CEPOS

InSilico

ParaSurf '07

Users´ Manual

TABLE OF CONTENTS

PROGRAM HISTORY

Release Date	Version	Platforms
1st July 2005	ParaSurf´05™ initial release (Revision A1)	32-bit Windows 32-bit Linux Irix
1st January 2006	ParaSurf´05™ Revision B1 (customer-feedback release)	
1st July 2006	ParaSurf´06™ Revision A1	32-bit Windows 32-bit Linux 64-bit Linux Irix
1st July 2007	ParaSurf´07™ Revision A1	32-bit Windows 32-bit Linux 64-bit Linux Irix

1 INTRODUCTION

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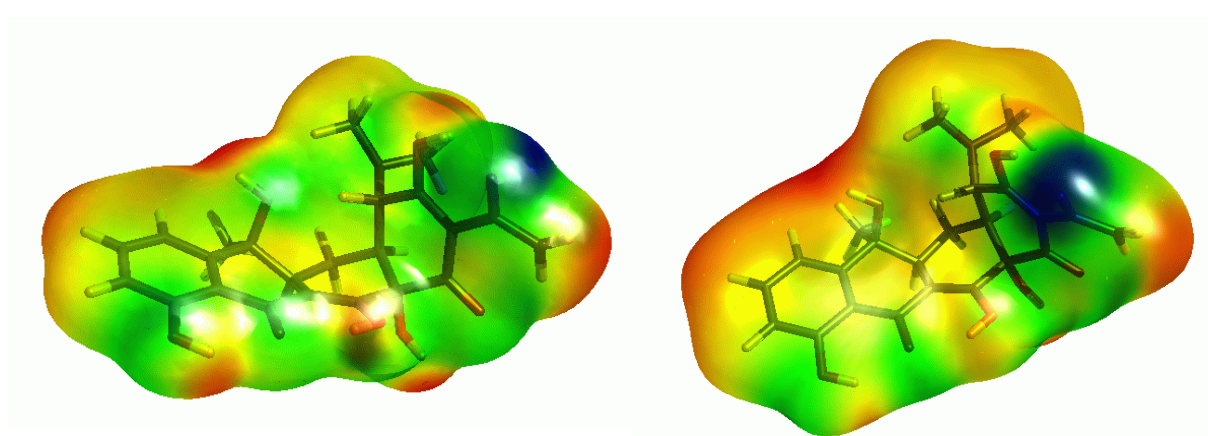
Figure 1:   Marching-cube (left) and shrink-wrap (right, fitted to a spherical-harmonic approximation) isodensity surfaces calculated with ParaSurf™ using the default settings.

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

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

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

ParaSurf™ is the first program to emerge from the ParaShift 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´06™

ParaSurf'07™ has been enhanced relative to its predecessor in order to provide better (=faster) performance, improved flexibility and a more comprehensive range of descriptors and features. The changes are outlined below:

1.1.1 Molecular surfaces

In ParaSurf'06™, isodensity surfaces were available for both shrink-wrap and marching-cube surfaces and solvent-excluded surfaces (SES) only for shrink-wrap. ParaSurf'07™ can also calculate solvent-excluded (SES) for both shrink-wrap and marching-cube and solvent-accessible surfaces (SAS) for marching cube. Using the SES marching-cube option gives a very significant speed enhancement compared with the isodensity marching-cube surface.

1.1.2 Surface options

The coarseness of the surface triangulation can now be defined by the user with the `mesh=n.nn` option and the radius of the solvent probe for SES and SAS can be set with the `rsol=n.nn` option.

1.1.3 Additional descriptors

ParaSurf'07™ can calculate an extended range of descriptors compared with ParaSurf'06™. On the one hand, additional descriptors (total integrated local property over the surface, skew, kurtosis) are available for the four established local properties (molecular electrostatic potential, local ionization energy, local electron affinity, local polarizability) and on the other hand, the equivalent range of descriptors is now calculated for the derived properties (local hardness and local electronegativity) or for the local properties defined in a SIM file for a surface-integral QSPR-model. The additional descriptors are described and defined in **1.9**.

1.1.4 User-definable molecular center

The user can define the molecular center to be used for the spherical-harmonic fitting either as Cartesian coordinates or as the center of gravity of listed atoms. This option is controlled *via* the input SDF file (see **2.2**).

1.1.5 Consistency checking

ParaSurf'07™ is stricter than ParaSurf'06™ when checking the input and the consistency of calculations. Unrecognized command-line arguments will, for instance, cause the program to abort, as will inconsistencies between the options defined for the external Shannon entropy or similarity calculations based on surface autocorrelations. The surface type, fitting, semiempirical Hamiltonian, solvent-probe radius or isodensity value as appropriate and the

method used to calculate the MEP must all be consistent before the external Shannon entropy or autocorrelation similarities can be calculated.

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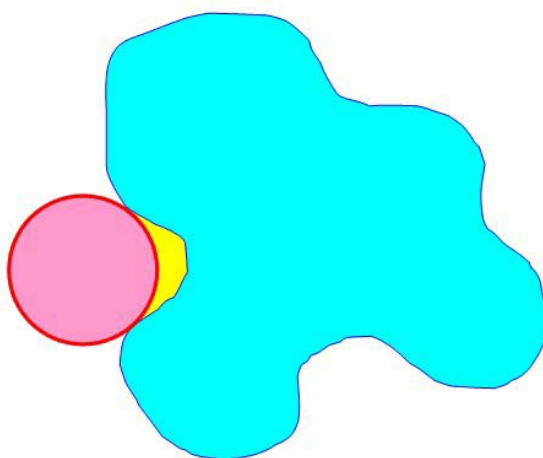
Figure 2: 2D-representation of a solvent-excluded surface.

1.4 Solvent-accessible surfaces

Solvent-accessible surfaces are obtained in the same way as solvent-excluded surfaces but the <u>outer</u> surface of the solvent sphere is used to define the molecular surface, as shown in **Figure 3**.
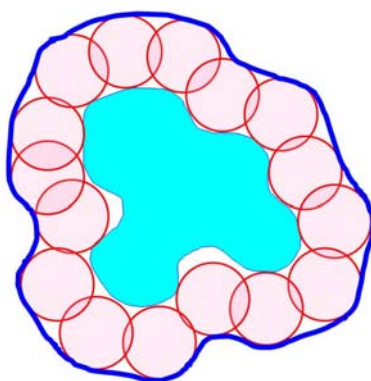
Figure 3: Marching-cube (left) and shrink-wrap (right, fitted to a spherical-harmonic approximation) isodensity surfaces calculated with ParaSurf™ using the default settings.

1.5 Shrink-wrap surface algorithm

Shrink-wrap surface algorithms [**3**] are used to determine single-valued molecular surfaces. Single-valued in this case means that for any given radial vector from the center of the molecule the surface is only crossed once (vectors **A** and **B** in **Figure 4**) and not multiply (vectors **C** and **D** in **Figure 4**):
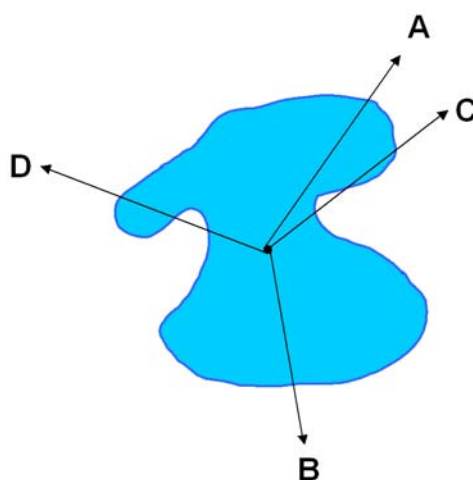
Figure 4: 2D-representation of a molecular surface with single-valued (**A** and **B**) and multiply valued (**C** and **D**) radial vectors from the center

Single-valued surfaces are necessary for spherical-harmonic fitting (see **1.4**). Thus, spherical-harmonic fitting is only available for shrink-wrap surfaces in ParaSurf™. The shrink-wrap algorithm works by starting outside the molecule (point **a** in **Figure 5**) and moving inwards along the radial vector until it finds the surface (in our case defined by the predefined level of the electron density, point **b** in **Figure 5**). Thus, the shrink-wrapped surface may contain areas (marked by dashed lines in **Figure 5**) for which the surface deviates from the true isodensity surface.

These areas of the surface, however, often have little consequence as they are situated above indentations in the molecule that are poorly accessible to solvents or other molecules. The shrink-wrapped surfaces generated by ParaSurf™ should normally be fitted to a spherical-harmonic series for use in HTVS, similarity, pattern-recognition or high-throughput docking applications. The default molecular center in ParaSurf™ is the center of gravity (CoG). In special cases in which the CoG lies outside the molecule, another center may be chosen.
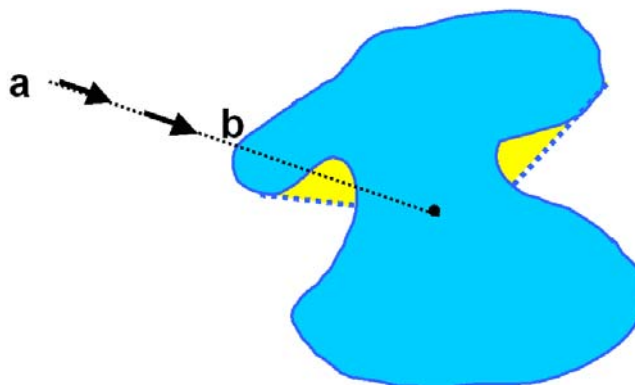
Figure 5: 2D-representation of the shrink-wrap algorithm. The algorithms scans along the vector from point **a** towards the center of the molecule until the electron density reaches the preset value (point **b**). The algorithm results in enclosures (marked yellow) for multi-valued radial vectors.

Figure 6 shows a spherical-harmonically fitted shrink-wrap surface for a difficult molecule. The areas shown schematically in **Figure 5** are clearly visible.
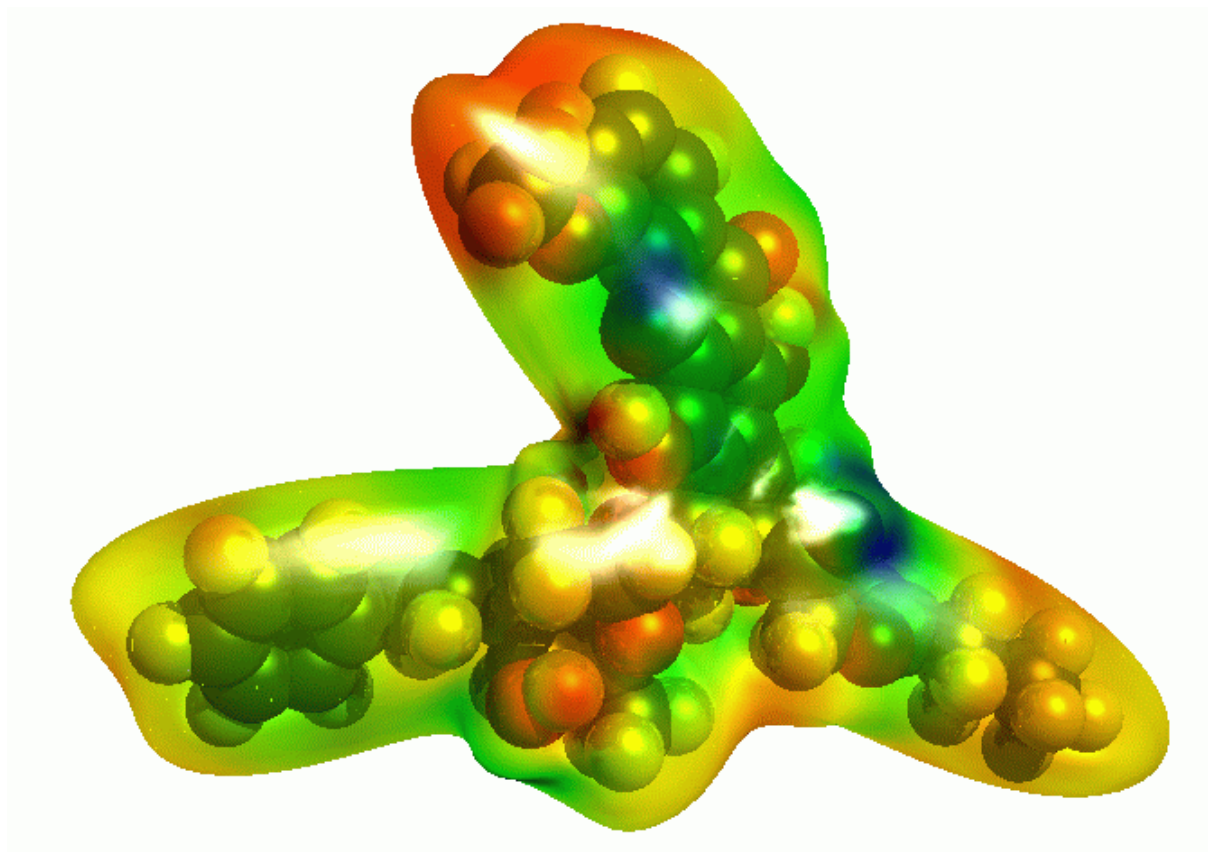
Figure 6: Spherical-harmonic approximation of a shrink-wrap isodensity surface. Note the areas where the surface does not follow the indentations of the moecule.

1.6 Marching-cube algorithm

The marching-cube algorithm [4] implemented in ParaSurf™ does not have the disadvantage of being single-valued like the shrink-wrap surface. It cannot, therefore, be fitted to a spherical harmonic series and is used as a purely numerical surface primarily for QSPR applications or surface-integral models. [10] The algorithm works by testing the electron density at the corners of cubes on a cubic lattice laid out through the molecular volume. The corners are divided into those "inside" the molecule (i.e. with a higher electron density than the preset value) and those "outside". The surface triangulation is then generated for each surface cube and the positions of the surface points corrected to the preset electron density.
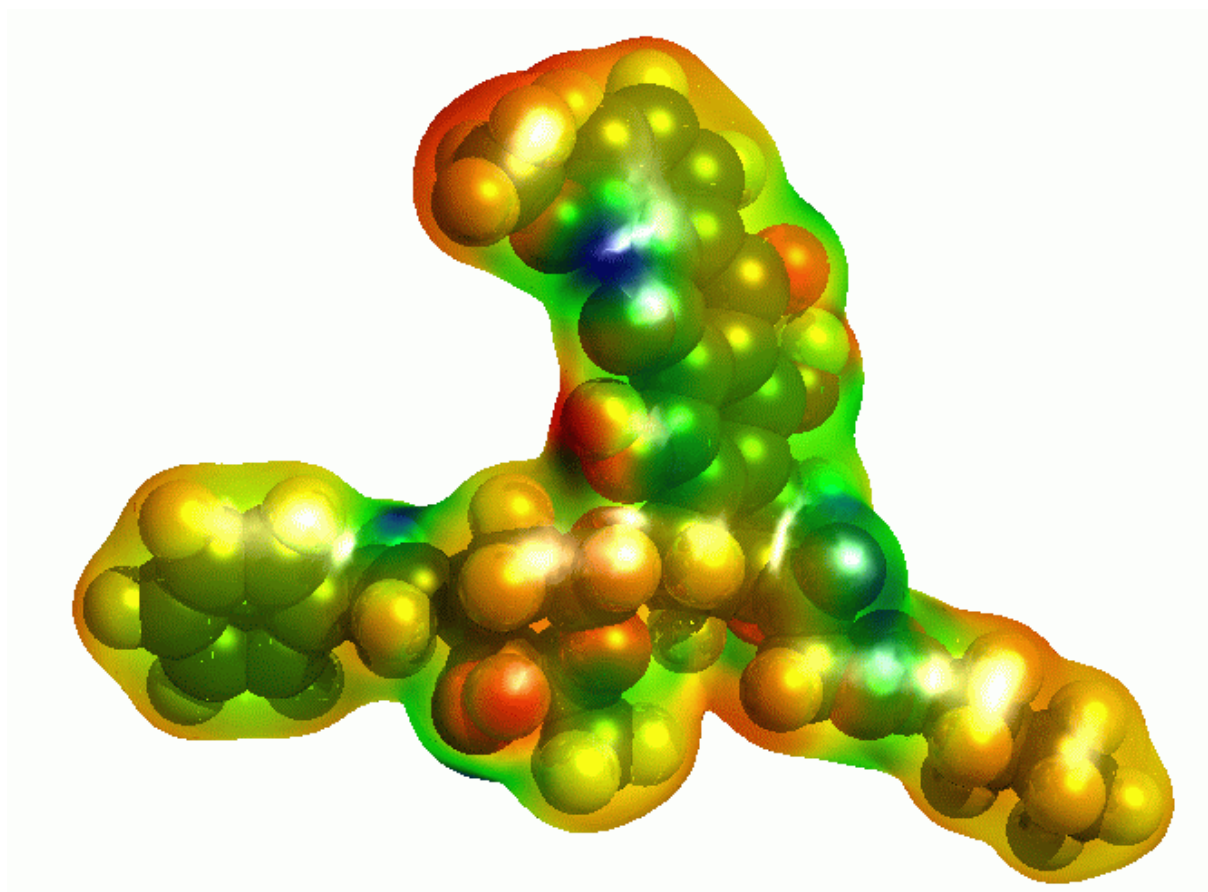
Figure 7: Marching-cube isodensity surface for the molecule shown in Figure 4. This surface is better suited for QSPR and surface-integral models

1.7 Spherical-harmonic fitting

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

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

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 $-l \le m \le l$. In the above form, spherical harmonics are complex functions. Duncan and Olson [12] have used the real functions

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

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

ParaSurf™ not only fits the surface itself (i.e. the radial distances) to spherical harmonic expansions, but also the four local properties (see **1.8**). 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 (0.0003 e⁻Å⁻³) is lower than that (0.008 e⁻Å⁻³) appropriate for an approximately van der Waals' surface using the marching-cube algorithm. The lower value avoids the surface coming too close to atoms. Note also that the fits are incremental, which means that the order chosen for a given application can be obtained by ignoring coefficients of higher order in the spherical-harmonic series.

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

1.8 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 [14,15] uses a total of nine point charges, one positive charge at the nucleus and eight negative ones distributed around it, to describe the electrostatics of a non-hydrogen atom with a valence-only *s*- and *p*-basis set for the semiempirical Hamiltonians MNDO, [16] AM1 [17] and PM3. [18] The negative charges are located at the charge centers of each lobe of the natural atomic orbitals, which are obtained by diagonalizing the one-atom blocks of the density matrix. [14] The NAO-PC charges are calculated by VAMP and output in the .sdf file for use in ParaSurf™. The NAO-PC model is therefore only available when using ParaSurf™ with VAMP .sdf input. NAO-PC charges are also not available for semiempirical Hamiltonians such as MNDO/d [19] or AM1* [20] that use *d*-orbitals in the basis set.

1.8.1.2 The multipole model

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

1.8.2 Local ionization energy, electron affinity, hardness and electronegativity

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

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

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

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

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

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

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

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

and the local electronegativity, χ_L:

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

1.8.3 Local polarizability

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

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

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

1.9 Descriptors

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

Table 1: The descriptors calculated by ParaSurf™.

Descriptor	Description	Formula/ Reference	Symbol in CSV file
μ	Dipole moment		**dipole**
μ_D	Dipolar density	[25]	**dipden**
α	Molecular electronic polarizabilty	[24]	**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**

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\overline{V}_+$	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	$\gamma_1^V = \dfrac{\sum_{i=1}^{N}\left(V_i - \overline{V}\right)^3}{(N-1)\sigma^3}$	`MEPskew`
γ_2^V	Kurtosis of the MEP-distribution	$\gamma_2^V = \dfrac{\sum_{i=1}^{N}\left(V_i - \overline{V}\right)^4}{(N-1)\sigma^4} - 3$	`MEPkurt`
$\int_V$	Integrated MEP over the surface	$\int_V = \sum_{i=1}^{N} V_i a_i$	`MEPint`
IE_L^{max}	Maximum value of the local ionization energy		`IELmax`
IE_L^{min}	Minimum value of the local ionization energy		`IELmin`
$\overline{IE_L}$	Mean value of the local ionization energy	$\overline{IE_L} = \dfrac{1}{N}\sum_{i=1}^{N} IE_L^i$	`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	$\sigma_{IE}^2 = \dfrac{1}{N}\sum_{i=1}^{N}\left[IE_L^i - \overline{IE_L} \right]^2$	`IELvar`
$\gamma_1^{IE_L}$	Skewness of the local ionization energy distribution	$\gamma_1^{IE_L} = \dfrac{\sum_{i=1}^{N}\left(IE_L^i - \overline{IE_L}\right)^3}{(N-1)\sigma^3}$	`IELskew`

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

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

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

Descriptor	Description	Formula/ Reference	Symbol in CSV file
$\overline{H_{ex}}$	Mean value of the external Shannon Entropy	$\overline{H_{ex}} = \dfrac{1}{N}\sum\limits_{i=1}^{N} H_{ex}^{i}$	**SHANEbar**
$\sigma^2_{H_{ex}}$	Variance in the external Shannon Entropy	$\sigma^2_{H_{ex}} = \dfrac{1}{N}\sum\limits_{i=1}^{N}\left[H_{ex}^{i} - \overline{H_{ex}} \right]^2$	**SHANEvar**
$\int_{H_{ex}}$	Integrated internal Shannon Entropy over the surface	$\int_{H_{ex}} = \sum\limits_{i=1}^{N} H_{ex}^{i} a_i$	**SHANEtot**

1.10 Surface-integral models

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

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

(10)

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

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

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

1.11 Spherical harmonic "hybrids"

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

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

(11)

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

1.12 Descriptors and moments based on surface-integral models

ParaSurf'06™ uses local properties defined in a surface-integral model (SIM, **see 1.10**) to calculate descriptors analogous to those listed in **Table 1**. Additionally, "dipolar moments of the local property are calculated. These are gauge-independent moments calculated by first shifting values of the local property so that their sum is zero and then calculating moments according to

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

(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 logP$_{OW}$ is shown below:

```
   Descriptors calculated for logP:

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

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

1.13 Shannon entropy

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

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

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

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

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

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

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

where $p(V, I, E, \alpha)$ 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

$$H = -\int p(V) \log_2 p(V) dV - \int p(I) \log_2 p(I) dI$$
$$-\int p(E) \log_2 p(E) dE - \int p(\alpha) \log_2 p(\alpha) d\alpha$$
(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

$$H = -\sum_{i=1}^{k} \left[p(V_i) \log_2 p(V_i) + p(I_i) \log_2 p(I_i) + p(E_i) \log_2 p(E_i) + p(\alpha_i) \log_2 p(\alpha_i) \right] \cdot A_i$$
(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'07™ offers two alternatives as sources for the probabilities $p(X_i)$. The first, known as the "external" Shannon entropy, is to use probabilities taken from an external dataset and defined in a

separate statistics file. The default "external" statistics file is called **bins.txt** and is read from the ParaSurf root directory. The statistics defined in **bins.txt** were derived from AM1 calculations of all the bound ligands defined in the PDBbind database [**29**] in their correct protonation states and at geometries obtained by optimizing with AM1 starting from the bound conformation**.** [**30**]

Alternatively, the user can define a custom "external" statistics file using the ParaSurf module **binner** (available free of charge for ParaSurf[TM] 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. [**30**]

1.14 Surface autocorrelations

Gasteiger et al. [**31**] introduced the concept of surface autocorrelations as powerful descriptions of molecular binding properties for quantitative structure-activity relationships (QSARs). In ParaSurf'07™, autocorrelations *A(R)* are defined as:

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

(18)

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

Four different autocorrelation functions are calculated by ParaSurf'07™. These are:

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

Autocorrelation functions based on the other three local properties correlate very strongly with the shape autocorrelation and are therefore not calculated.
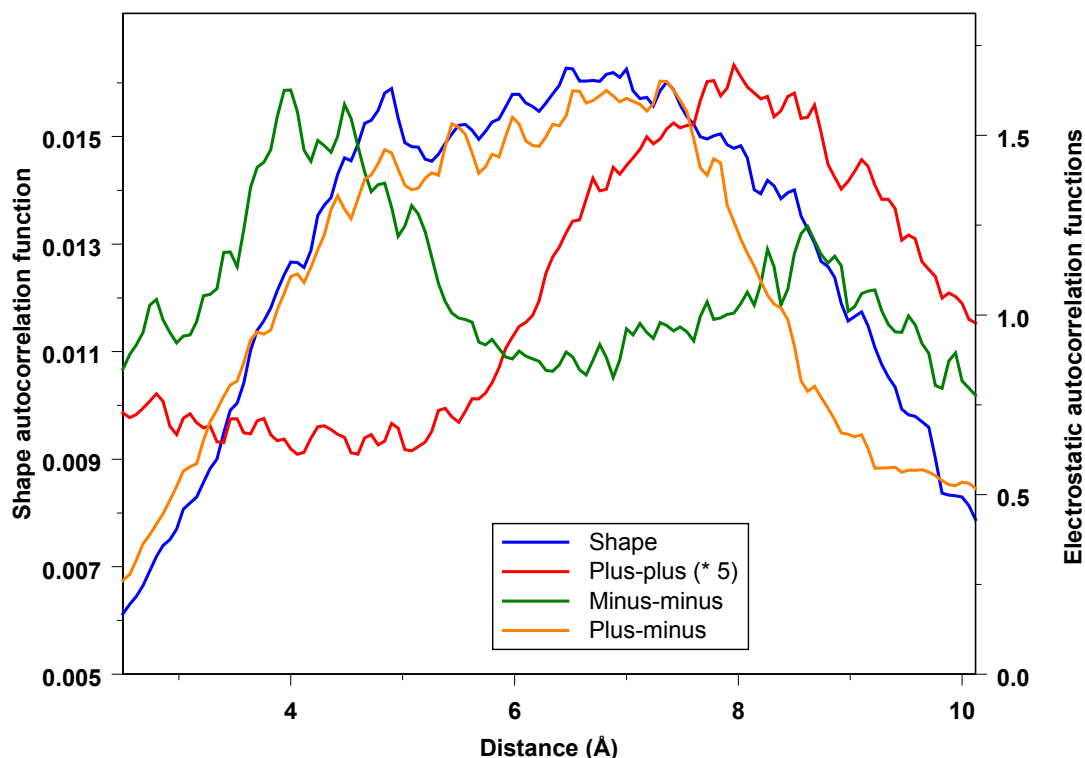
Figure 8: The four autocorrelation functions calculated using the AM1 Hamiltonian for trimethoprim.

ParaSurf'07™ calculates autocorrelations as vectors of *A(R)* values 128 elements long starting at an *R*-value of 2.5 Å and increasing in steps of 0.06 Å (i.e. up to a maximum value of 10.12 Å). **Figure 8** shows the four autocorrelation functions for trimethoprim calculated with AM1.

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

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

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

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

2 PROGRAM OPTIONS

2.1 Command-line options

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

Table 2: ParaSurf™ command-line options

`<name>`		Base name for the input file (must be the first argument)
		Using this option, the input file is assumed to be `<name>_v.sdf` if a file with this name exists.
		Otherwise the file `<name>.sdf` will be used as input.
		If neither of these files are found, the program will use an .sdf file written by the Cepos version of Mopac 6. These files are called `<name>_m.sdf`
		The output files are `<name>_p.out` `<name>_p.sdf` `<name>.psf` (optional) `<name>.asd` (optional) `<name>_p.vmp` (optional)
`surf=`	`wrap`	Shrink-wrap surface (default)
	`cube`	Marching-cube surface
`contour=`	`isoden`	The surface is defined by the electron density
	`solvex`	A solvent-excluded surface is used (only active for `surf=wrap` in this version)
`fit=`	`sphh`	Spherical-harmonic fitting (default for `surf=wrap`)
	`isod`	Smooth to preset isodensity value (default for `surf=cube`)
	`none`	No fitting
`iso=`	`n.nn`	Isodensity value set to `n.nn` $e^-Å^{-3}$ (default for shrink-wrap surface = `0.0002`; default for marching-cube surface = `0.003`; minimum possible value = `0.00001`)
`rsol=`	`n.nn`	A solvent-probe radius of `n.nn` Å is used for calculating the solvent-excluded or solvent-accessible surface (default=`1.0`, allowed range is from `0.0` to `2.0` Å)

mesh=	*n.nn*	The mesh size used to triangulate the surface is set to **n.nn** Å (default value = **0.2** Å, allowed range is from **0.1** to **1.0** Å)
estat=	**naopc**	Use NAO-PC electrostatics
	multi	Use multipole electrostatics (default)
psf=	**on**	Write .psf surface file
	off	Do not write **.psf** surface file (default)
asd=	**on**	Write anonymous SD (**.asd**) file
	off	Do not write .asd file (default)
vmp=	**on**	Write **.vmp** file for debugging. Map the MEP onto the surface
	off	Do not write .vmp file (default)
	mep	Write **.vmp** file for debugging. Map the MEP onto the surface
	iel	Write **.vmp** file for debugging. Map IE_L onto the surface
	eal	Write **.vmp** file for debugging. Map EA_L onto the surface
	pol	Write **.vmp** file for debugging. Map α_L onto the surface
	har	Write **.vmp** file for debugging. Map η_L onto the surface
	eng	Write **.vmp** file for debugging. Map χ_L onto the surface
	<MOD>	Write **.vmp** file for debugging. Map the local property with the three-character designator **<MOD>** defined in the SIM file onto the surface
grid=	**<filename>**	Read the Cartesian coordinates at which to calculate a grid of the four properties (MEP, IE_L, EA_L, α_L). See **3.8.1**
	auto	ParaSurf™ calculates an automatic grid (see **3.8.2**)
lattice=	*n.nn*	Sets the lattice spacing for the **grid=auto** option (see **3.8.2**)
sim=	**<filename>**	One or more surface-integralmodels will be read from the file **<filename>.sim** in the ParaSurf™ executable directory. **<filename>** can be upper or lower case or any mixture but must be exactly three characters long.
center= or **centre=**	**on**	The atomic and surface coordinates in the **.psf** output file will be centered for calculations that use spherical-harmonic fitting. Note that this means that the atomic coordinates in the **SDF-**output file (which are the input coordinates) will be different to those in the **PSF**-output file. This option is default.
	off	The atomic and surface coordinates in the **.psf** output file will not be centered and will correspond to the input coordinates and those in the **SDF**-output file.
shannon	**=<filename>**	Requests that Shannon entropies (both internal and external) be calculated. If no statistics file **<filename>** is given, the default file (**bins.txt** in the ParaSurf Root directory) will be used. If a statistics file is given that either does not exist, contains errors or is derived from ParaSurf™ runs using different options to the current one, only the internal Shannon entropy is calculated.
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 autocorrelation similarities will be written to the file `<filename>`. If `<filename>` exists, the values for the current molecule will be appended to the existing table, otherwise the file will be created.

Examples:

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

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

> `parasurf test surf=cube fit=none`

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

2.2 Options defined in the input SDF-file

2.2.1 Defining the center for spherical-harmonic fits

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

`<SPHH_CENTER>`

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

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

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

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

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

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

3 INPUT AND OUTPUT FILES

ParaSurf™ uses the following files for input and output:

Table 3: ParaSurf™ input and output files

File	Name	Description
Input	**`<filename>_v.sdf`** or **`<filename>.sdf`**	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 available) or **`<filename>_m.sdf`**	If no VAMP .sdf file is found, ParaSurf™ defaults to a Cepos Mopac 6 .sdf file. It is strongly recommended to use the **`EF`** option for geometry optimizations in Mopac.
Hamiltonian	**`Vhamil.par`**	The VAMP parameters file (also found in the VAMP executable directory). This file must be copied to the ParaSurf™ executable directory.
Output	**`<filename>_p.out`**	Always written.
SD-file	**`<filename>_p.sdf`**	Always written.
ASD-file	**`<filename>.asd`**	Anonymous SD-file. Requested by the option **`asd=on`**
PSF-file	**`<filename>.psf`**	ParaSurf™ surface file. Requested by the option **`psf=on`**
VMP-file	**`<filename>_p.vmp`**	Debug file.
SIM-file	**`<filename>.sim`**	Surface-integral model definition. **`<filename>`** must have exactly three characters and the file must reside in the ParaSurf™ executable directory.
Descriptor table file	**`User defined`**	An ascii, comma-separated file that contains a line of descriptors for each molecule. This file will be created if it does not exist or an extra line will be appended if it does exist.
Autocorrelation similarity file	**`User defined`**	An ascii, comma-separated file that contains a line of autocorrelation similarities for each molecule. This file will be created if it does not exist or an extra line will be appended if it does exist.

3.1 The VAMP .sdf file as input

VAMP .sdf files, an extension of the MDL .sdf file format, [32] are the primary communication channel between VAMP and ParaSurf™. The atomic coordinates and bond definitions are given in the MDL format as shown in **Figure 9**.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 9: The headers and titles, atomic coordinates and bond definitions from a VAMP .sdf file. The format follows the MDL definition. [29]

<HAMILTONIAN>

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

<NAO-PC>

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

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

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

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

`<VAMPBASICS>`

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

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

`<TOTAL COULSON CHARGE>`

The total charge of the molecule.

`<DENSITY MATRIX ELEMENTS>`

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

`<ORBITAL VECTORS>`

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

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

3.1.1 Multi-structure SD-files

ParaSurf'07™ 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'07™ 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'07™ functionality within the previous restrictions that:

- Polarizabilities are not yet available for Hamiltonians that use *d*-orbitals (MNDO/d and AM1*).

- NAO-PC electrostatics are only available if the NAO-PCs are present in the SD-file. Multipole electrostatics are available for all Hamiltonians.

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

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

3.2 The Cepos MOPAC 6.sdf file as input

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

3.3 The Vhamil.par file

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

3.4 The ParaSurf™ output file

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

3.4.1 For a spherical-harmonic surface

Figure 10 shows the output for a calculation using the options **surf=wrap fit=sphh** for trimethoprim, **1**.
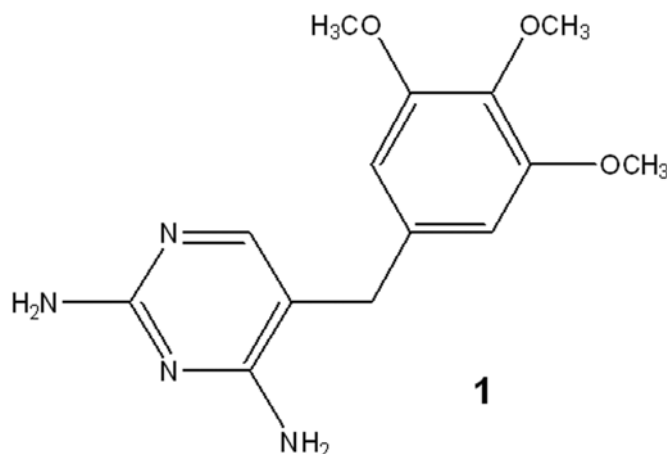

```
<> ParaSurf'07, Revision A1
<> Copyright (c) 2006,2007, Friedrich-Alexander-
Universitaet Erlangen-Nuernberg and
                        Cepos InSilico Ltd. All rights
reserved.

 <> Input = trimethoprim.sdf

 <> Program options :

    Using shrink-wrap isocontour surface
    Fitting surface to spherical harmonics
    Using an isodensity surface contour
    Isodensity value = 0.2000E-04 electrons/Angstrom**3
    Triangulation mesh =    0.20 Angstrom
    Using multipole electrostatics
```

Figure 10: ParaSurf® output for trimethoprim, **1**, using a spherical-harmonic surface.

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

      0       1.88757071
      1       1.93149693
      2       1.56818807
      3       1.15088494
      4       0.96172361
      5       0.75624534
      6       0.67247770
      7       0.58544266
      8       0.53423629
      9       0.49857261
     10       0.47886902
     11       0.45241646
     12       0.42436990
     13       0.40734665
     14       0.39757528
     15       0.38641552

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

      0      10.69283689
      1      10.67644860
      2       8.49899500
      3       6.72705215
      4       5.34777092
      5       4.50078452
      6       4.07825917
      7       3.33727705
      8       2.99574760
      9       2.69150818
     10       2.34647480
     11       2.05834398
     12       1.89975660
     13       1.69754320
     14       1.59755726
     15       1.43858209
     16       1.24892060
     17       1.07142579
     18       0.93838458
     19       0.86663917
     20       0.80393197
```

Figure 10: continued

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

      0       57.45749229
      1       51.13710977
      2       45.83807403
      3       45.27270405
      4       41.23895290
      5       36.66650506
      6       33.61785893
      7       27.34348870
      8       23.10749351
      9       19.55726604
     10       18.15965159
     11       16.77937642
     12       15.45879611
     13       14.67455377
     14       13.92771424
     15       13.44040657
     16       13.06653741
     17       12.47902099
     18       12.18875500
     19       12.18875500
     20       12.18875500

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

      0       12.04124628
      1       11.62731542
      2       11.42972402
      3        9.25104060
      4        8.57421604
      5        7.10175169
      6        6.95985005
      7        6.67722731
      8        6.35873742
      9        5.82786756
     10        5.54052467
     11        5.23489464
     12        4.89230363
     13        4.31390979
     14        3.87471992
     15        3.69197211
     16        3.41319944
     17        3.34063485
     18        3.13831701
     19        3.04024367
     20        2.94602890
```

Figure 10: continued

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

      0        0.02373155
      1        0.01651978
      2        0.01352180
      3        0.01094315
      4        0.00905235
      5        0.00801871
      6        0.00751616
      7        0.00716503
      8        0.00679400
      9        0.00634420
     10        0.00587568
     11        0.00565478
     12        0.00511256
     13        0.00511256
     14        0.00511256
     15        0.00511256
     16        0.00511256
     17        0.00511256
     18        0.00511256
     19        0.00511256
     20        0.00511256

 <> Property ranges:
    Density  :   0.2647E-05 to   0.9549E-04
    IE(l)    :       390.99 to       671.01
    EA(l)    :      -108.57 to       -38.44
    MEP      :       -47.44 to        15.98
    Alpha(l) :       0.2372 to        0.3376
```

Figure 10: continued

```
<> Descriptors :

   Dipole moment       :         1.2467 Debye
   Dipolar density     :     0.1887E-02 Debye.Angstrom**-3
   Molecular pol.      :       128.5408 Angstrom**3
   Molecular weight    :         290.32
   Globularity         :         0.7705
   Total surface area  :         475.99 Angstrom**2
   Molecular volume    :         660.49 Angstrom**3

   Most positive MEP   :          15.98 kcal/mol
   Most negative MEP   :         -47.44 kcal/mol
   Mean +ve MEP        :           5.38 kcal/mol
   Mean -ve MEP        :         -10.35 kcal/mol
   Mean MEP            :          -2.82 kcal/mol
   MEP range           :          63.42 kcal/mol
   MEP +ve Variance    :          10.81 kcal/mol
   MEP -ve Variance    :          89.72 kcal/mol
   MEP total variance  :         100.52 kcal/mol
   MEP balance parameter:        0.0959
   MEP balance*variance :        9.6451 kcal/mol
   MEP skewness        :        -1.2065
   MEP kurtosis        :         1.4632
   Integral MEP        : -1139.31      kcal.Angstrom**2/mol

   Maximum IE(l)       :         671.01 kcal/mol
   Minimum IE(l)       :         390.99 kcal/mol
   Mean IE(l)          :         473.73 kcal/mol
   IE(l) range         :         280.02 kcal/mol
   IE(l) variance      :        3276.81 kcal/mol
   IE(l) skewness      :         0.6491
   IE(l) kurtosis      :        -0.3030
   Integral IE(l)      :  9776.18      eV.Angstrom**2

   Maximum EA(l)       :         -38.44 kcal/mol
   Minimum EA(l)       :        -108.57 kcal/mol
   Mean +ve EA(l)      :           0.00 kcal/mol
   Mean -ve EA(l)      :         -94.33 kcal/mol
   Mean EA(l)          :         -94.33 kcal/mol
   EA(l) range         :          70.13 kcal/mol
   EA(l) +ve variance  :           0.00 kcal/mol
   EA(l) -ve variance  :         137.85 kcal/mol
   EA(l) total variance :        137.85 kcal/mol
   EA(l) skewness      :         1.8077
   EA(l) kurtosis      :         4.2701
   Integral EA(l)      : -1943.17      eV.Angstrom**2
   EA(l) balance param. :         0.0000
   Fraction pos. EA(l) :         1.0000 ( =  475.99 Angstrom**2)
```

Figure 10: continued

```
    Max. local Eneg.      :       299.56 kcal/mol
    Min. local Eneg.      :       143.14 kcal/mol
    Mean local Eneg.      :       189.70 kcal/mol
    Local Eneg. range     :       156.42 kcal/mol
    Local Eneg. variance  :       973.42
    Local Eneg. skewness  :         0.79
    Local Eneg. kurtosis  :        -0.01
    Integral local Eneg.  :  3916.50      eV.Angstrom**2

    Max. local hardness   :       371.45 kcal/mol
    Min. local hardness   :       247.42 kcal/mol
    Mean local hardness   :       284.03 kcal/mol
    Local hard. range     :       124.03 kcal/mol
    Local hard. variance  :       733.91
    Local hard. skewness  :         0.55
    Local hard. kurtosis  :        -0.57
    Integral local Hard.  :  5859.68      eV.Angstrom**2

    Maximum alpha(l)      :     0.3376    Angstrom**3
    Minimum alpha(l)      :     0.2372    Angstrom**3
    Mean alpha(l)         :     0.2819    Angstrom**3
    Alpha(l) range        :     0.1004    Angstrom**3
    Variance in alpha(l)  :     0.5474E-03 Angstrom**3
    Alpha(l) skewness     :        -0.8012
    Alpha(l) kurtosis     :        -0.6030
    Integral Alpha(l)     :   134.559    Angstrom**5

<> Spherical-Harmonic Hybridization:

   Shape hybrids         :
  17.725447    1.113716    3.467172    2.864957    1.413661    1.608636
   0.750132    0.692325    0.461673    0.396041    0.385152    0.314405
   0.245721    0.215156    0.211112    0.196789

   MEP hybrids           :
  12.873244    4.619311   25.550444   18.123098   13.554066   10.376171
   6.831225    7.849237    4.502953    4.736035    4.204671    3.450284
   2.353302    2.637218    1.602750    1.990410    1.862434    1.796871
   1.441618    0.995770    0.887875

   IE(l) hybrids         :
 1696.8928      91.5637     79.8973     62.1777     59.4791     54.6453
   52.4540      54.8521     39.9406     44.8333     24.9278     24.7114
   20.5216      17.0634     17.8577     14.6759     14.9262     17.7538
   15.4406       0.0000      0.0000

   EA(l) hybrids         :
  324.9128       6.2224     12.3595     25.2870     13.7906     16.5039
    8.3390       9.0377      7.7882      7.5410      6.0949      5.1157
    5.0100       7.3909      5.4569      4.2611      4.7894      3.8096
    3.7308       3.3984      3.1923

   Alpha(l) hybrids      :
 1.01338637  0.05270670  0.03322054  0.02969108  0.02167777  0.01484906
 0.01186878  0.00891221  0.00720500  0.00856844  0.00742688  0.00615349
 0.00673266  0.00000000  0.00000000  0.00000000  0.00000000  0.00000000
 0.00000000  0.00000000  0.00000000
```

Figure 10: continued

```
<> Atomic surface properties:


  Atom      Area        MEP            IE(l)           EA(l)        mean
                     max      min     max      min     max      min     pol.
 C     1   0.000
 O     2   0.073   -37.67  -40.72  534.28  531.02  -78.53  -80.78   0.261
 C     3   3.149    -6.35  -45.91  576.66  481.92  -44.41  -91.58   0.307
 C     4   1.002    -7.47  -14.46  569.68  494.06  -72.14  -90.58   0.323
 C     5   0.598    -8.07  -13.68  572.51  533.84  -84.84  -92.37   0.321
 C     6   0.000
 C     7   0.787    -9.47  -14.62  559.74  515.58  -69.96  -92.79   0.318
 C     8   4.188    -1.37  -21.91  584.46  484.63  -50.34  -98.61   0.295
 N     9   3.608   -14.57  -30.14  535.53  457.65  -80.69 -105.19   0.280
 C    10  10.127    -2.15  -27.13  634.48  533.15  -38.87  -88.28   0.284
 N    11   0.000
 N    12   1.266   -14.21  -31.88  540.14  475.28  -74.70  -99.75   0.268
 C    13   6.251    -8.23  -25.60  637.41  512.95  -38.44  -84.12   0.285
 N    14   0.000
 C    15   1.665    -9.15  -15.25  587.74  501.97  -64.39  -92.37   0.318
 C    16   3.762   -10.57  -41.98  566.59  477.64  -40.14  -98.42   0.311
 O    17   0.000
 C    18   0.000
 C    19   6.455   -17.23  -47.44  580.78  464.98  -47.22  -86.59   0.315
 O    20   1.707   -32.36  -42.43  531.52  447.38  -76.58  -93.16   0.247
 C    21   0.000
 H    22  33.853    12.47  -38.77  559.56  405.86  -83.05  -99.76   0.298
 H    23  21.782    13.72  -20.08  562.97  408.40  -81.95  -95.37   0.294
 H    24  24.164    13.72  -34.87  568.52  407.10  -71.28  -95.34   0.292
 H    25   6.907    10.98   -4.96  523.49  425.97  -84.81  -95.58   0.289
 H    26  18.928     7.80   -8.81  532.71  400.18  -88.39 -100.67   0.299
 H    27  18.168     7.23  -19.87  541.87  399.65  -69.14 -100.82   0.302
 H    28  27.487     8.72  -27.12  534.21  413.49  -62.62 -103.75   0.284
 H    29  34.200    15.98  -27.94  671.01  462.67  -71.89 -107.85   0.248
 H    30  33.730    15.73  -28.39  669.16  480.29  -71.38 -107.93   0.241
 H    31  34.865    11.40  -27.38  659.58  474.79  -68.82 -108.57   0.245
 H    32   9.589    11.30  -25.59  643.32  481.53  -73.06  -99.65   0.258
 H    33   5.605     9.42  -13.45  526.17  436.32  -73.75  -96.49   0.296
 H    34  30.322    11.00  -38.85  565.01  406.82  -81.91 -100.21   0.291
 H    35  24.185    11.18  -32.32  567.86  407.69  -67.97  -95.68   0.292
 H    36  23.002    11.12  -13.00  565.07  407.56  -80.52  -99.35   0.295
 H    37  34.556     4.48  -37.25  554.55  394.88  -88.71 -108.40   0.294
 H    38  19.255     4.18  -39.51  553.49  396.56  -83.75 -108.16   0.298
 H    39  26.373     3.82  -33.01  551.57  390.99  -88.07 -107.59   0.295

 Total    471.606

 <> ParaSurf used   29.91 seconds CPU time
```

Figure 10: continued

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

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

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

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

3.4.2 For a marching-cube surface

Figure 11 shows the output for a calculation using the options `surf=cube` for trimethoprim.

```
<> ParaSurf'07, Revision A1
<> Copyright (c) 2006,2007, Friedrich-Alexander-Universitaet
                            Erlangen-Nuernberg and Cepos InSilico
                            Ltd. All rights reserved.

 <> Input = trimethoprim.sdf

 <> Program options :

    Using marching-cube solvent-excluded surface
    Surface fitting turned off
    Using the solvent-excluded surface contour
    Solvent radius =     1.0000 Angstrom
    Triangulation mesh =    0.20 Angstrom
    Using multipole electrostatics


 <> AM1   calculation for  Trimethoprim

 <> Property ranges:
    Density  :   0.5005E-04 to   0.5096E-02
    IE(l)    :       394.23 to       683.82
    EA(l)    :      -111.55 to       -18.67
    MEP      :       -67.75 to        46.19
    Alpha(l) :       0.2266 to        0.3296
```

Figure 11: ParaSurf™ output for trimethoprim using a marching-cube surface

```
<> Descriptors :

    Dipole moment          :          1.2467 Debye
    Dipolar density        :     0.4622E-02 Debye.Angstrom**-3
    Molecular pol.         :        128.5408 Angstrom**3
    Molecular weight       :        290.32
    Globularity            :          0.6721
    Total surface area     :        300.35 Angstrom**2
    Molecular volume       :        269.70 Angstrom**3

    Most positive MEP      :          46.19 kcal/mol
    Most negative MEP      :         -67.75 kcal/mol
    Mean +ve MEP           :          14.17 kcal/mol
    Mean -ve MEP           :         -23.00 kcal/mol
    Mean MEP               :          -4.37 kcal/mol
    MEP range              :         113.94 kcal/mol
    MEP +ve Variance       :          77.90 kcal/mol
    MEP -ve Variance       :         260.10 kcal/mol
    MEP total variance     :         338.00 kcal/mol
    MEP balance parameter:            0.1774
    MEP balance*variance   :          59.9462 kcal/mol
    MEP skewness           :          -0.5246
    MEP kurtosis           :          -0.4608
    Integral MEP           : -1394.66       kcal.Angstrom**2/mol

    Maximum IE(l)          :         683.82 kcal/mol
    Minimum IE(l)          :         394.23 kcal/mol
    Mean IE(l)             :         497.18 kcal/mol
    IE(l) range            :         289.60 kcal/mol
    IE(l) variance         :        3923.55 kcal/mol
    IE(l) skewness         :          0.2972
    IE(l) kurtosis         :          -0.8192
    Integral IE(l)         :  6487.02       eV.Angstrom**2

    Maximum EA(l)          :         -18.67 kcal/mol
    Minimum EA(l)          :        -111.55 kcal/mol
    Mean +ve EA(l)         :           0.00 kcal/mol
    Mean -ve EA(l)         :         -86.15 kcal/mol
    Mean EA(l)             :         -86.15 kcal/mol
    EA(l) range            :          92.89 kcal/mol
    EA(l) +ve variance     :           0.00 kcal/mol
    EA(l) -ve variance     :         356.44 kcal/mol
    EA(l) total variance :           356.44 kcal/mol
    EA(l) skewness         :          1.2576
    EA(l) kurtosis         :          0.7070
    Integral EA(l)         : -1116.88       eV.Angstrom**2
    EA(l) balance param. :            0.0000
    Fraction pos. EA(l)  :            1.0000 ( =   300.35 Angstrom**2)

    Max. local Eneg.       :         301.48 kcal/mol
    Min. local Eneg.       :         143.83 kcal/mol
    Mean local Eneg.       :         205.52 kcal/mol
    Local Eneg. range      :         157.65 kcal/mol
    Local Eneg. variance :          1375.42
    Local Eneg. skewness :             0.37
    Local Eneg. kurtosis :            -0.98
    Integral local Eneg. :  2685.07       eV.Angstrom**2
```

Figure 11: continued

```
   Max. local hardness  :        384.79 kcal/mol
   Min. local hardness  :        248.64 kcal/mol
   Mean local hardness  :        291.67 kcal/mol
   Local hard. range    :        136.14 kcal/mol
   Local hard. variance :        764.57
   Local hard. skewness :          0.38
   Local hard. kurtosis :         -0.58
   Integral local Hard. :  3801.95     eV.Angstrom**2

   Maximum alpha(l)     :    0.3296    Angstrom**3
   Minimum alpha(l)     :    0.2266    Angstrom**3
   Mean alpha(l)        :    0.2810    Angstrom**3
   Alpha(l) range       :    0.1030    Angstrom**3
   Variance in alpha(l) :    0.4797E-03 Angstrom**3
   Alpha(l) skewness    :      -0.6529
   Alpha(l) kurtosis    :      -0.5869
   Integral Alpha(l)    :   84.4666    Angstrom**5

<> Atomic surface properties:

  Atom     Area      MEP           IE(l)         EA(l)        mean
                   max    min    max    min    max    min    pol.
C    1   1.552   34.27  -54.43 578.41 512.64 -77.19  -99.48  0.271
O    2   3.187  -19.02  -61.40 570.37 457.34 -66.69  -81.41  0.259
C    3   7.294   -5.32  -62.24 670.79 508.97 -22.25  -83.15  0.299
C    4   3.773    5.32  -25.07 679.79 500.02 -22.00  -93.08  0.309
C    5   2.760    4.79  -23.30 661.42 558.30 -35.32  -92.42  0.304
C    6   0.202    7.19  -12.51 600.62 555.85 -83.09  -99.69  0.297
C    7   3.305   -0.36  -32.00 637.47 513.62 -36.28  -89.88  0.316
C    8   6.483    7.98  -30.89 680.71 487.94 -29.35  -93.91  0.284
N    9   7.733  -17.33  -61.50 606.55 417.93 -45.55 -103.33  0.256
C   10   9.640    0.87  -54.90 683.82 547.29 -40.71  -84.68  0.276
N   11   1.110  -40.69  -62.87 615.73 576.98 -57.20  -81.52  0.272
N   12   7.186  -13.44  -59.38 609.54 415.22 -44.91  -99.67  0.244
C   13   8.035   -6.93  -51.64 667.98 545.05 -32.82  -87.44  0.280
N   14   1.362  -39.84  -67.75 625.17 574.61 -59.92  -83.67  0.283
C   15   5.226   -8.08  -29.49 681.96 489.72 -18.67  -86.79  0.311
C   16   6.967  -13.61  -61.14 682.14 524.73 -20.19  -85.58  0.302
O   17   0.905  -29.09  -61.20 562.84 466.85 -66.22  -85.51  0.247
C   18   1.508   26.85  -57.28 573.99 523.41 -73.25  -98.83  0.273
C   19   6.773  -13.93  -57.86 677.11 502.08 -32.47  -86.95  0.309
O   20   3.834  -31.54  -58.10 567.10 438.36 -70.54  -94.61  0.253
C   21   1.578   13.05  -55.89 569.64 503.52 -87.52 -111.55  0.276
H   22  13.673   32.71  -47.14 542.97 409.24 -81.92  -99.16  0.294
H   23  10.942   34.00  -46.68 577.57 409.42 -72.82  -98.04  0.292
H   24  10.885   34.03  -51.61 581.38 409.37 -66.80  -98.05  0.289
H   25   7.662   22.95  -10.58 571.20 426.79 -60.46  -98.95  0.287
H   26  10.162   18.86   -5.55 579.57 402.80 -82.29  -99.90  0.298
H   27   9.732   16.47  -33.57 599.43 401.60 -79.44 -100.32  0.300
H   28  11.975   21.45  -32.76 566.77 417.38 -57.05 -101.86  0.281
H   29  13.743   41.93  -48.27 644.62 489.50 -72.02 -107.49  0.247
H   30  13.682   45.21  -56.14 646.00 489.50 -74.29 -107.59  0.241
H   31  13.500   46.19  -57.18 650.47 485.39 -75.74 -108.18  0.245
H   32   9.315   44.86  -58.90 645.14 482.48 -76.42 -102.94  0.257
H   33   8.484   22.21  -25.81 541.67 423.16 -60.75  -99.13  0.292
H   34  13.279   25.07  -43.37 549.74 410.21 -81.02  -98.68  0.288
H   35  10.956   27.18  -52.92 580.80 409.72 -70.25  -98.01  0.289
H   36  11.026   27.05  -43.05 578.46 409.22 -70.51  -98.50  0.292
H   37  13.754   13.23  -43.36 538.24 398.03 -94.53 -110.24  0.291
H   38  12.306   12.87  -50.88 578.07 394.39 -78.57 -110.79  0.293
H   39  12.701   12.48  -40.29 567.09 394.23 -80.45 -111.02  0.292

Total   298.191

<> ParaSurf used   1.72 seconds CPU timetime
```

Figure 11: continued

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

3.4.3 For a job with Shannon entropy

Figures **12** and **13** show the relevant sections of the output for a calculation using the options surf=cube for trimethoprim with the extra *shannon* option, which requests internal and external Shannon entropies using the default *bins.txt* statistical background file from the PARASURF_ROOT directory. The output is identical to that shown in **Figure 11** except that an additional Shannon entropy block is printed after the descriptors, as shown in **Figure 12**:

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

Figure 12: Shannon entropy section of the ParaSurf® output for trimethoprim, **1**, using a marching-cube isodensity surface.

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

The Shannon entropy is also analyzed based on the surfaces assigned to the individual atoms to give the table shown in **Figure 13**:

```
    Shannon-entropy analysis :


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

 <> ParaSurf used    7.96 seconds CPU time
```

Figure 13: Shannon entropy analysis from the ParaSurf® output for trimethoprim, **1**, using a marching-cube isodensity surface.

3.4.4 For a job with autocorrelation similarity

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

parasurf captopril surf=cube autocorr

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

parasurf trimethoprim surf=cube autocorr=captopril_p.sdf

This leads to the following additional output from ParaSurf™:

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

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

   Total      : 0.8924    0.5535     0.6968     0.6675     0.1519     0.8364

   1. Quartal : 0.9039    0.3785     0.6334     0.7132     0.2781     0.8695
   2. Quartal : 0.8861    0.3713     0.6761     0.5637    -0.0821     0.8366
   3. Quartal : 0.9348    0.7294     0.8152     0.6597     0.0905     0.9109
   4. Quartal : 0.8450    0.7349     0.6623     0.7334     0.3211     0.7286


 Corr.Coeff.  : Shape      +/+        -/-        +/-        IE(L)      EA(L)

 (R)          : 0.7914   -0.7041     0.5585     0.9201     0.0016     0.58805
```

Figure 14: Similarity output using autocorrelation functions. The lead molecule is captopril, which is defined in captopril_p.sdf using the SMILES string.

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

3.5 ParaSurf™ SDF-output

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

<ParaSurf OPTIONS>

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

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

Where the individual variables can be:

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

`<MOLECULAR_CENTERS>`

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

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

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

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

`<SPHERICAL_HARMONIC_......>`

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

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

`Order = nn`	("Order = ",i4)
$l (C_l^m) m = -l$ `to l`	(I5, 10f8.4/5x,10f8.4/5x,10f8.4/5x,10f8.4) (One set of coefficients each for l = 1 to 15)
`RMSDs:` $l, RMSD^1, RMSD^2$	("RMSDs:") (i8, 2f12.8) (One line for each l for l = 1 to 15, where $RMSD^1$ is the area-weighted RMSD and $RMSD^2$ the simple RMSD)

There are five such blocks, indicated by the tags:

`<SPHERICAL_HARMONIC_SURFACE>`   The fitted molecular surface (radial distances) in Ångstrom

`<SPHERICAL_HARMONIC_MEP>`   The MEP values at the spherical-harmonic surface (l = 20) in kcal mol^{-1}

`<SPHERICAL_HARMONIC_IE(l)>`   The IE$_L$ values at the spherical-harmonic surface (l = 20) in kcal mol^{-1}

`<SPHERICAL_HARMONIC_EA(l)>`   The EA$_L$ values at the spherical-harmonic surface (l = 20) in kcal mol^{-1}

`<SPHERICAL_HARMONIC_ALPHA(l)>`   The α_L values at the spherical-harmonic surface (l = 20) in kcal mol^{-1}

`<ParaSurf Descriptors>`

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

Molecular:	μ, μ_D, α, MW, G, A, VOL
	`("Molecular ",5f10.4,2f10.2)`
MEP:	V_{max} , V_{min} , $\overline{V}_+$, $\overline{V}_-$, $\overline{V}$, ΔV , σ_+^2 , σ_-^2 , σ_{Tot}^2 , ν , $\sigma_{tot}^2 \nu$, 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 , $\sigma_{IE}^2 , \gamma_1^{IE} , \gamma_2^{IE} , \int_{IE}$
	`("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}^+ , $A_{EA}^+ , \gamma_1^{EA} , \gamma_2^{EA} , \int_{EA}$
	`("EA(l)      ",7f10.2/2f10.2,2f10.4,f10.2,2f10.4/12x,g12.6)`
Eneg(l):	χ_L^{max} , χ_L^{min} , $\overline{\chi_L}$, $\Delta \chi_L$, $\sigma_\chi^2 , \gamma_1^\chi , \gamma_2^\chi , \int_\chi$
	`("Eneg(l)    ",5f10.2,2f10.4/12x,g12.6)`
Hard(l):	η_L^{max} , η_L^{min} , $\overline{\eta_L}$, $\Delta \eta_L$, $\sigma_\eta^2 , \gamma_1^\eta , \gamma_2^\eta , \int_\eta$
	`("Hard(l)    ",5f10.2,2f10.4/12x,g12.6)`
Alpha(l):	α_L^{max} , α_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)`

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)`

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

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

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

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

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

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

3.5.1 Optional blocks in the SDF-output file

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

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

"Shannon(i)"	(4f10.4,f10.2,f10.4)

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

"Shannon(e)"	(4f10.4,f10.2,f10.4)

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

\<ATOMIC SURFACE MAXIMUM H (internal)\>

Maximum "internal" Shannon entropies	(10f8.4)

\<ATOMIC SURFACE MINIMUM H (internal)\>

Minimum "internal" Shannon entropies	(10f8.4)

\<ATOMIC SURFACE MEAN H (internal)\>

Mean "internal" Shannon entropies	(10f8.4)

\<ATOMIC SURFACE TOTAL H (internal)\>

Total "internal" Shannon entropies	(10f8.4)

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

\<ATOMIC SURFACE MAXIMUM H (external)\>

Maximum "external" Shannon entropies	(10f8.4)

\<ATOMIC SURFACE MINIMUM H (external)\>

Minimum "external" Shannon entropies	(10f8.4)

\<ATOMIC SURFACE MEAN H (external)\>

Mean "external" Shannon entropies	(10f8.4)

\<ATOMIC SURFACE TOTAL H (external)\>

Total "external" Shannon entropies	(10f8.4)

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

\<SURFACE AUTOCORRELATION PARAMETERS\>

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

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

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

3.6 The surface (.psf) file

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

Number of atoms	(i6)

One line per atom with the atomic surface properties:

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

Number of surface points	(i6)

One line per point with the local properties:

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

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

Number of surface triangles	(i6)

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

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

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

3.7 Anonymous SD (.asd) files

The .asd file contains only those blocks from the ParaSurf™ output SD file that do not pertain directly to the 2D-molecular structure. Its purpose is to allow a full descriptions of the intermolecular bonding

properties of the molecule without revealing its structure. The .asd file can only be written from a ParaSurf™ calculation using spherical-harmonic fitting. Its form is:

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

And the blocks defined by the following tags:

`<SPHERICAL_HARMONIC_SURFACE>`

`<SPHERICAL_HARMONIC_MEP>`

`<SPHERICAL_HARMONIC_IE(l)`

`<SPHERICAL_HARMONIC_EA(l)`

`<SPHERICAL_HARMONIC_ALPHA(l)>`

`<SHAPE HYBRIDS>`

`<MEP HYBRIDS>`

`<IE(L) HYBRIDS>`

`<EA(L) HYBRIDS>`
`<ALPHA(L) HYBRIDS>`

`<ParaSurf Descriptors>`

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

Molecular	μ, μ_D, α, 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)	$IE_L^{\max}$, $IE_L^{\min}$, 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)	$EA_L^{\max}$, $EA_L^{\min}$, EA_{L+}, EA_{L-}, 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)	$\chi_L^{\max}$, $\chi_L^{\min}$, χ_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)	$\eta_L^{\max}$, $\eta_L^{\min}$, η_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)	$\alpha_L^{\max}$, $\alpha_L^{\min}$, α_L, $\Delta\alpha_L$, σ_α^2, γ_1^α, γ_2^α, $\int_\alpha$ ("Alpha(l) ",5f10.2,2f10.4/12x,g12.6)

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

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

3.7.1 Optional blocks

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

\<SURFACE AUTOCORRELATION PARAMETERS\>

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

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

\<SHAPE AUTOCORRELATION\> (10f8.1)

\<PLUS-PLUS MEP AUTOCORRELATION\> (10f8.4)

\<MINUS-MINUS MEP AUTOCORRELATION\> (10f8.4)

\<PLUS-MINUS MEP AUTOCORRELATION\> (10f8.4)

3.8 Grid calculations with ParaSurf™

3.8.1 User-specified Grid

The command

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

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

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

Figure 15: Sample grid file

Gives the output shown in **Figure 16**.

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

 <> Program options :

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

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

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

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

 <> ParaSurf used    0.05 seconds CPU time
```

Figure 16: Sample grid outputfile

The name and the extension (if any) of the grid file are free. Only the output file is written. The units of the local properties are those used in the normal output (i.e. V, IEL, and EAL in kcal mol-1, αL in Ångstrom3.

3.8.2 Automatic grids

ParaSurf'07[TM] can generate grids automatically for lead compounds in ComFA[®]-like procedures. The **grid=auto** option generates a grid around the molecule (with a 4 Å margin around the positions of the atoms in each direction) and includes all points for which the electron density is lower than 10^{-2} (i.e. for points outside the molecule). The spacing of the grid is set to a default value of 1.0 Å, but can be set to any value up to a maximum of 2.0 Å by the command-line argument **lattice=*n.n***, which sets the lattice spacing to *n.n* Å. The grid thus generated is output (with the values of the local properties analogously to a calculation that uses an predefined grid and can be used for other molecules that have been aligned with the lead.

3.9 The SIM file format

SIM files must reside in the ParaSurf™ executable directory and are strictly fixed format. SIM files must be called **<filename>.sim**, where **<filename>** must have exactly three characters. A sample SIM file for a single model (the free energy of solvation in octabol) is shown in **Figure 17**:

```
> <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 17: 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:

$$\left[MEP^{m} \cdot IE_{L}^{n} \cdot EA_{L}^{o} \cdot \alpha_{L}^{p} \cdot \eta_{L}^{q} \right]^{r} \text{ if } \textbf{\textit{Abs}} \text{ is \textbf{false} and } \left[\left| MEP^{m} \cdot IE_{L}^{n} \cdot EA_{L}^{o} \cdot \alpha_{L}^{p} \cdot \eta_{L}^{q} \right| \right]^{r} \text{ if } \textbf{\textit{Abs}} \text{ is \textbf{true}.}$$

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

3.10 Output tables

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

Table 5

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

Column Header	Symbol[a]	Descriptor
MEPvar+	σ_+^2	Total variance in the positive MEP values
MEPvar-	σ_-^2	Total variance in the negative MEP values
MEPvartot	σ_{tot}^2	Total variance in the MEP
MEPbalance	ν	MEP balance parameter
var*balance	$\sigma_{tot}^2 \nu$	Product of the total variance in the MEP and the balance parameter
MEPskew	γ_1^V	Skewness of the distribution of the MEP
MEPkurt	γ_2^V	Kurtosis of the distribution of the MEP
MEPint	$\int_V$	Integral of the MEP*area over the surface
IELmax	IE_L^{max}	Maximum value of the local ionization energy
IELmin	IE_L^{min}	Minimum value of the local ionization energy
IELbar	$\overline{IE_L}$	Mean value of the local ionization energy
IELrange	ΔIE_L	Range of the local ionization energy
IELvar	σ_{IE}^2	Variance in the local ionization energy
IELskew	γ_1^{IE}	Skewness of the distribution of IE(L)
IELkurt	γ_2^{IE}	Kurtosis of the distribution of IE(L)
IELint	$\int_{IE}$	Integral of the IE(L)*area over the surface
EALmax	EA_L^{max}	Maximum of the local electron affinity
EALmin	EA_L^{min}	Minimum of the local electron affinity
EALbar+	$\overline{EA_{L+}}$	Mean of the positive values of the local electron affinity
EALbar-	$\overline{EA_{L-}}$	Mean of the negative values of the local electron affinity
EALbar	$\overline{EA_L}$	Mean value of the local electron affinity
EALrange	ΔEA_L	Range of the local electron affinity
EALvar+	σ_{EA+}^2	Variance in the local electron affinity for all positive values
EALvar-	σ_{EA-}^2	Variance in the local electron affinity for all negative values
EALvartot	σ_{EAtot}^2	Sum of the positive and negative variances in the local electron affinity
EALbalance	ν_{EA}	Local electron affinity balance parameter
EALfraction+	δA_{EA}^+	Fraction of the surface area with positive local electron affinity
EALarea+	A_{EA}^+	Surface area with positive local electron affinity
EALskew	γ_1^{EA}	Skewness of the distribution of the MEP
EALkurt	γ_2^{EA}	Kurtosis of the distribution of the MEP
EALint	$\int_{EA}$	Integral of the MEP*area over the surface
ENEGbar	$\overline{\chi_L}$	Mean value of the local electronegativity

Column Header	Symbol[a]	Descriptor
ENEGmax	χ_L^{max}	Maximum of the local electronegativity
ENEGmin	χ_L^{min}	Minimum of the local electronegativity
ENEGbar	$\overline{\chi_L}$	Mean value of the local electronegativity
ENEGrange	$\Delta\chi_L$	Range of the local electronegativity
ENEGvar	σ_χ^2	Variance in the local electronegativity
ENEGskew	γ_1^χ	Skewness of the distribution of the local electronegativity
ENEGkurt	γ_2^χ	Kurtosis of the distribution of the local electronegativity
ENEGint	$\int_\chi$	Integral of the χ(L)*area over the surface
HARDmax	η_L^{max}	Maximum of the local electronegativity
HARDmin	η_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
POLmax	α_L^{max}	Maximum value of the local polarizability
POLmin	α_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

[a]Symbols as used **1.9**.

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

Column Header	Symbol	Descriptor
SHANImax	H_{in}^{max}	Maximum internal Shannon entropy
SHANImin	H_{in}^{min}	Minimum internal Shannon entropy

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

3.11 Autocorrelation similarity tables

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

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

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

3.12 Shared files

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

4 TIPS FOR USING PARASURF´07

4.1 Choice of surface

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

4.2 ParaSurf™ and ParaFit™

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

4.3 QSAR using grids

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

5 SUPPORT

5.1 Contact

Questions regarding ParaSurf™ should be sent directly to:

support@ceposinsilico.com

5.2 Error reporting

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

5.3 CEPOS Insilico Ltd.

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

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

www.ceposinsilico.com/contact

6 REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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