DOI: 10.1002/cmdc.200700262

The Calculation of Polar Surface Area from First Principles: An Application of Quantum Chemical Topology to Drug Design

Ian Bytheway,*[a] Michael G. Darley,[b] and Paul L. A. Popelier*[b]

The calculation of polar surface areas (PSA) from the electron density using quantum chemical topology (QCT) and a newly developed algorithm to determine isodensity surface areas is described. PSA values were calculated from the atomic partitioning of B3LYP/6-311G wavefunctions and the results described herein represent the first application of this new algorithm. PSA values were calculated for forty drugs and compared to the topological polar surface area (TPSA) and those calculated by the QikProp program. Oral bioavailabilities predicted from the QCT PSA values for a subset of twenty drugs (the Palm set) were similar to those predicted by the dynamic polar surface area (DPSA) and in general, are in agreement with the observed values. Overall, PSA values obtained from QCT were generally similar to the DPSA, TPSA, and QikProp values, though differences in fragment contributions were found, with nitrogen-bearing functional groups showing the largest variation between methods. Differences between methods showed how the calculation of the PSA is dependent on the method used and, therefore, judicious application of the upper limits used in the prediction of oral bioavailability is warranted. These results also indicate that, because of the differences in the way PSA values are calculated, values from the different methods should not be used interchangeably.*

Introduction

The savings in time and cost to a drug discovery campaign from the early detection of lead molecules with problematic pharmacology is too important to ignore.[1–4] Of the physical properties that might be calculated, the polar surface area (PSA) provides a useful descriptor of bioavailability, an important consideration for progression of a candidate through the drug development process. In this work we consider the calculation of polar surface areas from a rigorous quantum mechanical viewpoint and compare the results with those obtained from fragment-based and semiempirical methods to examine how these methods compare, and how improvements might be made.

The correlation between the PSA and intestinal absorption was initially established by comparing dynamic PSA (DPSA) values, derived from Boltzmann-averaged ensembles of low-energy molecular conformations, with cell permeabilities and intestinal absorption.[5,6] The use of PSA values derived for single, low-energy conformers also correlated with effective intestinal permeability[7,8] and penetration of the blood brain barrier.[9] A fragment-based methodology which derived standardized contributions from functional groups and atom types[10] provides an efficient method to the calculation of molecular PSA values. Correlation with experiment increases only slightly when conformation-averaged PSA values were used over those derived from single conformers.

Examination of the quantitative relationship between the PSA and oral bioavailability suggested an upper limit of 140 Å^2 for reasonable bioavailability in the rat.[11] This limit is influenced by the subset of molecules used to correlate the PSA with experimental data,[12] the method of calculation employed and the constituent atoms included in the PSA total. Nevertheless, the development of a bioavailability score that combines the calculated PSA and the formal charge of a molecule provides an estimate of bioavailability that can be readily applied to drug design and discovery,[13] although the applicability of metrics such as the PSA to the prediction of bioavailability has been questioned.[14]

Different approaches to the calculation and partitioning of surface properties have been attempted to understand and harness their apparent predictivity. Deconstruction of the PSA into readily interpreted physicochemical properties (for example, hydrogen bond counts) provided a picture of their relative importance and provided a ready route to PSA estimation.[15] Extension of this model based on partitioning of the total molecular surface area provided an accurate description of membrane permeability, with accuracy comparable to descriptors derived from low level quantum chemical calculations.[15] Partitioning and scaling of the total molecular surface area, including atoms (halogen and aromatic C) not conventionally incorporated into PSA determinations, was also used to successfully predict octanol-water, chloroform-water, and cyclohexane-water partition coefficients.[16]

[a] *Dr. I. Bytheway*
Drug Design Group, Progen Pharmaceuticals Ltd.
PO Box 2403, Toowong 4066 Queensland (Australia)
Fax: (+61) 7-3375-1168
E-mail: ianb@progen-pharma.com

[b] *M. G. Darley, Dr. P. L. A. Popelier*
Manchester Interdisciplinary Biocentre, University of Manchester
131 Princess Street, Manchester M1 7DN (United Kingdom)
Fax: (+44) 161-306-5201
E-mail: paul.popelier@manchester.ac.uk

Prediction of human intestinal absorption rates using the distribution coefficient (log *D*) and PSA as descriptors further highlights the utility of the PSA in the drug discovery process.[17] The calculation of characteristics such as intestinal absorption, solubility, and permeability are important in the early stages of the drug design and discovery process. Molecular surface properties provide useful indicators of these characteristics[18] and suggest how an in silico biopharmaceutical classification scheme for the selection of drug candidates might be formulated.[19]

This work is concerned with the calculation of PSA values rather than their intrinsic relationships with measures of bioavailability. That is to say, given the importance of this molecular property to the estimation of bioavailability, how well do the largely (semi) empirical methods of calculating PSA compare with more accurate ab initio methods? The use of electronic structure methods in drug design is becoming more prominent[20–25] although use in high throughput methodologies is still not feasible. Nevertheless ab initio methods are well suited to, and computational resources are capable of, probing the properties of druglike molecules related to their activity and bioavailability. In addition to the calculation of all-electron wavefunctions the partitioning of molecules into constituent atoms using quantum chemical topology[26,27] is applied to the calculation of the PSA. This approach provides PSA values indicative of their chemical environment determined from the molecular charge density, which in turn may be used to validate or update the library values commonly used in PSA determination.

Table 1. PSA values (Å^2) calculated by the different methods for all of the molecules included in this study.

	QCT[a]	QCT	QikProp[a]	QikProp	TPSA
Acetaminophen	53.17	48.66	64.38	49.60	49.33
Alprenolol	43.70	43.70	38.17	38.16	41.90
Ascorbic	112.37	107.74	127.28	109.60	107.22
Aspirin	74.44	64.20	86.41	55.02	63.60
Atenolol	88.61	82.54	93.47	78.76	84.60
AZT	130.79	120.87	151.41	123.79	134.08
Cimetidine	94.94	88.36	94.62	83.83	84.60
Ciprofloxacin	96.88	83.63	97.61	69.20	74.60
Diazepam	67.82	49.57	47.07	27.63	32.70
Fenoprofen	50.54	45.94	56.40	42.36	46.53
Fluoxetine	26.22	26.22	21.14	21.14	21.26
Foscarnet	97.64	90.69	117.83	100.38	94.80
Ibuprofen	39.10	34.62	48.88	34.66	37.30
Isoniazid	68.79	68.79	81.46	68.49	68.01
Ketoprofen	62.68	51.64	75.04	50.66	54.37
Levodopa	95.59	91.04	115.88	102.21	103.78
Mannitol	109.62	109.62	123.11	123.11	121.40
Mercaptopurine	59.47	59.47	54.67	54.67	54.47
Methadone	34.55	26.13	23.27	17.39	20.31
Metolazone	96.47	91.32	103.64	93.10	92.50
Metoprolol	55.61	55.61	47.99	47.99	50.70
Naproxen	50.01	46.51	57.15	42.93	46.53
Nisoxetine	38.85	38.85	29.34	29.34	30.50
Nordiazepam	59.10	45.57	58.55	36.83	41.50
Norfluoxetine	33.54	33.54	33.29	33.29	35.26
Olsalazine	133.54	123.45	163.40	129.76	139.80
Oseltamivir	90.83	78.18	92.45	68.14	78.63
Oxazepam	75.44	64.02	78.30	57.26	61.70
Oxprenolol	57.46	57.46	45.11	45.11	50.70
Phenazone	39.81	34.92	35.20	21.41	23.55
Phenobarbitol	91.51	78.29	99.61	68.57	75.27
Pindolol	58.85	58.85	53.34	53.34	57.30
Practolol	79.41	74.45	81.63	81.63	70.60
Prontosil	119.28	119.28	139.45	139.45	136.94
Salbutamol	76.46	76.46	74.00	74.00	72.70
Sertraline	16.28	16.28	12.73	12.73	12.03
Sulfasalazine	137.76	132.44	155.72	138.90	141.30
Sulpiride	112.18	105.68	111.68	100.57	101.70
Tranexamic Acid	65.44	59.48	72.91	59.21	63.30
Triclosan	31.23	31.23	28.87	28.87	29.46

[a] Contributions from carbonyl and nitrile C atoms are included in the PSA total.

Theoretical Background and Computational Methods

A low-energy conformation for each of the molecules in Table 1 was obtained from 1000 steps of Monte Carlo multiple minimum (MCMM) searching[28] with all energy minimizations converged to 0.05 kcal mol^{-1} Å^{-1}. The molecular mechanics calculations were performed with the OPLS-AA forcefield[29] and incorporated solvation in water through the use of the GB/SA implicit solvation model[30] as implemented in the MacroModel program (version 9).[31] The low energy conformers obtained from MCMM searching were subsequently used in the calculation of PSA values with QikProp (version 2.5 and 3.0β)[32] and as starting points for the wavefunction calculations.

Wavefunctions for each molecule were calculated at the B3LYP/6-31G**//B3LYP/6-31G** level of theory[33] using the Jaguar program.[34] The final OPLS-AA/MCMM geometry for each molecule was used as the starting point for the corresponding B3LYP/6-31G** geometry optimizations which were converged to the default convergence parameters (maximum element of gradient = 4.5×10^{-4} au; rms of gradient elements = 3.0×10^{-4} au; maximum element of nuclear displacement = 1.8×10^{-3} au; rms of nuclear displacement elements = 1.2×10^{-3} au; difference between final energies from previous and current geometry optimization iterations = 5.0×10^{-5} au). Single-point energy calculations were performed with the optimized geometries by the program GAUSSIAN03[35] to generate the wavefunction files required for QCT analyses.

QCT proposes that topological analysis of the total electron density provides sufficient information necessary to describe the atoms and bonds within a molecule.[26,36,37] Topological atoms are finite portions of three-dimensional space each associated with an atomic volume. When occurring in free (that is, isolated or single) molecules, atoms are capped by constant electron density contours at the molecule's exterior and by interatomic surfaces inside the molecule. Interatomic surfaces consist of bundles of gradient paths, that is, paths of steepest ascent through the electron density, which originate at infinity and terminate at a critical point (where the gradient vanishes) somewhere in between the nuclei. Within QCT a molecule consists of non-overlapping atoms which together form, without gaps, the molecule from which they stem. QCT partitioning as-

signs each point in space to an atom, hence each point on a constant electron density contour capping the molecule also belongs to an atom. The electron density thus becomes partitioned into well-defined atomic surface regions.

Definition of atoms using capping isodensity values is not straightforward as the interatomic surfaces increase in complexity as the contour used to describe the outer boundary of the atom is decreased. This arises from two general causes:

1) Difficulty in completing the surfaces of ring atoms. These may be conventional heterocyclic rings or arise from the formation of intermolecular bonds. The plane formed by the ring atoms, the ring plane, must be determined. Once the ring plane is known a line perpendicular to it, and passing through the ring critical point, is used to form the boundary of the interatomic surfaces between the ring atoms. The calculation of a ring plane is more difficult when the ring is not planar and in extreme cases fails.[38]

2) Complicated surface areas can appear when two critical points are close together. This can affect nearby atomic surfaces and result in surfaces with high curvature. The gradient paths of surfaces with high curvature can diverge rapidly and the tracing of points onto the path can fail. In both cases, failure to trace the ring plane or gradient path, triangulation of the surface may be incomplete.

An example of both cases occurs in prontosil where interatomic surfaces are complicated by the formation of a bond critical point between an azo nitrogen and a hydrogen atom of the ortho amine group, that is, an intramolecular hydrogen bond is formed. This results in the formation of a set of topologically connected atoms in a closed ring and the formation of a ring critical point. This ring critical point is formed in close proximity to the bond critical point, as shown in Figure 1, which leads to the formation of a complex interatomic surface. In such situations the interatomic surfaces are difficult, and for some of the molecules considered in this study, impossible to calculate as the capping isodensity values are decreased.

The increase in interatomic surface complexity with decreasing capping isodensity is shown pictorially for prontosil in Figure 2a. In this Figure the interatomic surfaces obtained with the 10^{-2} au and 10^{-6} au capping isodensities for the azo and amino N atoms are superimposed. The surface complexity in the region of the intramolecular hydrogen bond is shown, as is the change in surface area as the capping isodensity contour is decreased. The corresponding van der Waals surfaces of these atoms, generated using the Maestro graphical user interface (version 7.5),[39] are shown in Figure 2b for comparison.

In light of this surface complexity at lower capping isodensities only the 10^{-2} au contour was chosen for the purposes of the calculations described here. Preliminary calculations indicated that the surfaces defined by this isodensity were calculable across all molecules, whereas the surfaces with lower isodensity contours were not. Whereas lower capping isodensities might seem preferable, analysis of the atomic surface areas for the atoms of fluoxetane (see Table 2) showed that surface areas of heavy atoms at the 10^{-2} au capping isodensity are approximately 98% of the surface area at the 10^{-6} au capping isodensity. The comparative surface area of H atoms at the

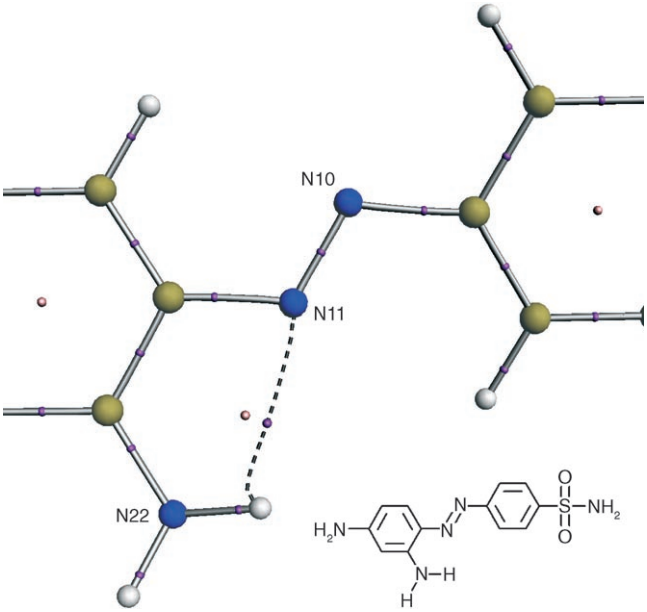


Figure 1. The atomic graph of prontosil in the region of the azo group (atoms N10 and N11) and a nearby amino group (atom N22) showing the close proximity of the bond and ring critical points that complicate the calculation of interatomic surfaces. Bond paths which connect the nuclei (colored spheres) are shown as gray lines and a bond path denoting an intramolecular hydrogen bond connecting an azo nitrogen with a hydrogen of the ortho amine is shown as a black dashed line. Bond critical points are shown as purple spheres, and ring critical points as pink spheres.

10^{-2} au isodensity is lower, approximately 89% of the 10^{-6} au surfaces. Importantly, the H atoms make smaller contributions to the PSA than the heavy atoms so the lower fraction of surface coverage is less important.

The acceptability of the 10^{-2} au capping isodensity in this study is also confirmed by examining the total electron count obtained from integration of the atomic basins with the different capping isodensities. With the 10^{-2} au capping isodensity the total integrated charge of the molecule is 97% (that is, 157.8 e of the 162 e in total), which increased to 99% or better for the smaller capping isodensities. These results are in agreement with earlier calculations that showed the 10^{-3} au contour encompasses 96.0 to 96.6% of the electron density of hydrogen atoms in hydrocarbons,[40] and over 99% of the electronic charge of the atoms carbon to neon.[36] Thus the choice of the 10^{-2} au capping isodensity allows for surface definition without significant loss of resolution and is sufficient for the calculation of QCT PSA values.

The radius of the QCT surface used to evaluate the PSA values was calculated by averaging the distances from the nucleus to points on the isodensity contour capping the atomic basin. Approximately 1500 points typically defined the isodensity surface. Averaged radii obtained in this way are given in Table 4.

Solution of the problems associated with the calculating interatomic surface areas required the development of a new algorithm to allow the application of QCT to the calculation of PSA. This algorithm constructs atoms by means of a finite element meshing algorithm,[41] which enables the calculation of

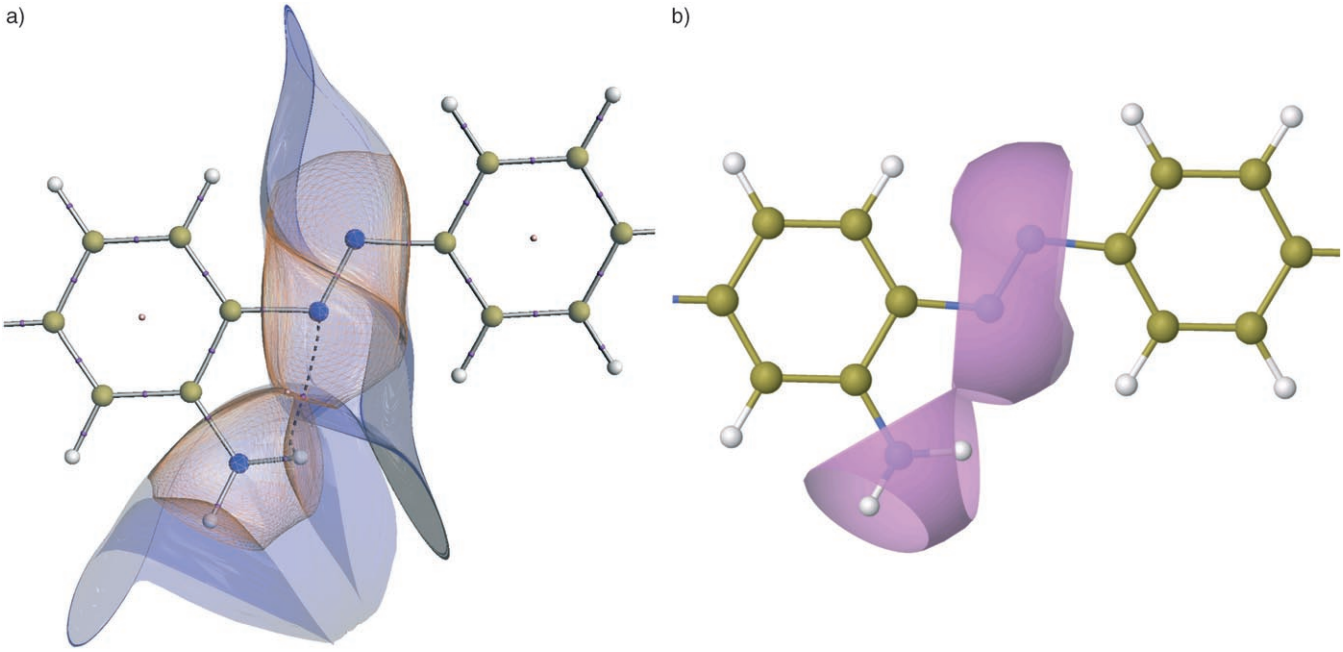


Figure 2. Plots of the QCT and van der Waals surfaces of prontosil in the region of the azo and nearby amino N atoms. a) A plot showing the increased interatomic surface complexity that occurs when the capping isodensity is decreased. Superimposed plots of the total electron density at the 10^{-2} au (orange wireframe) and 10^{-6} au (blue transparent) capping isodensities are shown. Bond paths, and bond and ring critical points are depicted as in Figure 1. b) The van der Waals surfaces for the same atoms.

Table 2. Analysis of the total charge density of fluoxetine.[a]

	% Coverage of the 10^{-6} (au) Capping Isodensity Isodensity Contour (au)			
	10^{-2}	10^{-3}	10^{-4}	10^{-5}
C	98.43	99.89	100.00	100.01
H	88.91	98.33	99.86	99.98
N	98.31	99.83	99.99	100.01
O	98.41	99.90	100.00	100.00
F	98.86	99.83	99.98	100.00
% Total Charge	97.43	99.70	99.98	100.0

[a] The average atomic surface areas contained within the atomic basins capped by the different isodensity contours expressed as a percentage of the 10^{-6} au surface area. The percent total charge was obtained by comparing the summed electron populations obtained using the different isodensity contours with the total electron population for fluoxetine.

surface areas of each atom as defined by the topology of the charge density, from which the PSA could be derived. Refinements and extensions to this algorithm were needed to tackle atoms involved in heterocyclic rings. Initially only molecules with polar atoms involved in noncyclic bonding were chosen, with the set expanded once the algorithmic issues were resolved. All of the QCT surface area calculations were performed with a local version of the program MORPHY.[42]

The calculated PSA values are collected in Table 1. The QCT PSA values obtained from summation of the appropriate surface area contributions obtained from the QCT analysis are compared with those calculated with QikProp PSA and the topological PSA (TPSA[10,43]). It is important to note that the standard PSA definition employed by QikProp includes surface area contributions from carbonyl and nitrile C atoms. PSA values for both the QCT and QikProp methods were, therefore, calculated both with and without the contributions of these

Table 3. Averaged PSA values (Å^2) of fragments derived from the QCT and QikProp calculations for the set of molecules studied herein compared with the TPSA fragment values.[10]

Fragment	QCT	QikProp	TPSA	Fragment	QCT	QikProp	TPSA
[N](–*)(–*)–*	8.69	2.82	3.24	[nH](:*):*	14.46	13.99	15.79
[N](–*)=*	12.71	11.32	12.36	[O](–*)–*	11.44	6.97	9.23
[NH](–*)–*	14.77	11.47	12.03	[O]=*	16.67	16.57	17.07
[NH_2]–*	21.46	25.91	26.02	[OH]–*	18.39	20.21	20.23
[N](=*)*	9.22	10.54	13.60	[S](–*)(–*)(=*)=*	5.24	4.92	8.38
[N]#*	15.51	25.70	23.79	[P](–*)(–*)(–*)=*	4.60	8.15	9.81
[n](:*):*	12.64	13.16	12.89	[C](=O)(–*)–*	5.50	13.81	-
[n](–*)(:*):*	8.80	2.74	4.93	[C](=N)(–*)–*	8.18	8.31	-

atoms, whilst all TPSA values were calculated without any contributions from these C atoms.

Results and Discussion

QCT predicted oral bioavailabilities

A plot of the fraction absorbed after oral administration in humans (FA) versus QCT-derived PSA (Figure 3a) is similar to that of Palm et al.[6] for the same set of molecules (with the exclusion of lactulose and raffinose). The corresponding plot of observed versus QCT-predicted values also reflects the agreement between experiment and theory (Figure 3b). There are outliers, tranexamic acid and foscarnet, for which the QCT PSA values differ from the DPSA by approximately 10–20 Å^2 (Figure 3c). Foscarnet is also an outlier[10] when the TPSA (90.7 Å^2) and DPSA (115.3 Å^2[6]) values are compared. The QikProp PSA, 117.8 Å^2 is closer to the Palm et al. value, however, this is fortuitous as the QikProp value includes a contribution from the carbonyl C atom, which is ignored in the calculation of the DPSA. When the carbonyl C is not included in the QikProp PSA determination a value of 100.38 Å^2 was obtained, closer to the TPSA value.

Comparison of TPSA and QCT PSA

Agreement between these two methods of calculating PSA is generally excellent, which can be seen in Figure 4a. There are, however, a number of molecules that deviate from ideal agreement, which is a consequence of differences in the surface areas attributed to the different component groups.

The PSA attributed to the OH functional group is the source of discrepancy between the two methods for olsalazine and mannitol. The TPSA model assigns 20.23 Å^2 to this group whereas the QCT values range from 17-19 Å^2, which although a small difference, contributes significantly to the difference between the two models: approximately 16 Å^2 for olsalazine (9 Å^2 of this from the four OH groups) and 12 Å^2 for mannitol (from the six OH groups). A similar explanation also holds for the difference in PSA values for levodopa. Thus small differences in the PSA of the constituent functional groups can lead to a large overall systematic difference between the two methods.

A difference of approximately 14 Å^2 between the two methods for AZT is largely attributable to the different treatment of the azide group by the two methods. A PSA of 34.6 Å^2 was cal-

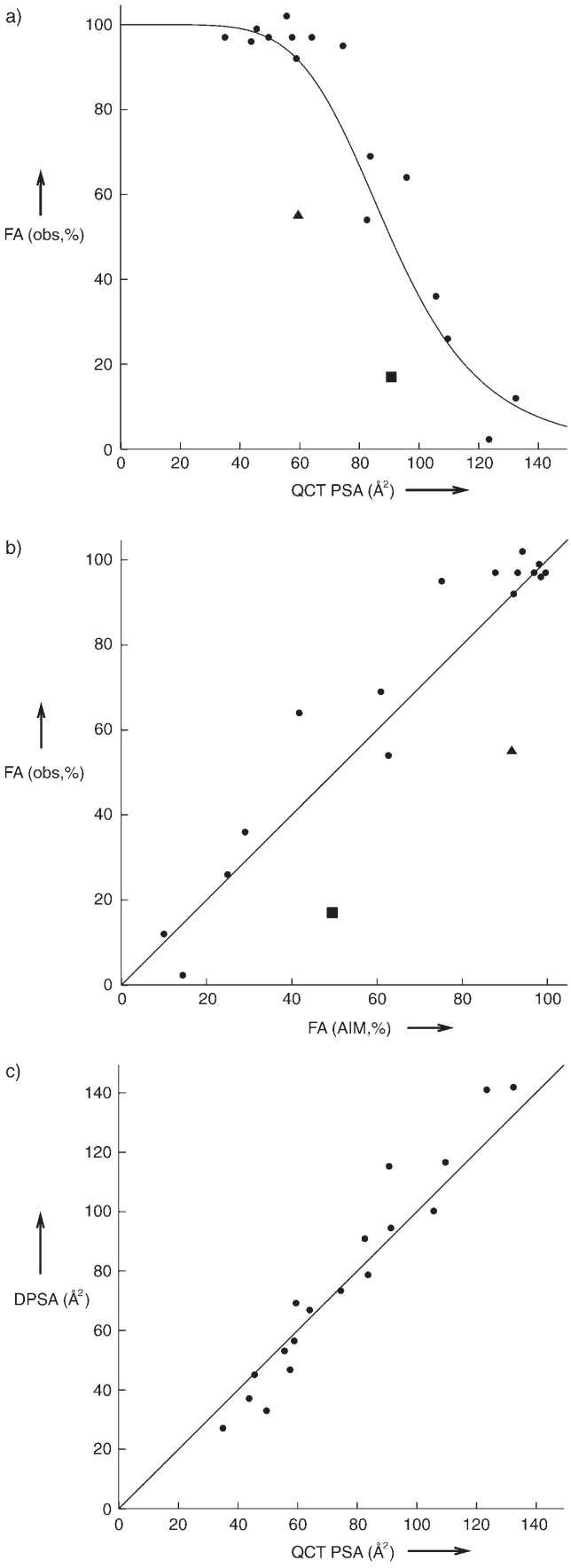
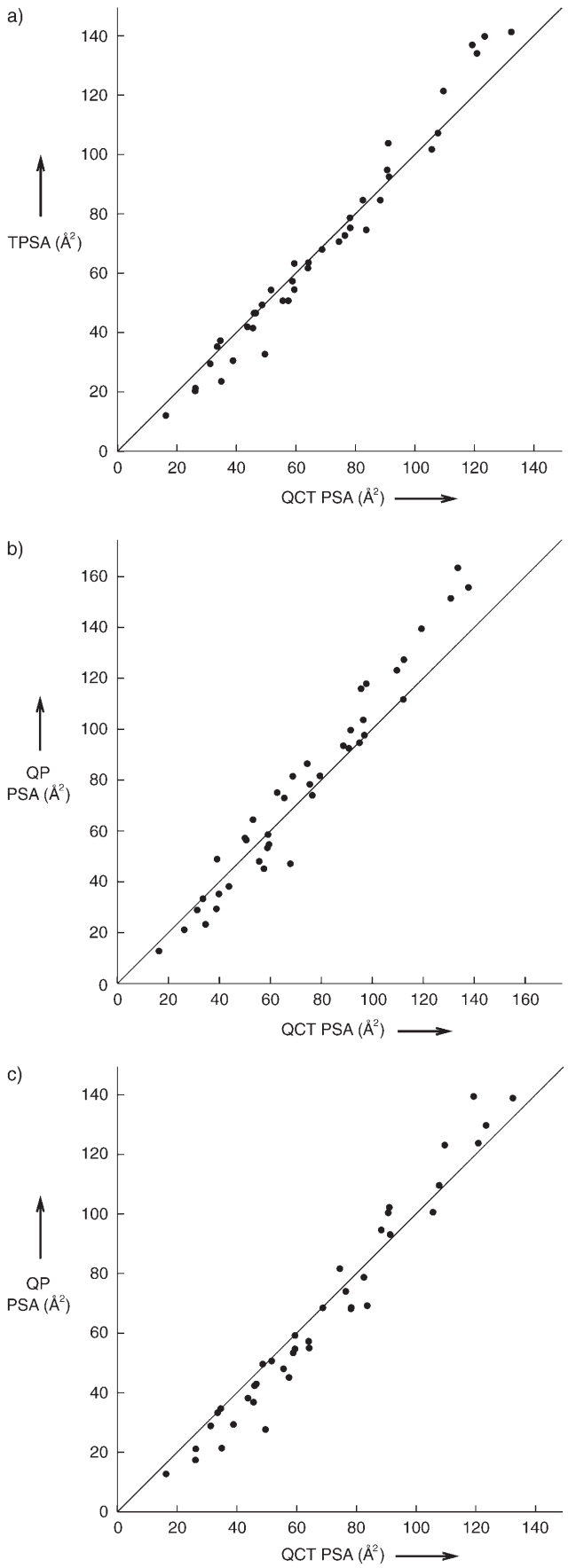


Figure 3. a) A plot of the observed fraction absorbed (FA) after oral administration against the QCT partitioned PSA. The sigmoidal curve % FA = 100/(1+(PSA/PSA$_{50\%}$)$^{\gamma}$) was fitted to the data and is also shown (PSA$_{50\%}$ = 90.40 Å^2, γ = 5.71). b) A plot of the observed FA against that predicted by the QCT PSA, R^2 = 0.82. The outliers discussed in the text, foscarnet and tranexamic acid, are depicted by the square and triangle, respectively. c) A plot of DPSA versus QCT PSA, R^2 = 0.95. FA and DPSA values are for the set of molecules discussed by Palm et al.[6] with the exception of lactulose and raffinose. QCT PSA values were not calculated for these molecules because of their size and complexity with respect to the calculation of ab initio wavefunctions.



culated for the azide group by the QCT method whereas the TPSA method estimates a value of 49.8 Å^2. These values can be further partitioned into their N component values, which shows that, using Lewis notation, the –N= values are similar (11.4 Å^2 and 12.4 Å^2, QCT and TPSA estimates, respectively), whereas the middle azide nitrogen atom =N$^+$= (9.2 Å^2 and 13.6 Å^2, QCT and TPSA estimates, respectively) and =N$^-$ (14.0 Å^2 and 23.8 Å^2, QCT and TPSA estimates, respectively) deviate more significantly. The QCT results, therefore, predict similar PSA values for these latter two N atoms, indicating that the formal charge separation invoked to describe the bonding is not reflected in the size of their PSA contribution.

Another source of discrepancy in the AZT comparison is the PSA attributed to the tertiary (ring) N atom of 8.7 Å^2 by the QCT method and 3.2 Å^2 by the TPSA method. The contribution of this fragment to the PSA of diazepam also differs significantly between the two methods: 11.7 Å^2 (QCT) versus 3.2 Å^2 (TPSA).

Two other molecules for which there are differences between the QCT and TPSA values are diazepam and phenazone. For diazepam the bulk of the difference in the PSA values are due to the contributions from the two nitrogen atoms. The tertiary N contributes 11.7 Å^2 to the QCT PSA but only 3.2 Å^2 to the TPSA, whereas the imino N contributes 17.5 Å^2 to the QCT PSA and 12.4 Å^2 to the TPSA. The difference for phenazone is due to the contributions from the two aromatic nitrogen atoms: 17.8 Å^2 and 9.8 Å^2 to the QCT PSA and TPSA, respectively.

Comparison of QikProp and QCT PSA

The QCT polar surface areas are strongly correlated with the corresponding QikProp values (see Figure 4b) although the latter tend to be larger even when carbonyl and nitrile C contributions are included in both methods of PSA estimation. This difference is attributable to the larger contribution of the carbonyl C to the QikProp PSA, averaging 13.9 Å^2 whereas the average QCT contribution is only 5.6 Å^2 for the same atom. Upon removal of the carbonyl carbons, the QikProp values are mostly similar to their QCT counterparts (Figure 4c).

There are some differences, however. The PSA values for mannitol differ by 14 Å^2, attributable to the accumulation of small differences in the PSA of the OH group. Ciprofloxacin provides an interesting case where removal of the carbonyl and nitrile C contributions results in worse agreement between the two methods. This is attributable to differences in the contributions of the secondary N atom (QikProp = 15.7 Å^2, QCT = 10.0 Å^2), the tertiary N atom (QikProp = 2.7 Å^2, QCT = 8.40 Å^2), and the aromatic ring N atom (QikProp = 2.5 Å^2, QCT = 8.60 Å^2). These differences, along with the minor differences in contri-

Figure 4. a) A plot of TPSA versus QCT PSA values: $R^2 = 0.97$. b) A plot of QikProp PSA versus QCT PSA calculations with surface area contributions from carbonyl C atoms included: $R^2 = 0.96$. c) A plot of QikProp PSA versus QCT PSA calculations with surface area contributions from carbonyl C atoms excluded: $R^2 = 0.96$. PSA values for all molecules listed in Table 1 are included in these plots.

butions from the carbonyl O and hydroxy groups, are amply masked by the different estimates of the PSA attributable to the two carbonyl C atoms (QikProp = 28.4 Å^2, QCT = 13.20 Å^2).

As noted when comparing the QCT PSA with the TPSA, diazepam and phenazone also show differences between the QCT and QikProp PSA values. For diazepam the bulk of the difference in the PSA values are due to the contributions from the two nitrogen atoms. The tertiary N contributes 11.7 Å^2 to the QCT PSA but only 1.9 Å^2 to the QikProp value, whereas the imino N contributes 17.5 Å^2 to the QCT PSA but only 9.8 Å^2 to the QikProp value. The difference for phenazone is similarly attributable to the contributions from the two aromatic nitrogen atoms: 17.8 Å^2 to the QCT PSA and only 4.6 Å^2 to the QikProp value.

Contributions from fragments

The discrepancy between methods is interesting for two reasons, the cancellation of differences and the PSA attributed to functional groups by the various methods. Thus for ciprofloxacin similar overall QCT and QikProp PSA values resulted from different contributions from each of the polar functional groups. With this in mind the QCT and QikProp PSA values for the set of molecules considered here have been extracted and averaged to understand how the different methods treat the different functional groups. Averaged QCT and QikProp PSA values per functional group are given in Table 3 along with the standard values used to calculate the TPSA.[10]

Averaged values are generally similar though notable differences do exist. The QCT average of 8.7 Å^2 for a tertiary N, SMILES string [N](–*)(–*)–*, is nearly three times the value ascribed to this functional group by the TPSA and QikProp methods, whereas the QCT average of 15.5 Å^2 for a triply bonded N, [N]#*, is approximately 10 Å^2 lower than the corresponding TPSA and QikProp values. The average contributions from the imine nitrogen, [N](–*)=*, are similar across all three methods, though it was the difference in this contribution that accounted for differences in the PSA values ascribed to diazepam. This is a potential problem with using averages and a small sample of molecules as PSA contributions from the N atoms in diazepam occurring at the upper end of the range for QCT and at the lower end of the range for QikProp.

The difference in PSA contributions from N atoms ascribed by the different methods may be examined more closely by considering the nature of the surfaces defined by the QCT method. Whereas the asphericity of the QCT interatomic surfaces for prontosil is clear in Figure 2, calculation of the distances from the N nuclei to the QCT interatomic surfaces, given in Table 4, provide a clear indication of the asymmetry of the surface. The average QCT radii for the two azo nuclei of approximately 1.35 Å are smaller than the 1.824 Å used by QikProp or 1.82 Å used by Stenberg et al.[15] Similarly, the average amino N radius of 1.4 Å is smaller than the radius used by QikProp (1.852 Å) whereas the PSA contribution calculated for this amino group by the QCT method is higher than either of the QikProp or TPSA values. This variation in radii used to define the atomic surface, and therefore its shape, is an important difference between the QCT and other methods.

Contributions from oxygen-bearing functional groups are generally more similar across methods. Hydroxy group contributions from the QCT analysis are slightly lower than the corresponding TPSA and QikProp values, as noted above for the case of mannitol, whereas doubly bonded O, [O]=*, are similar across all three methods.

The PSA contribution of the ether O, [O](–*)–*, from the averaged QCT values and the TPSA method are similar, differing by approximately 2 Å^2 whereas the QikProp contribution is lower by approximately 3–5 Å^2. To examine this further a value of 1.375 Å was calculated for the average radius of the QCT surface for the ether O in atenolol (see Table 4). As found for the N atoms in prontosil, this value is smaller than those used by other methods in the evaluation of PSA (for example, Stenberg et al.[15] used 1.74 Å and QikProp used 1.628 Å). The variation in this radius, approximately 0.4 Å, suggests that the QCT surface is not uniformly spherical which, as discussed with respect to the N atoms above, accounts for the differences in the PSA contributions for this atom. Importantly, these results for the ether O are not particular to atenolol with average, minimum and maximum radii obtained from all 21 ether O atoms of 1.377 Å, 1.252 Å, and 1.694 Å, respectively.

It is important to consider the meaning of these differences in PSA fragment contributions. If one is interested in the likely bioavailability of molecules and the PSA is being used as a guide (along with other factors) then, as long as upper limits are not rigidly adhered to and methods are not used interchangeably, these differences are perhaps not very important. If, however, one is interested in formulating quantitative relationships between PSA and biological processes (for example, intestinal absorption, oral bioavailability, blood-brain barrier crossing) then it is important that the most realistic PSA contributions are considered and incorporated.

These results suggest differences between quantum topological and traditional (van der Waals radii based) derived atomic surface areas and their use in the calculation of the PSA. The complexity of the

Table 4. The average, minimum and maximum distances (Å) from the nuclei to the 10^{-2} au capping isodensity surfaces.[a]

	QCT Nucleus-Surface Distance (Å)			PSA Contribution (Å^2)		
	Average[b]	Minimum	Maximum	QCT	QikProp	TPSA
N10 (azo)	1.351	1.242	1.782	12.14	11.91	12.36
N11 (azo)	1.362	1.233	1.742	11.63	11.19	12.36
N22 (amino)	1.411	1.272	1.565	21.52	25.10	26.02
O (ether)	1.375	1.271	1.677	11.80	7.38	9.23

[a] Values are provided for the three N nuclei in prontosil shown in Figure 1 and the ether oxygen atom in atenolol. The PSA contributions (Å^2) for these atoms obtained from the various methods are also shown for comparison. [b] Obtained from approximately 1500 points on the outer (nonbonded) surface.

quantum chemical calculations and the complexity of interatomic isodensity surfaces that require QCT analysis preclude the examination of a large number of molecules typical of earlier studies in this area. With the generation of this algorithm capable of providing PSA from ab initio wavefunctions based on electron density partitioning, however, continued analyses of more molecules is warranted to establish the importance of the differences between the models described here.

Conclusions

The calculation of polar surface areas has received attention as a potential descriptor for a number of different biological processes. Whereas previous calculations have based the calculation of PSA values from surfaces created with standard van der Waals radii, herein we described the calculation of polar surface areas from the total electron density partitioned according to quantum chemical topology. This necessitated the development of a new algorithm to enable the definition of the complex surfaces, for example resulting from intramolecular hydrogen bonding, often found in druglike species.

Comparison of PSA values obtained from this method with other readily available methods, TPSA and QikProp, showed how the different atomic and functional group descriptions affected the calculation of the PSA. Comparisons between methods showed differences in PSA, with contributions from the various nitrogenous functional groups showing the greatest variation between methods. Comparisons also highlighted the importance of including similar atom sets in PSA determination and the possibility of fortuitous agreement from dissimilar sets of PSA contributors. While the applications of the PSA to drug design continue to be exploited it is important that the methods used to obtain it are also explored, and where necessary, updated to reflect the results from quantum chemical studies such as these.

Acknowledgements

We thank Dr. Matt Repasky (Schrödinger LLC) for useful discussions and for providing a beta release of QikProp 3.0, and Dr. Michel Rafat for developing the finite element algorithm to calculate QCT PSA. M.D. and P.L.A.P. thank the EPSRC for providing a studentship.

Keywords: atoms in molecules (AIM) · drug design · molecular modeling · polar surface area (PSA) · quantum chemical topology (QCT)

[1] H. van de Waterbeemd, E. Gifford, *Nat. Rev. Drug Discovery* **2003**, *2*, 192–204.
[2] M. C. Wenlock, R. P. Austin, P. Barton, A. M. Davis, P. D. Leeson, *J. Med. Chem.* **2003**, *46*, 1250–1256.
[3] I. Kola, J. Landis, *Nat. Rev. Drug Discovery* **2004**, *3*, 711–715.
[4] S. Whitebread, J. Hamon, D. Bojanic, L. Urban, *Drug Discovery Today* **2005**, *10*, 1421–1433.
[5] K. Palm, K. Luthman, A. L. Ungell, G. Strandlund, P. Artursson, *J. Pharm. Sci.* **1996**, *85*, 32–39.
[6] K. Palm, P. Stenberg, K. Luthman, P. Artursson, *Pharm. Res.* **1997**, *14*, 568–571.
[7] S. Winiwarter, N. M. Bonham, F. Ax, A. Hallberg, H. Lennernas, A. Karlen, *J. Med. Chem.* **1998**, *41*, 4939–4949.
[8] D. E. Clark, *J. Pharm. Sci.* **1999**, *88*, 807–814.
[9] D. E. Clark, *J. Pharm. Sci.* **1999**, *88*, 815–821.
[10] P. Ertl, B. Rohde, P. Selzer, *J. Med. Chem.* **2000**, *43*, 3714–3717.
[11] D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, *J. Med. Chem.* **2002**, *45*, 2615–2623.
[12] J. J. Lu, K. Crimin, J. T. Goodwin, P. Crivori, C. Orrenius, L. Xing, P. J. Tandler, T. J. Vidmar, B. M. Amore, A. G. Wilson, P. F. Stouten, P. S. Burton, *J. Med. Chem.* **2004**, *47*, 6104–6107.
[13] Y. C. Martin, *J. Med. Chem.* **2005**, *48*, 3164–3170.
[14] T. Hou, J. Wang, W. Zhang, X. Xu, *J. Chem. Inf. Model.* **2007**, *47*, 460–463.
[15] P. Stenberg, U. Norinder, K. Luthman, P. Artursson, *J. Med. Chem.* **2001**, *44*, 1927–1937.
[16] R. A. Saunders, J. A. Platts, *New J. Chem.* **2004**, *28*, 166–172.
[17] J. Linnankoski, J. M. Makela, V. P. Ranta, A. Urtti, M. Yliperttula, *J. Med. Chem.* **2006**, *49*, 3674–3681.
[18] C. A. Bergstrom, M. Strafford, L. Lazorova, A. Avdeef, K. Luthman, P. Artursson, *J. Med. Chem.* **2003**, *46*, 558–570.
[19] S. R. Johnson, W. Zheng, *AAPS J.* **2006**, *8*, E27–40.
[20] R. Jones, P. C. Connolly, A. Klamt, M. Diedenhofen, *J. Chem. Inf. Model.* **2005**, *45*, 1337–1342.
[21] A. Cavalli, P. Carloni, M. Recanatini, *Chem. Rev.* **2006**, *106*, 3497–3519.
[22] P. L. A. Popelier, P. J. Smith, *Eur. J. Med. Chem.* **2006**, *41*, 862–873.
[23] P. L. A. Popelier, U. A. Chaudry, P. J. Smith, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 709–718.
[24] P. J. Smith, P. L. A. Popelier, *J. Comput.-Aided Mol. Des.* **2004**, *18*, 135–143.
[25] B. Buttingsrud, E. Ryeng, R. D. King, B. K. Alsberg, *J. Comput.-Aided Mol. Des.* **2006**, *20*, 361–373.
[26] P. L. A. Popelier, *Atoms in Molecules. An Introduction*, Pearson, London, **2000**.
[27] P. L. A. Popelier, F. M. Aicken, *ChemPhysChem* **2003**, *4*, 824–829.
[28] G. Chang, W. C. Guida, W. C. Still, *J. Am. Chem. Soc.* **1989**, *111*, 4379–4386.
[29] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
[30] W. C. Still, A. Tempczyk, R. C. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
[31] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440–467.
[32] *QikProp* version 2.5 Schrödinger, LLC, New York, New York, **2006**.
[33] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
[34] *Jaguar* version 6.5 Schrödinger, LLC, New York, New York, **2006**.
[35] GAUSSIAN 03, Revision C.02, M. J. Frisch; G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Wallingford, Connecticut, **2004**.
[36] R. F. W. Bader, *Atom in Molecules. A Quantum Theory.*, Oxford University Press, Oxford, **1990**.
[37] R. J. Gillespie, P. L. A. Popelier, *Chemical Bonding and Molecular Geometry from Lewis to Electron Densities*, Oxford University Press, New York, USA, **2001**.
[38] M. Rafat, M. Devereux, P. L. A. Popelier, *J. Mol. Graphics Modell.* **2005**, *24*, 111–120.
[39] *Maestro* version 7.5 Schrödinger, LLC, New York, New York, **2006**.

[40] R. F. W. Bader, M. T. Carroll, J. R. Cheeseman, C. Chang, *J. Am. Chem. Soc.* **1987**, *109*, 7968–7979.
[41] M. Rafat, P. L. A. Popelier, *J. Comput. Chem.* **2007**, *28*, 2602–2617.
[42] MORPHY: a program written by P. L. A. Popelier with contributions from M. Rafat and R. G. A. Bone, University of Manchester, England, EU, **2007**.
[43] Molinspiration Property Calculation Service, http://www.molinspiration.-com, **2006**.

Received: September 21, 2007
Revised: November 20, 2007
Published online on December 27, 2007

