19.13 -13.24  176.52 150.21  277.19 230.01
 H   29 12.557  28.32 -32.30  537.65 465.98  -73.96 -111.21  0.370  58.46 -39.59  231.55 177.41  312.13 278.59
 H   30 12.354  28.35 -28.15  538.82 465.58  -77.89 -111.25  0.349  62.75 -29.94  230.46 177.29  311.69 274.57
 H   31 12.658  26.33 -29.11  539.60 461.88  -75.50 -112.06  0.364  50.74 -27.22  230.86 174.94  309.67 278.15
 H   32  7.359  26.07 -11.11  539.23 414.89  -75.34 -106.39  0.394  32.98 -26.16  230.48 157.36  308.74 257.53
 H   33  6.884  15.97  -8.00  446.99 389.30  -77.35  -99.71  0.450  14.37 -13.14  175.53 148.86  271.89 237.01
 H   34 13.443  13.65 -24.83  466.77 384.57  -85.30  -96.78  0.424  15.59 -26.06  189.59 145.87  277.17 238.69
 H   35 11.745  15.14  -8.63  467.45 382.69  -72.65  -95.71  0.412  13.76  -0.45  189.59 145.33  277.86 237.36
 H   36 11.570  14.77  -5.76  474.04 381.97  -77.50  -96.16  0.421  12.80  -1.84  198.27 145.06  278.15 236.91
 H   37 13.434   5.48 -36.63  446.92 370.92  -97.15 -109.63  0.420  10.58 -16.68  174.74 133.24  273.45 237.64
 H   38 11.343   4.12 -36.90  453.74 365.57  -92.78 -109.05  0.439  12.36 -32.90  175.88 129.49  277.86 236.09
 H   39 13.105   4.36 -33.58  487.05 365.53  -85.22 -108.13  0.427  25.62 -10.59  200.91 130.38  286.14 235.15

  Total   316.139

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

          x           y           z       value
```

ParaSurf™ output for trimethoprim using a marching-cube surface.

```
<>     13 MEP Maxima      :
     4.4569      0.9397     -3.2568     13.65
    -0.9431      3.5397     -2.9285     15.84
    -5.1681     -4.1904     -2.2285     13.64
     1.6569      1.8397     -2.5118     15.97
    -5.2189     -3.5960     -2.0285     13.79
    -4.3431     -1.6853     -1.1618     18.89
     3.6569      0.7480     -0.8285     10.06
    -3.3931     -0.3603     -0.2618     12.48
    -0.7431     -0.4686      0.3715     11.75
    -1.9181     -1.4603      1.3548     26.07
    -1.9431     -1.4103      1.6933     26.33
     1.2069      6.1397      2.4298     21.06
     1.2569      4.7204      3.8715     28.35
<>      5 MEP Minima      :
     0.4819     -3.8270     -4.1785    -60.72
     2.2944     -2.2603     -3.5285    -52.21
    -0.4931     -4.2603     -3.2436    -64.18
     0.2569      5.3397      0.0840    -61.59
    -1.2431      1.7397      3.3715    -58.48

<>     10 IEL Maxima      :
     1.6861     -2.6103     -3.4142     525.3
    -0.5431     -4.2436     -2.2285     524.5
     1.1569      4.5115      1.3715     519.8
    -1.8214     -1.4103      1.5715     526.5
    -2.7431     -1.2436      1.5715     539.6
    -0.6431      5.3397      1.5441     533.2
    -2.4431      1.3215      2.5715     533.5
     0.4569      2.5397      3.0715     521.6
    -1.2931      3.3564      3.4048     533.1
     0.6726      5.1397      4.0715     538.8
<>      3 IEL Minima      :
     0.0569      5.4397      0.1848     315.7
    -1.4431      1.9137      3.3715     313.4
    -0.9514      1.6746      3.3715     313.0

<>      6 EAL Maxima      :
    -1.4714     -2.0603     -3.9285    -22.48
     0.0069     -1.2520     -4.1285    -19.89
    -2.5431     -0.8270     -3.5285    -27.99
     0.8569     -1.1103     -0.6285    -25.12
    -2.4931      3.9397     -0.7067    -27.37
    -0.7431     -2.4270     -0.4785    -28.56
<>     16 EAL Minima      :
     1.4619     -4.8103     -3.9285    -107.3
     1.8319     -4.4936     -3.9285    -107.7
     2.1569     -4.2603     -3.8285    -108.6
     2.6569     -3.7186     -3.4285    -109.8
     0.2569     -5.4603     -2.4285    -111.1
     3.4569     -2.7603     -1.3952    -106.6
     1.5286     -2.8603     -0.5285    -107.4
    -2.5931     -1.5603      0.2298    -105.9
    -1.0431     -1.0603     -0.0035    -106.4
     0.9569      3.1397      0.2798    -106.0
    -2.1431     -1.6436      0.3715    -106.0
     0.2569      5.3397      0.8215    -105.6
     0.4402      1.7397      1.3465    -105.7
     1.2569      5.9397      1.9557    -111.2
    -1.7598     -0.3603      3.3715    -112.1
     0.2569      3.1397      4.6715    -111.3

<>      3 Alpha(l) Maxima :
    -0.7431     -3.7603     -3.4785     1.339
    -0.8648     -3.8603     -3.2785     1.370
     0.3370      5.3397      0.4215     1.563
<>     54 Alpha(l) Minima :
     2.0569      0.1397     -4.9285    0.3178
     2.2569      0.2447     -4.9785    0.3160
    -3.8514     -3.4603     -4.3285    0.3187
```

ParaSurf™ output for trimethoprim using a marching-cube surface.

```
   -3.3981   -3.2603   -4.3285  0.3178
    4.1569   -0.6936   -4.4285  0.3231
    2.1402   -5.0603   -3.9285  0.3183
    4.8069    0.5397   -3.4285  0.3797
   -2.3931    1.5347   -3.7285  0.3309
   -2.7431    1.7647   -3.6118  0.3267
   -1.1431    2.0480   -3.5285  0.3534
    1.4319    1.7230   -3.3918  0.3390
   -3.0955   -5.4603   -3.3785  0.3368
    3.0569   -5.0603   -3.3285  0.3313
   -5.1431   -2.7603   -3.2152  0.3362
   -5.0431   -2.4103   -2.9952  0.3331
   -3.1264    2.1647   -3.2285  0.3250
   -2.7431   -5.6436   -2.8535  0.3358
    3.4569   -3.3103   -2.6285  0.3429
    4.9569    0.1623   -2.5785  0.3595
   -0.2598    3.1397   -2.7285  0.3384
   -2.7223   -5.7603   -2.2545  0.3337
    3.0569   -5.7603   -2.2285  0.3457
   -4.1431   -1.0103   -2.5285  0.3418
    4.6569   -0.9695   -2.2285  0.3184
    0.0902    3.3397   -2.5785  0.3334
   -1.8431    4.3454   -2.2285  0.3198
    2.0652   -6.2603   -1.8285  0.3643
   -4.1431    1.5573   -2.1285  0.3377
    2.1069    1.7397   -1.6285  0.3607
   -3.9064   -5.5603   -1.2285  0.3498
   -0.3431   -4.9603   -1.1118  0.3573
   -4.8931   -2.4603   -1.0285  0.3431
    1.6569    1.1897   -1.0285  0.3529
   -0.3264   -4.2153   -0.6285  0.3554
    3.1361   -3.5493   -0.6285  0.3377
    3.1319   -0.0603   -0.6285  0.3092
    1.3402   -5.6603   -0.4285  0.3542
   -3.3931   -0.4603   -0.0335  0.2865
   -3.9431    1.5340   -0.3285  0.3554
    1.6402   -4.4353    0.1715  0.3239
   -3.5931   -3.8603    0.0715  0.3174
   -3.2931   -3.6603    0.0715  0.3131
   -1.3107   -1.3603    0.5257  0.2593
   -0.7431   -0.4686    0.3715  0.3961
    1.7569    4.9397    2.0882  0.2451
    0.0569    6.0897    1.9382  0.2527
   -0.8431   -0.6603    2.4215  0.2322
    0.2569    6.2397    2.9715  0.2828
   -2.6431    0.3590    3.1715  0.2427
    1.6569    4.8680    3.2590  0.2709
    1.5569    5.1897    3.3715  0.2611
    1.2569    3.2397    3.7998  0.2319
   -0.7431    4.0397    4.5715  0.2467
    0.4069    4.5397    4.6715  0.2668

 <>     28 F(N)  Maxima     :
    1.1569   -0.0803   -4.2285   13.38
   -2.7431   -2.6803   -3.9285   13.90
    4.9569   -0.4404   -3.8035   13.94
   -2.2181    2.5397   -3.6285   10.90
   -3.0880   -5.7603   -3.0285   13.71
    2.4902   -2.8103   -3.1285   10.38
    3.9402   -1.7603   -3.2535   14.07
   -4.3431   -5.8436   -2.6285   13.91
   -3.6848   -5.9603   -2.8785   13.91
   -2.6431   -5.6603   -2.6118   14.22
   -0.9431    4.5397   -2.6285   18.65
    1.2569    2.0897   -1.7902   14.21
   -0.4431   -4.3770   -1.6285   22.72
   -0.4931   -4.1653   -1.4285   23.01
   -4.2931   -1.0603   -1.0285   15.70
    2.0569    0.5730   -0.6785   12.24
   -4.7431   -2.8445   -0.3285   10.19
   -3.6431    0.8897   -0.0285   13.70
```

ParaSurf™ output for trimethoprim using a marching-cube surface.

```
     -1.6931     -1.5603      0.6498     31.72
      0.4569      2.3397      0.6215     11.69
     -2.1264      3.5397      1.1715     15.99
      0.8152      5.2897      1.2715     44.97
      1.6069      5.7397      2.3715     34.29
     -1.7014     -1.1603      2.7715     31.95
     -1.2931      0.5397      3.0715     40.69
     -0.3681      2.5564      3.5715     43.03
      0.3402      3.7397      4.8715     32.81
      0.6819      3.7197      4.7548     33.42
 <>      16 F(N) Minima       :
     -0.0157     -4.2603     -3.8285    -72.08
      0.7684     -3.6603     -4.2118    -72.28
      2.5018     -2.2603     -3.4285    -64.88
     -2.4264     -0.6603     -3.5285    -15.74
     -0.5931      0.5397     -3.7285    -15.80
      3.0402     -2.0853     -2.2285    -58.31
      2.8569     -2.0153     -1.8285    -58.39
     -1.5256     -4.6103     -1.4285    -69.79
     -1.5431     -2.0103     -0.2285    -16.73
     -0.0014     -0.0603     -0.2785    -18.93
      0.3569      1.5897     -0.1952    -16.03
     -3.0931      2.7397     -0.1702    -19.34
      0.2569      5.3397      0.0840    -92.35
     -3.5431      0.8373      1.2307    -64.12
     -0.6931      1.7397      3.3715    -94.22
     -1.3431      4.9397      3.1840    -59.11

 <> ParaSurf used         6.02 seconds CPU time
```

Figure 15 ParaSurf™ output for trimethoprim using a marching-cube surface.

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 0.007 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.

3.5.3 For a job with Shannon entropy

Figure 16 and **Figure 17** 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 statistical background file from the PARASURF_ROOT directory. The output is identical to that shown in **Figure 15** except that an additional Shannon entropy block is printed after the descriptors, as shown in **Figure 16**:

```
                                internal     external
     Maximum Shannon H      :     0.4229       0.3305     bits Angstrom**-2
     Minimum Shannon H      :     0.0464       0.1296     bits Angstrom**-2
     Mean Shannon H         :     0.2126       0.2186     bits Angstrom**-2
     Variance Shannon H     :     0.0073       0.0013     bits Angstrom**-2
     Molecular Shannon H    :      68.76        69.98     bits
```

Figure 16 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. “Internal” Shannon entropy is calculated using the statistical distribution of the local properties on the surrface of the molecule itself as reference, whereas “external” Shannon entropy uses pre-calculated background statistics from a database of drug-like ligands. [43]

Note that external statistics files are only provided for the AM1 Hamiltonian.

The Shannon entropy is also analyzed based on the surfaces assigned to the individual atoms to give the table shown in **Figure 17**:

Shannon entropy section of the ParaSurf™ output for trimethoprim, 1, using a marching-cube isodensity surface

```
   Shannon-entropy analysis :


                                         Shannon Entropy
                            Internal                        External
  Atom    Area     max     min    mean   total     max     min    mean   total
C     1   3.633  0.2135  0.1191  0.1784  0.6481  0.2589  0.1902  0.2189  0.7952
O     2   5.872  0.1466  0.0545  0.1120  0.6577  0.2334  0.1350  0.1985  1.1656
C     3   6.547  0.1954  0.0744  0.1383  0.9054  0.2881  0.1481  0.2100  1.3751
C     4   5.042  0.2322  0.0982  0.1639  0.8263  0.3080  0.1634  0.2373  1.1966
C     5   2.637  0.2285  0.1216  0.1634  0.4307  0.3065  0.1978  0.2332  0.6148
C     6   0.736  0.2277  0.1525  0.1854  0.1365  0.2571  0.2264  0.2415  0.1778
C     7   3.866  0.2273  0.0856  0.1503  0.5810  0.3133  0.1582  0.2203  0.8519
C     8   7.747  0.2071  0.1233  0.1623  1.2570  0.3041  0.1672  0.2325  1.8010
N     9   8.821  0.2147  0.0555  0.1166  1.0287  0.2639  0.1375  0.1943  1.7138
C    10   9.463  0.1988  0.0877  0.1453  1.3751  0.2781  0.1468  0.2087  1.9747
N    11   3.906  0.1953  0.1013  0.1297  0.5068  0.2528  0.1556  0.2101  0.8209
N    12   8.304  0.1955  0.0595  0.1193  0.9905  0.2588  0.1461  0.1960  1.6276
C    13   7.528  0.1900  0.0922  0.1460  1.0993  0.2746  0.1488  0.2095  1.5768
N    14   3.839  0.1935  0.0971  0.1296  0.4974  0.2418  0.1475  0.2131  0.8181
C    15   6.243  0.2357  0.1006  0.1504  0.9391  0.3079  0.1690  0.2229  1.3916
C    16   6.375  0.2046  0.0788  0.1425  0.9084  0.2821  0.1543  0.2052  1.3081
O    17   4.315  0.2239  0.0690  0.1249  0.5389  0.2435  0.1566  0.2072  0.8941
C    18   3.756  0.2255  0.1233  0.1845  0.6931  0.2578  0.1917  0.2224  0.8355
C    19   5.797  0.2178  0.0676  0.1214  0.7038  0.2909  0.1307  0.1970  1.1422
O    20   5.835  0.1675  0.0587  0.1190  0.6945  0.2434  0.1609  0.2143  1.2503
C    21   4.118  0.2069  0.1260  0.1528  0.6292  0.2498  0.1553  0.1911  0.7870
H    22  13.548  0.3922  0.1129  0.2909  3.9414  0.2870  0.1983  0.2393  3.2424
H    23  11.587  0.3890  0.1496  0.3034  3.5158  0.2795  0.1795  0.2357  2.7315
H    24  11.815  0.4071  0.1264  0.3175  3.7516  0.3038  0.1826  0.2457  2.9026
H    25   6.767  0.3387  0.1741  0.2669  1.8066  0.3114  0.1730  0.2048  1.3856
H    26  10.613  0.3405  0.1561  0.2671  2.8346  0.2673  0.2038  0.2373  2.5183
H    27   9.786  0.3281  0.1500  0.2615  2.5588  0.2760  0.1891  0.2388  2.3368
H    28  11.191  0.4103  0.1267  0.3037  3.3982  0.3079  0.1736  0.2416  2.7035
H    29  12.557  0.2481  0.1108  0.1692  2.1243  0.2587  0.1401  0.1736  2.1795
H    30  12.354  0.2736  0.1009  0.1649  2.0367  0.2564  0.1392  0.1724  2.1301
H    31  12.658  0.2601  0.0908  0.1631  2.0650  0.2512  0.1390  0.1741  2.2040
H    32   7.359  0.2899  0.0994  0.1792  1.3188  0.2651  0.1393  0.1795  1.3214
H    33   6.884  0.3384  0.1664  0.2591  1.7834  0.3096  0.1834  0.2267  1.5607
H    34  13.443  0.4137  0.1221  0.3069  4.1249  0.2824  0.1974  0.2386  3.2078
H    35  11.745  0.4136  0.1311  0.3171  3.7246  0.3046  0.1880  0.2443  2.8692
H    36  11.570  0.3938  0.1278  0.3115  3.6039  0.3041  0.1868  0.2443  2.8270
H    37  13.434  0.3376  0.1440  0.2558  3.4362  0.2668  0.1668  0.2349  3.1552
H    38  11.343  0.3255  0.1204  0.2391  2.7124  0.2671  0.1680  0.2312  2.6222
H    39  13.105  0.3345  0.1122  0.2579  3.3800  0.2850  0.1639  0.2382  3.1220
```

Figure 17 Shannon entropy section of the ParaSurf™ output for trimethoprim, **1**, using a marching-cube isodensity surface.

3.5.4 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
```

Then 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 (Figure 18) from ParaSurf™:

Similarity output using autocorrelation functions

```
  <> Surface Autocorrelation vectors written to the SD-File

<> Calculating autocorrelation similarities to captopril_p.sdf
<> Lead molecule = Captopril

<> Individual autocorrelation similarities;

   Shape     MEP(tot)  MEP(+-)   MEP(++)   MEP(--)   IE(l)     EA(l)    Alpha(l)
   0.7833    0.6309    0.7873    0.8122    0.5133    0.9720    0.3303    0.9634

<>                        Total autocorrelation fingerprint similarity = 0.9716
```

Figure 18 Similarity output using autocorrelation functions. The lead molecule is captopril, which is defined in captopril_p.sdf.

The “Total autocorrelation fingerprint similarity” refers to the shape, MEP(+-), MEP(++), MEP(--), IE(L), EA(L) and Alpha(L) autocorrelation functions (a total of 448 bins). It is, however dominated by IE(L) and EA(L) because their numerical values are far larger than the other autocorrelation functions.

3.6 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 centres 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 centre of the molecule used for the spherical-harmonic fit and the centre of gravity, respectively. These two centres are usually identical, but may be different if the centre 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:

`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:	$V_{\max}$, $V_{\min}$, $\bar{V}_+$, $\bar{V}_-$, $\bar{V}$, 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_{EA+}^2, σ_{EA-}^2, σ_{EA}^2, 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_χ^2, 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_η^2, γ_1^η, γ_2^η, $\int_\eta$ `("Hard(l)    ",5f10.2,2f10.4/12x,g12.6)`
Alpha(l):	$\alpha_L^{\max}$, $\alpha_L^{\min}$, $\overline{\alpha_L}$, $\Delta\alpha_L$, σ_α^2, γ_1^α, γ_2^α, $\int_\alpha$ `("Alpha(l)   ",5f10.2,2f10.4/12x,g12.6)`
F$_N$	$F_N^{\max}$, $F_N^{\min}$, $\overline{F_N}$, σ_F^2, σ_{F+}^2, σ_{F-}^2, ν_F, $\gamma_1^{F_N}$, $\gamma_2^{F_N}$, $\int_{F_N}$, $\int_{F_N}^+$, $\int_{F_N}^-$, $\int_{F_N}$ `("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 `i` 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 [36] is printed as a list of 54 floating point numbers (6g12.6). The first 41 are those defined in reference [36] and the last 13 are the square roots of the hybridization coefficients for the normal field from l=0-12.

3.6.1 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 number of autocorrelation points	`("ncorr   = ",i6)`
The lower end of the autocorrelation range	`("rmin    = ",f10.6)`
The bin size	`("dcorr   = ",f10.6)`

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

Table 7 Column headings and definitions for autocorrelation tables.

Column heading	Contents
R	Reference distance (R in **Equation (18)**)
shape	Shape autocorrelation
MEP(Tot)	Total MEP autocorrelation
MEP(+-)	MEP +/- autocorrelation
MEP(++)	MEP +/+ autocorrelation
MEP(--)	MEP -/- autocorrelation
IE(L)	IEL autocorrelation
EA(L)	EAL autocorrelation
Alpha(L)	Alpha(L) autocorrelation

The format of the columns is (`f8.2,2x,8g15.6`)

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 centre plus eight translated ones) for **`TRANSLATE`** and 16 for **`TRANSLATE2`**. The original centre 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)

3.7 The surface (.psf) file

The .psf file can be used to derive properties and descriptors from the ParaSurf™ results. It includes the coordinates and properties of the atoms, surface points and surface triangles in the following format. This format has been extended compared to that used by ParaSurf’11™.

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, total number (N_{models}) of surface-integral models (normal and binned)	(i6,1x,i5)

The three-letter codes for the individual models	N_{models}*(1x,a3)

One line per point with the local properties:

x-coordinate, y-coordinate, z-coordinate, MEP, IE_L, EA_L, α_L, $atom_L$, local value of each model	(3f10.5,3f8.2,f8.4,i6,N_{models}*(2x,g12.4))

(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)

3.8 Anonymous SD (.asd) files

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

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

And the blocks defined by the following tags:

```
<SPHERICAL_HARMONIC_SURFACE>
<SPHERICAL_HARMONIC_MEP>
<SPHERICAL_HARMONIC_IE(l)>
<SPHERICAL_HARMONIC_EA(l)>
<SPHERICAL_HARMONIC_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	$V_{\max}$, $V_{\min}$, $\bar{V}_+$, $\bar{V}_-$, $\bar{V}$, 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_{EA+}^2, σ_{EA-}^2, σ_{EA}^2, 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)	$\chi_L^{\max}$, $\chi_L^{\min}$, $\overline{\chi_L}$, $\Delta\chi_L$, $\sigma_\chi^2, \gamma_1^\chi, \gamma_2^\chi, \int_\chi$ `("Eneg(l)   ",5f10.2,2f10.4/12x,g12.6)`
Hard(l)	$\eta_L^{\max}$, $\eta_L^{\min}$, $\overline{\eta_L}$, $\Delta\eta_L$, $\sigma_\eta^2, \gamma_1^\eta, \gamma_2^\eta, \int_\eta$ `("Hard(l)   ",5f10.2,2f10.4/12x,g12.6)`
Alpha(l)	$\alpha_L^{\max}$, $\alpha_L^{\min}$, $\overline{\alpha_L}$, $\Delta\alpha_L$, $\sigma_\alpha^2, \gamma_1^\alpha, \gamma_2^\alpha, \int_\alpha$ `("Alpha(l)  ",5f10.2,2f10.4/12x,g12.6)`
F_N	$F_N^{\max}$, $F_N^{\min}$, ΔF_N , $\overline{F_N}$, σ_F^2 , σ_{F+}^2 , σ_{F-}^2 , 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)	$H_{in}^{\max}$, $H_{in}^{\min}$, $\overline{H_{in}}$, $\sigma_{H_{in}}^2$, $\int_{H_{in}}$ `("Shannon(i)  ",4f10.4,f10.2,f10.4)`
Shannon(e)	$H_{ex}^{\max}$, $H_{ex}^{\min}$, $\overline{H_{ex}}$, $\sigma_{H_{ex}}^2$, $\int_{H_{ex}}$ `("Shannon(e)  ",4f10.4,f10.2,f10.4)`

3.8.1 Optional blocks

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

`<SURFACE AUTOCORRELATION PARAMETERS>`

The number of autocorrelation points	`("ncorr   = ",i6)`
The lower end of the autocorrelation range	`("rmin    = ",f10.6)`
The bin size	`("dcorr   = ",f10.6)`

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

Table 8 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
MEP(Tot)	Total MEP autocorrelation
MEP(+-)	MEP +/- autocorrelation
MEP(++)	MEP +/+ autocorrelation
MEP(--)	MEP -/- autocorrelation
IE(L)	IEL autocorrelation
EA(L)	EAL autocorrelation
Alpha(L)	Alpha(L) autocorrelation

The format of the columns is (`f8.2,2x,8g15.6`)

3.9 Grid calculations with ParaSurf™

3.9.1 User-specified Grid

The command

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

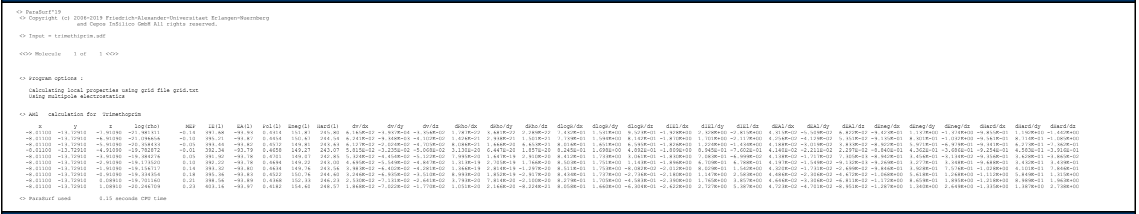
instructs ParaSurf™ to read a set of Cartesian coordinates from the file grid.dat and to calculate the $\log_{10}$ of the electron density, the four local properties, the electric field and the derivatives of the local properties (log(ρ), MEP (=V), IE(l), EA(l) Pol(l), Eneg(l), Hard(l), dV/dx, dV/dy, dV/dz, dρ/dx, dρ/dy, dρ/dz, dlog(ρ)/dx, dlog(ρ)/dy, dlog(ρ)/dz, dIE(l)/dx, dIE(l)/dy, dIE(l)/dz, dEA(l)/dx, dEA(l)/dy, dEA(l)/dz, dEneg/dx, dEneg/dy, dEneg/dz, dHard/dx, dHard/dy, dHard/dz). The format of the file grid.dat (which must be in the same directory as the input) is one line per point containing the x, y and z coordinates in free format, comma-separated, maximum line length 80 with no trailing comma. For instance, the following grid file (Figure 19):

```
-8.01100 , -13.72910 , -7.91090
-8.01100 , -13.72910 , -6.91090
-8.01100 , -13.72910 , -5.91090
-8.01100 , -13.72910 , -4.91090
-8.01100 , -13.72910 , -3.91090
-8.01100 , -13.72910 , -2.91090
-8.01100 , -13.72910 , -1.91090
-8.01100 , -13.72910 , -0.91090
-8.01100 , -13.72910 ,  0.08910
-8.01100 , -13.72910 ,  1.08910
```

Figure 19 Sample grid file

gives the output shown in **Figure 20**.

Sample grid output file (original)



```
<> ParaSurf'19
 <> Copyright (c) 2006-2019 Friedrich-Alexander-Universitaet Erlangen-Nuernberg
                  and Cepos InSilico GmbH All rights reserved.

 <> Input = trimethiprim.sdf


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



 <> Program options :

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


 <> AM1   calculation for  Trimethoprim

          x          y         z    log(rho)      MEP   IE(l)   EA(l)  Pol(l)  Eneg(l)  Hard(l)      dv/dx      dv/dy      dv/dz    dRho/dx    dRho/dy    dRho/dz   dlogR/dx   dlogR/dy   dlogR/dz    dIEl/dx    dIEl/dy    dIEl/dz    dEAl/dx    dEAl/dy    dEAl/dz   dEneg/dx   dEneg/dy   dEneg/dz   dHard/dx   dHard/dy   dHard/dz
   -8.01100  -13.72910  -7.91090  -21.981311    -0.14  397.68  -93.93  0.4314   151.87   245.80  6.165E-02 -3.937E-04 -3.356E-02  1.787E-22  3.681E-22  2.289E-22  7.432E-01  1.531E+00  9.523E-01 -1.928E+00  2.328E+00 -2.815E+00  4.315E-02 -5.509E-02  6.822E-02 -9.423E-01  1.137E+00 -1.374E+00 -9.855E-01  1.192E+00 -1.442E+00
   -8.01100  -13.72910  -6.91090  -21.096656    -0.10  395.21  -93.87  0.4454   150.67   244.54  6.241E-02 -9.348E-03 -4.102E-02  1.426E-21  2.938E-21  1.501E-21  7.739E-01  1.594E+00  8.142E-01 -1.870E+00  1.701E+00 -2.117E+00  4.256E-02 -4.129E-02  5.351E-02 -9.135E-01  8.301E-01 -1.032E+00 -9.561E-01  8.714E-01 -1.085E+00
   -8.01100  -13.72910  -5.91090  -20.358433    -0.05  393.44  -93.82  0.4572   149.81   243.63  6.127E-02 -2.024E-02 -4.705E-02  8.086E-21  1.666E-20  6.653E-21  8.016E-01  1.651E+00  6.595E-01 -1.826E+00  1.224E+00 -1.434E+00  4.188E-02 -3.019E-02  3.833E-02 -8.922E-01  5.971E-01 -6.979E-01 -9.341E-01  6.273E-01 -7.362E-01
   -8.01100  -13.72910  -4.91090  -19.782872    -0.01  392.34  -93.79  0.4658   149.27   243.07  5.815E-02 -3.235E-02 -5.068E-02  3.130E-20  6.447E-20  1.857E-20  8.245E-01  1.698E+00  4.892E-01 -1.809E+00  8.945E-01 -7.602E-01  4.140E-02 -2.211E-02  2.297E-02 -8.840E-01  4.362E-01 -3.686E-01 -9.254E-01  4.583E-01 -3.916E-01
   -8.01100  -13.72910  -3.91090  -19.384276     0.05  391.92  -93.78  0.4701   149.07   242.85  5.324E-02 -4.454E-02 -5.122E-02  7.995E-20  1.647E-19  2.910E-20  8.412E-01  1.733E+00  3.061E-01 -1.830E+00  7.083E-01 -6.999E-02  4.138E-02 -1.717E-02  7.305E-03 -8.942E-01  3.456E-01 -3.134E-02 -9.356E-01  3.628E-01 -3.865E-02
   -8.01100  -13.72910  -2.91090  -19.173520     0.10  392.22  -93.78  0.4694   149.22   243.00  4.695E-02 -5.549E-02 -4.847E-02  1.313E-19  2.705E-19  1.766E-20  8.503E-01  1.751E+00  1.143E-01 -1.896E+00  6.709E-01  6.788E-01  4.197E-02 -1.549E-02 -9.132E-03 -9.269E-01  3.277E-01  3.348E-01 -9.688E-01  3.432E-01  3.439E-01
   -8.01100  -13.72910  -1.91090  -19.156717     0.14  393.32  -93.80  0.4634   149.76   243.56  3.983E-02 -6.402E-02 -4.281E-02  1.366E-19  2.814E-19 -1.297E-20  8.511E-01  1.753E+00 -8.082E-02 -2.012E+00  8.029E-01  1.542E+00  4.320E-02 -1.731E-02 -2.699E-02 -9.846E-01  3.928E-01  7.576E-01 -1.028E+00  4.101E-01  7.846E-01
   -8.01100  -13.72910  -0.91090  -19.334354     0.18  395.36  -93.83  0.4522   150.76   244.60  3.246E-02 -6.935E-02 -3.510E-02  8.993E-20  1.852E-19 -2.917E-20  8.434E-01  1.737E+00 -2.736E-01 -2.180E+00  1.147E+00  2.583E+00  4.486E-02 -2.306E-02 -4.672E-02 -1.068E+00  5.618E-01  1.268E+00 -1.112E+00  5.849E-01  1.315E+00
   -8.01100  -13.72910   0.08910  -19.701160     0.21  398.56  -93.89  0.4368   152.33   246.23  2.530E-02 -7.131E-02 -2.641E-02  3.793E-20  7.814E-20 -2.100E-20  8.279E-01  1.705E+00 -4.583E-01 -2.390E+00  1.765E+00  3.857E+00  4.646E-02 -3.306E-02 -6.811E-02 -1.172E+00  8.659E-01  1.895E+00 -1.218E+00  8.989E-01  1.963E+00
   -8.01100  -13.72910   1.08910  -20.246709     0.23  403.16  -93.97  0.4182   154.60   248.57  1.868E-02 -7.022E-02 -1.770E-02  1.051E-20  2.166E-20 -8.224E-21  8.058E-01  1.660E+00 -6.304E-01 -2.622E+00  2.727E+00  5.387E+00  4.723E-02 -4.701E-02 -8.951E-02 -1.287E+00  1.340E+00  2.649E+00 -1.335E+00  1.387E+00  2.738E+00

 <> ParaSurf used         0.15 seconds CPU time
```

Sample grid output file (.txt)

```
<> ParaSurf'19
 <> Copyright (c) 2006-2019 Friedrich-Alexander-Universitaet Erlangen-Nuernberg
                  and Cepos InSilico GmbH All rights reserved.

 <> Input = trimethiprim.sdf


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



 <> Program options :
```

```
     Calculating local properties using grid file grid.txt
     Using multipole electrostatics


  <> AM1   calculation for  Trimethoprim

       x           y           z        log(rho)         MEP     IE(l)     EA(l)     Pol(l)
Eneg(l)  Hard(l)     dv/dx       dv/dy       dv/dz      dRho/dx     dRho/dy     dRho/dz
dlogR/dx   dlogR/dy   dlogR/dz    dIEl/dx    dIEl/dy    dIEl/dz    dEAl/dx    dEAl/dy
dEAl/dz   dEneg/dx   dEneg/dy   dEneg/dz   dHard/dx   dHard/dy   dHard/dz

   -8.01100   -13.72910    -7.91090   -21.981311       -0.14    397.68    -93.93    0.4314
151.87     245.80   6.165E-02  -3.937E-04  -3.356E-02   1.787E-22   3.681E-22   2.289E-22
7.432E-01  1.531E+00  9.523E-01 -1.928E+00  2.328E+00 -2.815E+00  4.315E-02 -5.509E-02
6.822E-02 -9.423E-01  1.137E+00 -1.374E+00 -9.855E-01  1.192E+00 -1.442E+00

   -8.01100   -13.72910    -6.91090   -21.096656       -0.10    395.21    -93.87    0.4454
150.67     244.54   6.241E-02  -9.348E-03  -4.102E-02   1.426E-21   2.938E-21   1.501E-21
7.739E-01  1.594E+00  8.142E-01 -1.870E+00  1.701E+00 -2.117E+00  4.256E-02 -4.129E-02
5.351E-02 -9.135E-01  8.301E-01 -1.032E+00 -9.561E-01  8.714E-01 -1.085E+00

   -8.01100   -13.72910    -5.91090   -20.358433       -0.05    393.44    -93.82    0.4572
149.81     243.63   6.127E-02  -2.024E-02  -4.705E-02   8.086E-21   1.666E-20   6.653E-21
8.016E-01  1.651E+00  6.595E-01 -1.826E+00  1.224E+00 -1.434E+00  4.188E-02 -3.019E-02
3.833E-02 -8.922E-01  5.971E-01 -6.979E-01 -9.341E-01  6.273E-01 -7.362E-01

   -8.01100   -13.72910    -4.91090   -19.782872       -0.01    392.34    -93.79    0.4658
149.27     243.07   5.815E-02  -3.235E-02  -5.068E-02   3.130E-20   6.447E-20   1.857E-20
8.245E-01  1.698E+00  4.892E-01 -1.809E+00  8.945E-01 -7.602E-01  4.140E-02 -2.211E-02
2.297E-02 -8.840E-01  4.362E-01 -3.686E-01 -9.254E-01  4.583E-01 -3.916E-01

   -8.01100   -13.72910    -3.91090   -19.384276        0.05    391.92    -93.78    0.4701
149.07     242.85   5.324E-02  -4.454E-02  -5.122E-02   7.995E-20   1.647E-19   2.910E-20
8.412E-01  1.733E+00  3.061E-01 -1.830E+00  7.083E-01 -6.999E-02  4.138E-02 -1.717E-02
7.305E-03 -8.942E-01  3.456E-01 -3.134E-02 -9.356E-01  3.628E-01 -3.865E-02

   -8.01100   -13.72910    -2.91090   -19.173520        0.10    392.22    -93.78    0.4694
149.22     243.00   4.695E-02  -5.549E-02  -4.847E-02   1.313E-19   2.705E-19   1.766E-20
8.503E-01  1.751E+00  1.143E-01 -1.896E+00  6.709E-01  6.788E-01  4.197E-02 -1.549E-02
-9.132E-03 -9.269E-01  3.277E-01  3.348E-01 -9.688E-01  3.432E-01  3.439E-01

   -8.01100   -13.72910    -1.91090   -19.156717        0.14    393.32    -93.80    0.4634
149.76     243.56   3.983E-02  -6.402E-02  -4.281E-02   1.366E-19   2.814E-19  -1.297E-20
8.511E-01  1.753E+00 -8.082E-02 -2.012E+00  8.029E-01  1.542E+00  4.320E-02 -1.731E-02
-2.699E-02 -9.846E-01  3.928E-01  7.576E-01 -1.028E+00  4.101E-01  7.846E-01

   -8.01100   -13.72910    -0.91090   -19.334354        0.18    395.36    -93.83    0.4522
150.76     244.60   3.246E-02  -6.935E-02  -3.510E-02   8.993E-20   1.852E-19  -2.917E-20
8.434E-01  1.737E+00 -2.736E-01 -2.180E+00  1.147E+00  2.583E+00  4.486E-02 -2.306E-02
-4.672E-02 -1.068E+00  5.618E-01  1.268E+00 -1.112E+00  5.849E-01  1.315E+00


   -8.01100   -13.72910     0.08910   -19.701160        0.21    398.56    -93.89    0.4368
152.33     246.23   2.530E-02  -7.131E-02  -2.641E-02   3.793E-20   7.814E-20  -2.100E-20
8.279E-01  1.705E+00 -4.583E-01 -2.390E+00  1.765E+00  3.857E+00  4.646E-02 -3.306E-02
-6.811E-02 -1.172E+00  8.659E-01  1.895E+00 -1.218E+00  8.989E-01  1.963E+00

   -8.01100   -13.72910     1.08910   -20.246709        0.23    403.16    -93.97    0.4182
154.60     248.57   1.868E-02  -7.022E-02  -1.770E-02   1.051E-20   2.166E-20  -8.224E-21
8.058E-01  1.660E+00 -6.304E-01 -2.622E+00  2.727E+00  5.387E+00  4.723E-02 -4.701E-02
-8.951E-02 -1.287E+00  1.340E+00  2.649E+00 -1.335E+00  1.387E+00  2.738E+00


  <> ParaSurf used         0.15 seconds CPU time
```

Figure 20 Sample grid output file

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.

3.9.2 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 0.5 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 a predefined grid and can be used for other molecules that have been aligned with the lead. An additional output file named <filename>_p.grid. There are two further variations of the automatic grid-generation procedure: **`grid=auto`** excludes any points that are within 0.5 Å of a nucleus, whereas **`grid=vdw`** excludes all grid points within the van der Waals volume of the molecule and **`grid=box`** calculates all points regardless of their proximity to a nucleus.

3.10 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 octanol) is shown in **Figure 21**:

```
> <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 21 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.

3.11 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 9 Definitions and order of the descriptors printed to the descriptor table if requested.

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

Column Header	Symbol	Descriptor
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+	$\delta\mathrm{A}_{EA}^{+}$	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
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

Column Header	Symbol	Descriptor
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
FNvar-	σ_{F-}^2	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:

Table 10 Additional descriptors (Shannon entropy)

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

3.12 Binned SIM descriptor tables

If the option “`desfile=<filename>`” is used, a user-defined file with binned SIM-descriptors is written. The bin limits are taken from installed models using the command-line option “`desmodel=<code>`”, where `<code>` is the model code taken from **Fehler! Verweisquelle konnte nicht gefunden werden.**. If the table file does not exist, a new one with a header line will be written, otherwise the results for the current molecule will be appended. The descriptors are denoted by a two- or three-letter code to denote the property followed by the number of the descriptor (currently 1-12). The letter codes are:

Table 11 Letter codes (Binned SIM descriptor tables)

Code	Property	Code	Property	Code	Property
MEP	MEP	IEL	IE_L	EAL	EA_L
POL	α_L	FN	F_N	HD	η_L
ENG	χ_L	MI	$\mathrm{MEP}\times\mathrm{IE}_L$	MA	$\mathrm{MEP}\times\mathrm{EA}_L$
MP	$\mathrm{MEP}\times\alpha_L$	MF	$\mathrm{MEP}\times F_N$	MH	$\mathrm{MEP}\times\eta_L$
ME	$\mathrm{MEP}\times\chi_L$	IA	$\mathrm{IE}_L\times\mathrm{EA}_L$	IP	$\mathrm{IE}_L\times\alpha_L$
IF	$\mathrm{IE}_L\times F_N$	IH	$\mathrm{IE}_L\times\eta_L$	IE	$\mathrm{IE}_L\times\chi_L$
AP	$\mathrm{EA}_L\times\alpha_L$	AF	$\mathrm{EA}_L\times F_N$	AH	$\mathrm{EA}_L\times\eta_L$
AE	$\mathrm{EA}_L\times\chi_L$	PF	$\alpha_L\times F_N$	PH	$\alpha_L\times\eta_L$
PE	$\alpha_L\times\chi_L$	FH	$F_N\times\eta_L$	FE	$F_N\times\chi_L$
HE	$\eta_L\times\chi_L$				

The descriptor file is comma-separated.

3.13 Autocorrelation fingerprint and similarity tables

If the option “`aclist=<filename>`” is used, a user-defined file with the autocorrelation fingerprint 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 and contains the molecular identifier followed by 448 binned autocorrelation values in the order Shape, MEP(+-), MEP(++), MEP(--), IE(L), EA(L), Alpha(L) (64 bins each).

The option “`aclist=<filename>`” requests a user-defined file with the autocorrelation similarities to the lead compound defined using the “`autocorr = <filename>`” keyword. If this file does not exist, it is created and the header line written, otherwise entries are appended. The ASCII file is has a fixed format. The header line is

```
Molid                         Shape       MEP(tot)  MEP(+-)   MEP(++)   MEP(-
-)    IE(l)      EA(l)     Alpha(l)    Fingerprint
```

and the similarities are written in format (a20,9f10.4). If the molecular identifier is longer than 20 characters, it will be truncated. The “Fingerprint” similarity considers all 448 autocorrelation values (but see Section **1.15**.

3.14 Shared files

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

4 TIPS FOR USING PARASURF19™

4.1 Choice of surface

ParaSurf™ was originally written to use isodensity surfaces. However, calculations that use a solvent-excluded 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-excluded 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 Local Properties

The improved local properties implemented in ParaSurf'12™ generally give better QSAR and QSPR models than the earlier ones available up to ParaSUrf'11™. It is therefore recommended that new projects use the ParaSurf'12™ local properties.

4.3 QSAR using grids

As outlined in Section **3.9.2**, 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 Section **1.1**

The automatic grid generated for a lead molecule is now written to the file `<filename>_p.grid` for use with the remainder of the dataset.

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 GmbH

Waldstraße 25
90587 Obermichelbach
Germany

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

www.ceposinsilico.com/contact

6 LIST OF TABLES

7 LIST OF FIGURES

8 REFERENCES

[1] T. Clark, A. Alex, B. Beck, J. Chandrasekhar, P. Gedeck, A. H. C. Horn, M. Hutter, B. Martin, G. Rauhut, W. Sauer, T. Schindler, T. Steinke, VAMP 8.2, **2002**, Erlangen. Available from Accelrys Inc., San Diego, USA (**https://www.3dsbiovia.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] a) J. H. Van Drie. 'Shrink-wrap' surfaces: a new method for incorporating shape into pharmacophoric 3D database searching*, Journal of Chemical Information & Computer Sciences* **1997**, *37*, 38-42; b) J. H. Van Drie, 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; c) J. Erickson, D. J. Neidhart, J. H. 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, J. Brickmann. Fast generation of molecular surfaces from 3D data fields with an enhanced "marching cube" algorithm*, J. Comput. Chem.* **1993**, *14*, 246-250

[5] D. W. Ritchie, 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*, P. Politzer, D. G. Truhlar (Eds.), Plenum Press, New York, NY, **1981**.

[7] P. Sjoberg, J. S. Murray, T. Brinck, P. Politzer. Average local ionisation energies on the molecular surfaces of aromatic systems as guides to chemical reactivity*, Can. J. Chem.* **1990**, *68*, 1440-1443

[8] B. Ehresmann, B. Martin, A. H. Horn, T. Clark. Local molecular properties and their use in predicting reactivity*, Journal of Molecular Modeling* **2003**, *9*, 342-347. doi:10.1007/s00894-003-0153-x

[9] B. Ehresmann, M. J. de Groot, A. Alex, T. Clark. New molecular descriptors based on local properties at the molecular surface and a boiling-point model derived from them*, J Chem Inf Comput Sci* **2004**, *44*, 658-668. doi:10.1021/ci034215e

[10] B. Ehresmann, M. J. de Groot, T. Clark. Surface-integral QSPR models: local energy properties*, J Chem Inf Model* **2005**, *45*, 1053-1060. doi:10.1021/ci050025n

[11] L. M. Loew, 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*, 1019-1025

[12] B. S. Duncan, A. J. Olson. Approximation and visualization of large-scale motion of protein surfaces*, J Mol Graph* **1995**, *13*, 250-257

[13] J. H. Lin, T. Clark. An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties*, J Chem Inf Model* **2005**, *45*, 1010-1016. doi:10.1021/ci050059v

[14] a) G. Rauhut, T. Clark. Multicenter Point Charge Model for High Quality Molecular Electrostatic Potentials from AM1 Calculations*, J. Comput. Chem.* **1993**, *14*, 503-509; b) B. Beck, G. Rauhut, T. Clark. The Natural Atomic Orbital Point Charge Model for PM3: Multipole Moments and Molecular Electrostatic Potentials*, J. Comput. Chem.* **1994**, *15*, 1064-1073

[15] a) M. J. S. Dewar, W. Thiel. MNDO*, J. Am. Chem. Soc.* **1977**, *99*, 4899-4907; 4907-4917; b) W. Thiel. MNDO, in *Encyclopedia of Computational Chemistry, Vol. 1*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, P. R. Schreiner (Eds.), Wiley, Chichester, **1998**, p. 1599

[16] a) M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, 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. doi:10.1021/ja00299a024; b) A. J. Holder. AM1, in *Encyclopedia of Computational Chemistry, Vol. 1*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, P. R. Schreiner (Eds.), Wiley, Chichester, **1998**, pp. 8-11

[17] a) J. J. P. Stewart. Optimization of parameters for semiempirical methods I. Method*, Journal of Computational Chemistry* **1989**, *10*, 209-220. doi:10.1002/jcc.540100208; b) J. J. P. Stewart. Optimization of parameters for semiempirical methods II. Applications*, Journal of Computational Chemistry* **1989**, *10*, 221-264. doi:10.1002/jcc.540100209; c) J. J. Stewart. PM3, in *Encyclopedia of*

Computational Chemistry, Vol. 1, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, P. R. Schreiner (Eds.), Wiley, Chichester, **1998**, pp. 2080-2086

[18] P. Winget, A. C. Horn, C. Selçuki, B. Martin, T. Clark. AM1* parameters for phosphorus, sulfur and chlorine, *Journal of Molecular Modeling* **2003**, *9*, 408-414. doi:10.1007/s00894-003-0156-7

[19] a) W. Thiel, A. A. Voityuk. Extension of the MNDO formalism to d orbitals: integral approximations and preliminary numerical results, *Theor. Chim. Acta* **1992**, *81*, 391-404; b) W. Thiel, A. A. Voityuk. Extension of MNDO to d orbitals: parameters and results for the silicon, *J. Mol. Struct.* **1994**, *313*, 141-154; c) W. Thiel, A. A. Voityuk. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group, *The Journal of Physical Chemistry* **1996**, *100*, 616-626. doi:10.1021/jp952148o; d) W. Thiel, A. A. Voityuk. Extension of MNDO to d orbitals: parameters and results for the halogens, *Theor. Chim. Acta* **1996**, *93*, 315-315; e) W. Thiel. MNDO/d, in *Encyclopedia of Computational Chemistry, Vol. 3*, P. v. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, III, P. R. Schreiner (Eds.), Wiley, Chichester, **1998**, pp. 1604-1606

[20] a) B. Martin, M. Gedeck, T. Clark. An Additive NDDO-Based Atomic Polarizability Model, *Int. J. Quant. Chem.* **2000**, *77*, 473-497; b) T. Clark. The local electron affinity for non-minimal basis sets, *Journal of Molecular Modeling* **2010**, *16*, 1231-1238. doi:10.1007/s00894-009-0607-x

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

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

[23] a) D. Rinaldi, 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*, 57-57; b) D. Rinaldi, J.-L. Rivail. Calculation of molecular electronic polarizabilities. Comparison of different methods, *Theor. Chim. Acta* **1974**, *32*, 243-251; c) J.-L. Rivail, D. Rinaldi. Variational calculation of multipole electric polarizabilities, *ComptesRendus, Serie B: Sciences Physiques* **1976**, 283-283

[24] a) R. J. Abraham, B. D. Hudson, M. W. Kermode, J. R. Mines. A general calculation of molecular solvation energies, *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1988**, *84*, 1911-1917. doi:10.1039/f19888401911; b) J. Tomasi, B. Mennucci, R. Cammi. Quantum mechanical continuum solvation models, *Chem Rev* **2005**, *105*, 2999-3093. doi:10.1021/cr9904009

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

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

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

[28] CORINA, Molecular Networks GmbH, Erlangen. **https://www.mn-am.com/products/corina**

[29] LogKOW dataset. **http://www.tds-tds.com/c5/application/files/7314/7360/9080/LOGKOWFS2016e.pdf**

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

[31] a) R. Wang, X. Fang, Y. Lu, S. Wang. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures, *J Med Chem* **2004**, *47*, 2977-2980. doi:10.1021/jm030580l; b) R. Wang, X. Fang, Y. Lu, C. Y. Yang, S. Wang. The PDBbind database: methodologies and updates, *J Med Chem* **2005**, *48*, 4111-4119. doi:10.1021/jm048957q

[32] L. Mavridis, B. D. Hudson, D. W. Ritchie. Toward high throughput 3D virtual screening using spherical harmonic surface representations, *J Chem Inf Model* **2007**, *47*, 1787-1796. doi:10.1021/ci7001507

[33] A. J. Jakobi, H. Mauser, T. Clark. ParaFrag--an approach for surface-based similarity comparison of molecular fragments, *Journal of Molecular Modeling* **2008**, *14*, 547-558. doi:10.1007/s00894-008-0302-3

[34] a) M. Hennemann, T. Clark. A QSPR-approach to the estimation of the pK(HB) of six-membered nitrogen-heterocycles using quantum mechanically derived descriptors, *Journal of Molecular Modeling* **2002**, *8*, 95-101. doi:10.1007/s00894-002-0075-z; b) M. Hennemann, A. Friedl, M. Lobell, J. Keldenich, A. Hillisch, T. Clark, A. H. Goller. CypScore: Quantitative prediction of reactivity toward cytochromes P450 based on semiempirical molecular orbital theory, *ChemMedChem* **2009**, *4*, 657-669. doi:10.1002/cmdc.200800384

[35] M. Connelly. The molecular surface package, *J. Mol. Graph.* **1992**, *11*, 139-141

[36] A. Bondi. van der Waals Volumes and Radii, *J. Phys. Chem.* **1964**, *68*, 441-451

[37] CTFile Formats. **https://www.daylight.com/meetings/mug05/Kappler/ctfile.pdf**

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

[39] G. B. Rocha, R. O. Freire, A. M. Simas, J. J. P. Stewart. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, *Journal of Computational Chemistry* **2006**, *27*, 1101-1111. doi:10.1002/jcc.20425

[40] J. J. P. Stewart. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements, *Journal of molecular modeling* **2007**, *13*, 1173-1213. doi:10.1007/s00894-007-0233-4

[41] H. B. Thomas, M. Hennemann, P. Kibies, F. Hoffgaard, S. Gussregen, G. Hessler, S. M. Kast, T. Clark. The hpCADD NDDO Hamiltonian: Parametrization, *J Chem Inf Model* **2017**, *57*, 1907-1922. doi:10.1021/acs.jcim.7b00080

[42] a) M. Kriebel, K. Weber, T. Clark. A Feynman dispersion correction: a proof of principle for MNDO, *J Mol Model* **2018**, *24*, 338. doi:10.1007/s00894-018-3874-6; b) M. Kriebel, A. Heßelmann, M. Hennemann, T. Clark. The Feynman dispersion correction for MNDO extended to F, Cl, Br and I, *Journal of Molecular Modeling* **2019**, *25*, 156. doi:10.1007/s00894-019-4038-z; c) M. Kriebel, A. Heßelmann, M. Hennemann, T. Clark. Correction to: The Feynman dispersion correction for MNDO extended to F, Cl, Br and I, *Journal of Molecular Modeling* **2019**, *25*, 257. doi:10.1007/s00894-019-4142-0

[43] T. Clark, K. G. Byler, M. J. De Groot. Biological Communication via Molecular Surface, in *Proceedings of the International Beilstein Workshop, Bozen, Italy, 2008)*, Logos Verlag, Bozen, Italy**,** 129-146. (**https://www.beilstein-institut.de/en/publications/proceedings/bozen-2006**)