The local electron affinity for non-minimal basis sets

Timothy Clark

Editorial Manager(tm) for Journal of Molecular Modeling
Manuscript Draft

Manuscript Number: JMMO1035R1

Title: The Local Electron Affinity for Non-Minimal Basis Sets

Article Type: Original paper

Keywords: Semiempirical molecular orbital theory; local electron affinity; intensity filtering; AM1*

Corresponding Author: Prof. Tim Clark,

Corresponding Author's Institution: Universitaet Erlangen-Nurnberg

First Author: Tim Clark

Order of Authors: Tim Clark

Abstract: A technique known as intensity filtering is introduced to select valence-like virtual orbitals for calculating the local electron affinity, EAL. Intensity filtering allows EAL to be calculated using semiempirical molecular orbital techniques that include polarization functions. Without intensity filtering, such techniques yield spurious EAL values that are dominated by the polarization functions. As intensity filtering should also be applicable for ab initio or DFT calculations with large basis sets, it also makes EAL available for these techniques

Response to Reviewers: All changes suggested by the referees have been made

The local electron affinity for non-minimal basis sets

Received: 25.09.2009 / **Accepted**: 08.10.2009

Timothy Clark[1,2,✉]

[1]Centre for Molecular Design, University of Portsmouth, Mercantile House, Portsmouth PO1 2EG, United Kingdom
[2]Computer-Chemie-Centrum and Interdisciplinary Center for Molecular Materials der Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany

✉Email: Tim.Clark@port.ac.uk

Abstract

A technique known as intensity filtering is introduced to select valence-like virtual orbitals for calculating the local electron affinity, EA_L. Intensity filtering allows EA_L to be calculated using semiempirical molecular orbital techniques that include polarization functions. Without intensity filtering, such techniques yield spurious EA_L values that are dominated by the polarization functions. As intensity filtering should also be applicable for *ab initio* or DFT calculations with large basis sets, it also makes EA_L available for these techniques.

Keywords Semiempirical molecular orbital theory · Local electron affinity · Intensity filtering · AM1*

Introduction

Local properties in the vicinity of molecules can be used as a useful alternative to atom-centered potentials for describing intermolecular interactions. [1]. The best known of these local properties is the molecular electrostatic potential (MEP) [2]. The MEP clearly governs the strength of electrostatic (Coulomb) interactions between molecules. This is the dominant intermolecular interaction in the gas phase, but is made less important by solvation in solvents of high dielectric constant. Weaker intermolecular interactions in the gas phase therefore become more important in polar solutions, crystals or biological systems. Additional local properties are therefore necessary to describe interactions such as electron donor-acceptor (Lewis acid-base) and dispersion. This was first recognized by Sjoberg *et al.*[3], who introduced the local ionization energy, IE_L, which is defined as a density-weighted Koopmans' theorem ionization potential:

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

where HOMO is the highest occupied molecular orbital, ε_i is the Eigenvalue of molecular orbital (MO) *i* and $\rho_i(\mathbf{r})$ is the electron density assignable to MO *i* at position $(\mathbf{r})$. The local ionization energy is a non-equilibrium property that describes the propensity of the molecule to donate electrons at the position $(\mathbf{r})$. It has been linked with the polarizability [4] and has proven to be useful in a variety of *in silico* approaches to predicting physical and chemical properties [5-7].

We [8] later extended this idea to define the local electron affinity for semiempirical MO techniques that use a minimal basis set. EA_L is the equivalent of IE_L for the virtual orbital space:

$$IE_L(\mathbf{r}) = \frac{\sum_{i=LUMO}^{Norbs} -\varepsilon_i \rho_i(\mathbf{r})}{\sum_{i=LUMO}^{Norbs} \rho_i(\mathbf{r})} \qquad (2)$$

where the sum now runs over the virtual orbital space from the lowest unoccupied molecular orbital (*LUMO*) to the total number of orbitals (*Norbs*). EA_L is designed to provide an

electron-acceptor (Lewis acid) pendant to IE_L. Note that EA_L is defined analogously to the electron affinity itself as the ionization potential of the reduced species. The strongest electron-accepting capacity is therefore indicated by the most positive (or least negative) EA_L values. EA_L has proven to be a very important, and often the dominant local property in quantitative structure-property relationship (QSPR) models [9-14] and predictions of biological activity [15]. Indeed, in such applications, IE_L and EA_L often play a statistically more significant role than the MEP, presumably because solvation effects shield electrostatic interactions and because the difference between electrostatic interactions in the receptor and in bulk aqueous solution is small. IE_L and EA_L can be combined in the spirit of Mulliken [16, 17] and Pearson [18] to give the local electronegativity and hardness, respectively. [8]. They represent an extension of the idea of the Fukui function[19], which, however is limited by the frontier-orbital approximation to considering only the HOMO and LUMO of the molecule in question.

It is often pointed out that virtual orbitals are meaningless because they do not affect the energy and are therefore not optimized. This is strictly true, but virtual orbitals have in practice played a significant role in qualitative molecular orbital treatments [20]. They are useful because they are constrained by the requirement that they are orthogonal to occupied orbitals (and each other). This leads to the well known correspondence between bonding occupied orbitals and their antibonding virtual equivalents. We will call this the "orthogonalization constraint" in the following. This orthogonalization constraint results in virtual orbitals that are valence-like antibonding equivalents of one or a combination of several (optimized) bonding orbitals in the occupied space. Thus, we can distinguish five subsets of molecular orbitals, as shown in Fig. 2.

- *Figure 2 here* –

The term Rydberg in Fig. 2 reflects the nomenclature used in the natural bond order (NBO) analysis [21, 22] and indicates polarization functions, diffuse orbitals *etc*. that are only represented weakly in the occupied space. Ideally, we would like to include the bonding and Lewis-acid orbitals in the calculation of EA_L but exclude the Rydberg block. This happens automatically for minimal basis sets because no basis functions are available to describe Rydberg-type orbitals. If we now assume, as is approximately the case, that the antibonding block spans the same orbital space as the bonding one, we can define a criterion for filtering the antibonding block from the total virtual space. This criterion is only approximately correct

for split-valence basis sets because antibonding orbitals tend to be more diffuse than their bonding counterparts (*i.e*. they "borrow intensity" from the Rydberg space), but proves to work well in practice. Note that there is no orthogonality constraint between lone pairs and localized Lewis-acid acceptor orbitals because they are not bonding-antibonding pairs. This would lead to exclusion of the Lewis-acid acceptors from the valence-like virtual space if they were strictly localized. In practice they are delocalized enough to interact with the bonding block for all but the smallest molecules (see below).

Eq. 2 has so far been used exclusively for semiempirical molecular orbital theory calculations [23] using techniques such as MNDO [24], AM1 [25] or PM3 [26] that use minimal basis sets. This is because large basis sets in *ab initio* or density functional theory (DFT) calculations, or even semiempirical techniques such as AM1* [27-32] that use *d*-orbitals as polarization functions contain many basis functions that are very weakly occupied and therefore dominate the virtual space. The effect of such "non-valence" orbitals is to dominate the sum in Eq. 2 locally in the vicinity of the atom concerned. This leads to spuriously large negative EA_L values in the affected area. The problem is that the virtual space is often considerably larger than the occupied one so that many virtual orbitals are not limited sufficiently by the orthogonality requirement with occupied orbitals. This has the effect that, for instance, the EA_L calculated with AM1* differs from that calculated with AM1 by having strong negative peaks around atoms with polarization functions, as shown in Fig. 1 for a typical drug-like molecule, **1**.

- *Scheme 1 (no title) here –*
- *Figure 1 here –*

This is clearly a severe limitation, especially when the importance of EA_L in QSAR and QSPR applications is taken into account. We therefore now report a practical technique that can be used to calculate EA_L for semiempirical methods with *d*-polarization functions and for Hartree-Fock or DFT calculations with large basis sets.

Methods

The filtering criterion outlined qualitatively above is analogous to the density-overlap requirement that determines the oscillator strength for electronic excitations [33], so that we

have named it “intensity filtering”. Within the zero differential overlap (ZDO) approximation the density overlap $|O|_{ij}^{ZDO}$ between virtual and occupied orbitals *i* and *j*, respectively, is

$$|O|_{ij}^{ZDO} = \sum_{k=1}^{norbs} |c(i,k) \cdot c(j,k)| \tag{3}$$

where $c(i,k)$ is the coefficient of atomic orbital *k* in the LCAO-Eigenvector for MO *i*.

One possible corresponding expression for a non-orthogonal basis is

$$|O|_{ij} = \langle \Psi_i^2 | \Psi_j^2 \rangle \tag{4}$$

where Ψ_i denotes molecular orbital *i* and the squares ensure absolute values of the overlap. We now examine the use of $|O|_{ij}^{ZDO}$ as a criterion (intensity filtering) for calculating EA_L for semiempirical MO techniques that use polarization functions. In a later paper, we will investigate the use of Eqs. 3 and 4 for Hartree-Fock *ab initio* and DFT calculations with large basis sets.

Each virtual orbital whose maximum $|O|_{ij}^{ZDO}$ with an occupied orbital is larger than an arbitrary threshold is included in the calculation of EA_L.

ZDO Selection Criterion for AM1*

Table 1 shows details of the $|O|_{ij}^{ZDO}$ analysis of molecule **2**. The highest five virtual orbitals are excluded if we use an $|O|_{ij}^{ZDO}$ threshold of 0.5. However, in this case, the choice of threshold is not critical because the gap in $|O|_{ij}^{ZDO}$ values between these five orbitals and the remainder stretches from 0.23 to 0.76. The five virtual orbitals with no close occupied equivalent all consist predominantly of *d*-polarization functions on the sulfur.

- *Table 1 here* -

Fig. 3 shows the EAL projected onto the same molecular surface for the AM1* wavefunction as that shown in Fig. 1 using a slightly modified version of Eq. 2 in which the five MOs are excluded:

$$IE_L(\mathbf{r}) = \frac{\sum_{i=LUMO}^{Norbs} -\varepsilon_i \rho_i(\mathbf{r}) \delta_i}{\sum_{i=LUMO}^{Norbs} \rho_i(\mathbf{r}) \delta_i}$$

$$\delta_i = 1 \text{ for maximum } \left|O\right|_i^{ZDO} \geq 0.5$$

$$\delta_i = 0 \text{ otherwise} \qquad (5)$$

The large negative peak around the sulfur no longer occurs and the map resembles that given by AM1. However, there are some differences caused by the importance of polarization functions for describing heavy elements.

The separation observed for **1** is quite general. Fig. 4 shows a scatter plot of $\left|O\right|_{ij}^{ZDO}$ against ε_i for all virtual orbitals of a dataset of 74 neutral compounds containing S, P, Cl, Br and I taken from a logP dataset [34]. For this set of neutral compounds, there is one exception to the $\left|O\right|_{ij}^{ZDO} \geq 0.5$ criterion, which will be discussed below. Note that for this set of compounds, a simple energy criterion ($\varepsilon_i \leq 8$ eV) would work well and also treat the outlier correctly. However, if we simply deprotonate one of the compounds to give anion **3**, the energy criterion is no longer appropriate. The outlier (the red point) corresponds to a CH–antibonding orbital of the methoxy-substituent on the ring of compound **4**.

- *Figure 3 here* -

- Scheme 2 (no title) here –

Fig. 5 shows the "outlier" orbital with the occupied orbital with which it has the highest density overlap.

- *Figure 5 here* -

The low density overlap (0.48) between these two orbitals is caused by the fact that the occupied orbital is strongly delocalized, whereas its virtual counterpart is localized on the methyl group. This situation is not uncommon as only delocalization of the occupied orbital leads to stabilization; the virtual orbital has no reason to delocalize as long as it remains orthogonal to all others. A further source of exceptions to the $\left|O\right|_{ij}^{ZDO} \geq 0.5$ criterion is a Lewis-acceptor orbital such as that indicated in Fig. 2. In the worst case, for instance the unoccupied *p*-orbital on aluminum in planar AlH_3, the density overlap with occupied orbitals is zero. In practice, delocalization increases $\left|O\right|_{ij}^{ZDO}$ for all but the smallest molecules (the maximum density overlap for the corresponding Lewis-acceptor orbital in $Al(CH_3)_3$ is 0.4), but may not increase it above the 0.5 threshold.

We have therefore implemented a further check to capture such exceptions. Quite simply, if virtual orbital $i+1$ satisfies the $\left|O\right|_{ij}^{ZDO} \geq 0.5$ criterion, but not orbital i, orbital i is included in the virtual subspace used to calculate EA_L. Thus, a continuous block of virtual orbitals is selected. This is equivalent to a variable energetic criterion, but automatically allows for the effects of charge on the orbital energies.

Conclusions

The suggested ZDO-based intensity-filtering method provides a fast and effective technique for selecting “valence-like” virtual orbitals in semiempirical calculations using basis sets that are larger than minimal. When used to calculate the local electron affinity, the filtering technique leads to results similar to those given by techniques that use minimal basis sets. The intensity-filtering technique is in principle applicable to DFT or *ab initio* calculations that uses extended basis sets, although it may be necessary to use Eq. 4 rather than the ZDO-based Eq. 3 for such calculations. We are currently investigating the use of Eqs. 3 and 4 and their basis-set dependence for DFT calculations.

Alternative, even simpler schemes such as using as many virtual as occupied orbitals or simple energy filtering do not lead to the same results because the valence-like space may not be balanced between occupied and virtual orbitals (*e.g.* polyhalogen compounds have many more occupied valence-like orbitals than virtual ones) and because of charge effects.

Above all, intensity filtering now means that EA_L can be used as a useful and easily calculated index of electrophilicity as the pendant to IE_L as an index of nucleophilicity. Together, they avoid the limitation of the frontier-orbital approximation inherent in the Fukui function [19].

The technique introduced here has been implemented in ParaSurf'10 [35].

Acknowledgments

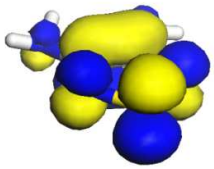
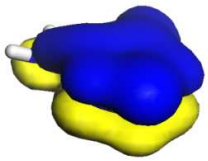
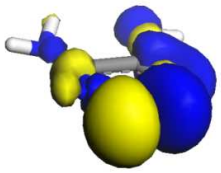
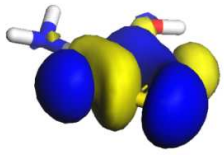
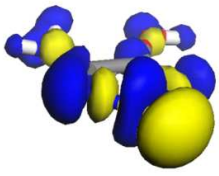
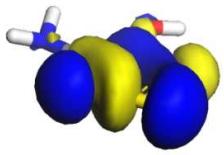
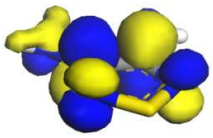
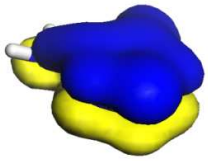
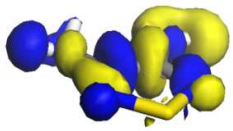
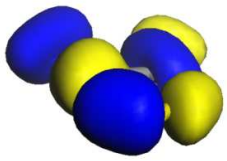
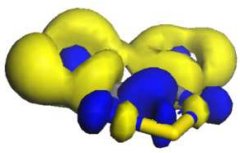
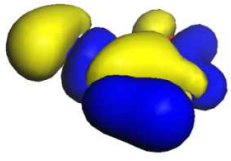
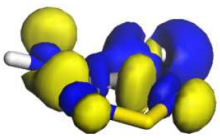
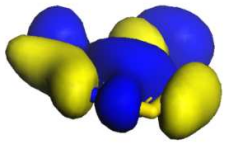
I am grateful for constructive criticism and many stimulating discussions with Jane Murray, Peter Politzer and Felipe Bulat. This work was supported by the Deutsche Forschungsgemeinschaft as part of SFB583 *"Redox-Active Metal Complexes: Control of Reactivity via Molecular Architecture"*

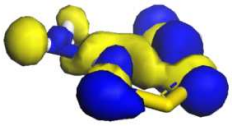
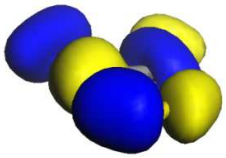
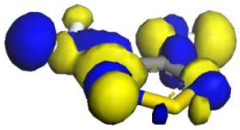
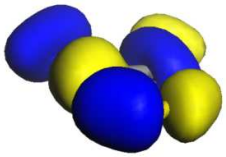
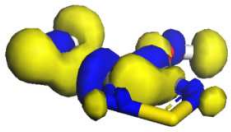
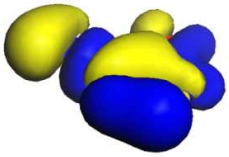
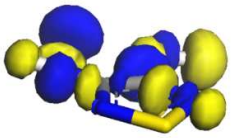
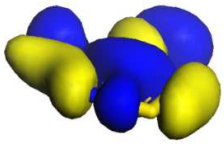
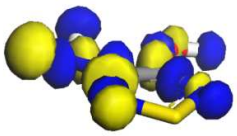
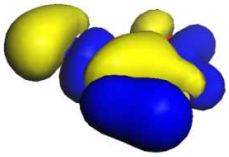
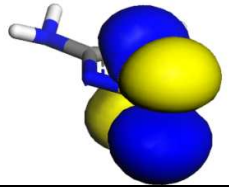
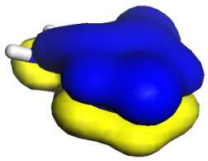
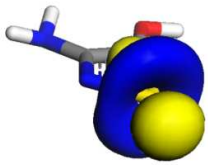
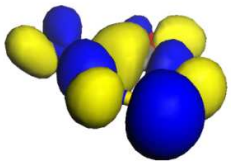
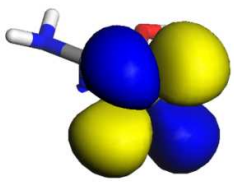
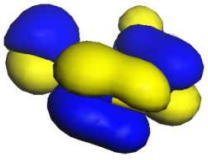
References

1. Clark T, Byler KG, de Groot MJ (2008) In: Molecular Interactions - Bringing Chemistry to Life. Logos Verlag, Berlin, pp 129-146
2. Politzer P, Murray JS (1991) In: Lipkowitz K, Boyd RB (eds) Rev Comput Chem 2:273-312
3. Ehresmann B, Martin B, Horn AHC, Clark T (2003) J Mol Model 9:342-347
4. Politzer PA, Murray JS, Grice ME, Brinck T, Ranganathan S (1991) J Chem Phys 95:6699-6704
5. Politzer PA, Murray JS, Concha MC (2002) Int J Quant Chem 88:19-27 and references therein
6. Hussein W, Walker CJ, Peralta-Inga Z, Murray JS (2001) Int J Quant Chem 82:160-169
7. Murray JS, Abu-Awwad F, Politzer PA (2000) THEOCHEM 501-502:241
8. Ehresmann B, de Groot MJ, Alex A, Clark T (2004) J Chem Inf Comp Sci 44:658-668
9. Clark T (2004) J Mol Graph Model 22:519-525
10. Ehresmann B, de Groot MJ, Clark T (2005) J Chem Inf Model 45:1053-1060
11. Clark T, Ford MG, Essex JW, Richards WG, Ritchie DW (2006) QSAR and Molecular Modelling in Rational Design of Bioactive Molecules, EuroQSAR 2004 Proceedings. In: Aki E, Yalcin I (eds) CADDDS in Turkey, Ankara, pp 536-537
12. Kramer C, Beck B, Kriegl JM, Clark T (2008) Chem Med Chem 3:254-265
13. Jakobi A-J, Mauser H, Clark T (2008) J Mol Model 14:547-558
14. Hennemann H, Friedl A, Lobell M, Keldenich J, Hillisch A, Clark T, Göller AH (2009) Chem Med Chem 4:657-669
15. Manallack DT (2008) J Mol Model 14:797-805
16. Mulliken RS (1934) J Chem Phys 2:782-793
17. Mulliken RS (1955) J Chem Phys 23:1833-1840
18. Pearson RG (1991) Chemtracts, Inorganic Chemistry 3:317-333
19. Parr RG, Yang W (1989) Density Functional Theory in Atoms and Molecules. Oxford University Press, New York
20. Fukui K, Yonezawa T, Nagata C (1957) J Chem Phys 26:831-841
21. Foster JP, Weinhold F (1980) J Am Chem Soc 102:7211-7218
22. Reid AE, Weinstock RB, Weinhold F (1985) J Chem Phys 83:735-746

23. Clark T, Stewart JJP (2010) In: Reimers JR (ed) Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology. Wiley, New York, to be published
24. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4899-4907;
Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4907-4917;
Thiel W (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger, J, Kollman PA, Schaefer HF III, Schreiner PR (eds) Encyclopedia of Computational Chemistry. Wiley, Chichester, 3:1599-1604
25. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc 107*:3902-3909;
Holder AJ (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger, J, Kollman PA, Schaefer HF III, Schreiner PR (eds) Encyclopedia of Computational Chemistry. Wiley, Chichester, 1:8-11
26. Stewart JJP (1989) J Comp Chem 10:209; Stewart JJP (1989) J Comp Chem 10:221; Stewart JJP (1998) Encyclopedia of Computational Chemistry, Schleyer PvR, Allinger NL, Clark T, Gasteiger, J, Kollman PA, Schaefer HF III, Schreiner PR (Eds), Wiley, Chichester 3:2080
27. Winget P, Horn AHC, Selçuki C, Martin B, Clark T (2003) J Mol Model 9:408-414
28. Winget P, Clark T (2005) J Mol Model 11:439−456
29. Kayi H, Clark T (2007) J Mol Model 13:965-979
30. Kayi H, Clark T (2009) J Mol Model 15:295-308
31. Kayi H, Clark T (2009) J Mol Model 15:online first (DOI 10.1007/s00894-009-0489-y)
32. Kayi H, Clark T (2009) J Mol Model 15:online first (DOI 10.1007/s00894-009-0503-4)
33. Orchin M, Jaffé HH (1971) Symmetry, Orbitals, and Spectra. Wiley, New York, NY
34. Kramer C, Beck B, Clark T (2009) submitted to Chem Med Chem
35. ParaSurf'10, Cepos InSilico Ltd, Kempston, UK, to be released on July 1st 2010; www.ceposinsilico.com

Table 1 The virtual MOs of **2** (AM1*) with their highest $|O|_{ij}^{ZDO}$ values and the occupied orbitals that give these values

Virtual MO		Occupied MO		$\|O\|_{ij}^{ZDO}$
20		11		0.869
21		17		0.764
22		17		0.826
23		11		0.832
24		6		0.823
25		9		0.688
26		8		0.858

27		6		0.764
28		6		0.799
29		9		0.821
30		8		0.828
31		9		0.815
32		11		**0.123**
33		14		**0.146**
34		18		**0.181**

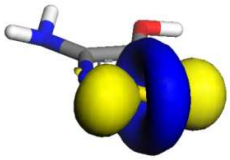
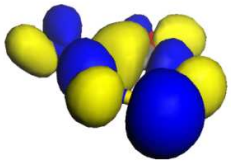
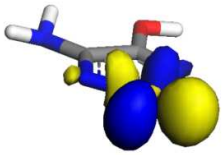
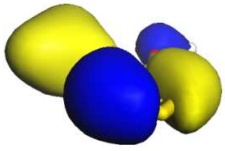
35		14		**0.203**
36		4		**0.232**

Figure captions

Fig. 1 EA_L calculated using Eq. 2 projected onto equivalent isodensity surfaces for molecule **1** using the AM1 and AM1* Hamiltonians

Fig. 2 Schematic representation of the five subsets of molecular-orbital space

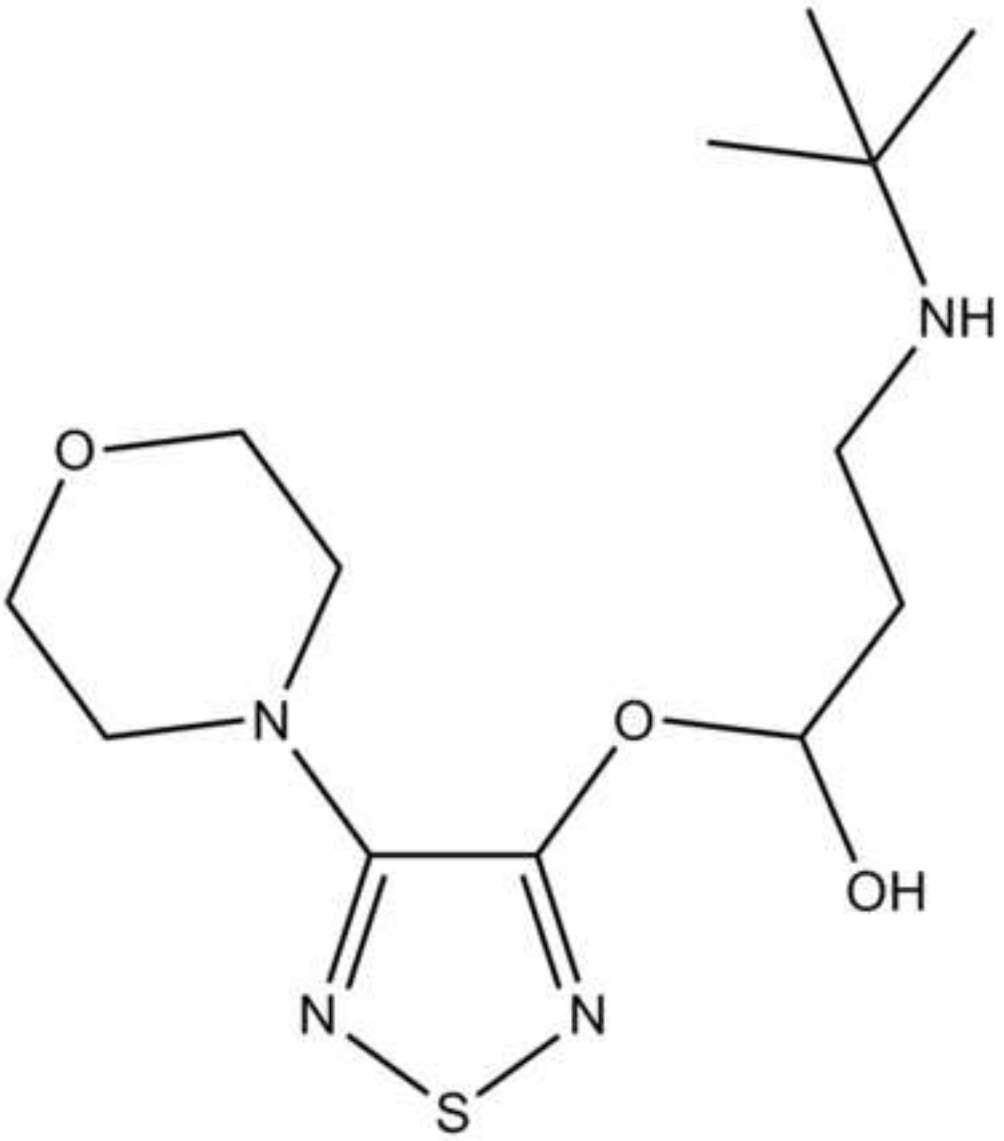
Fig. 3 EA_L calculated using Eq. 5 projected onto the same isodensity surface as used in Fig. 1 for molecule **1** using the AM1* Hamiltonian

Fig. 4 Scatter plot of $|O|_{ij}^{ZDO}$ *vs.* ε_i for some typical S, P, Cl, Br and I-containing compounds. The large squares are data for anion **3**. The vertical red dotted line represents the $|O|_{ij}^{ZDO} \geq 0.5$ criterion. The red point corresponds to the orbital discussed for **4** below

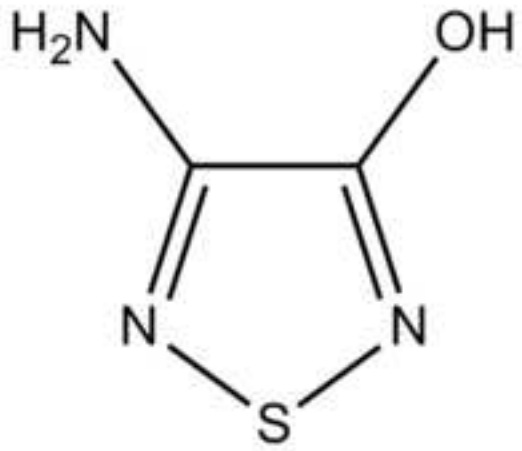
Fig. 5 The "outlier" virtual orbital indicated by the red point in Fig. 4 and the occupied orbital with which it has the highest density overlap

scheme 1

Click here to download high resolution image



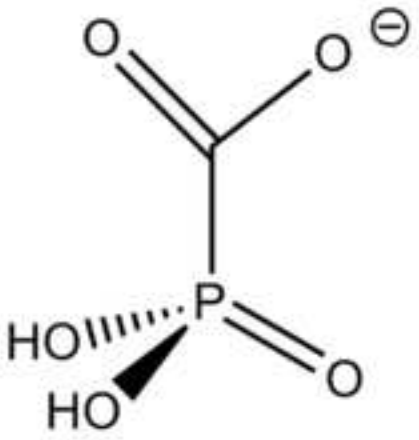
1



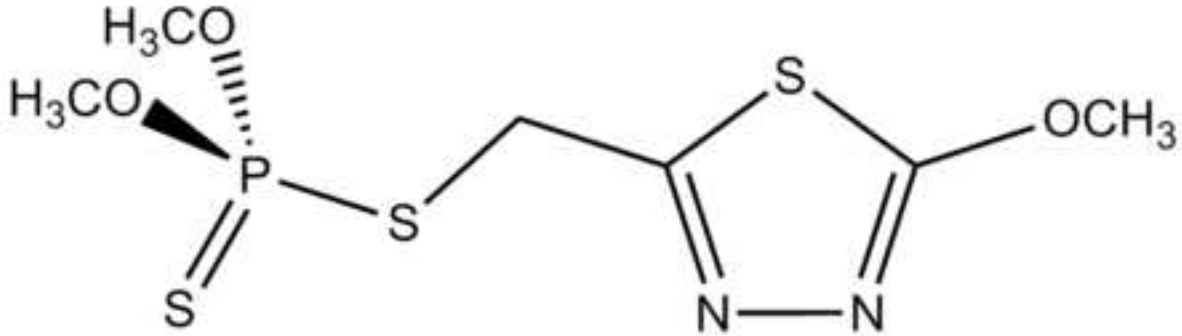
2

scheme 2

Click here to download high resolution image



3



4

figure 1
Click here to download high resolution image

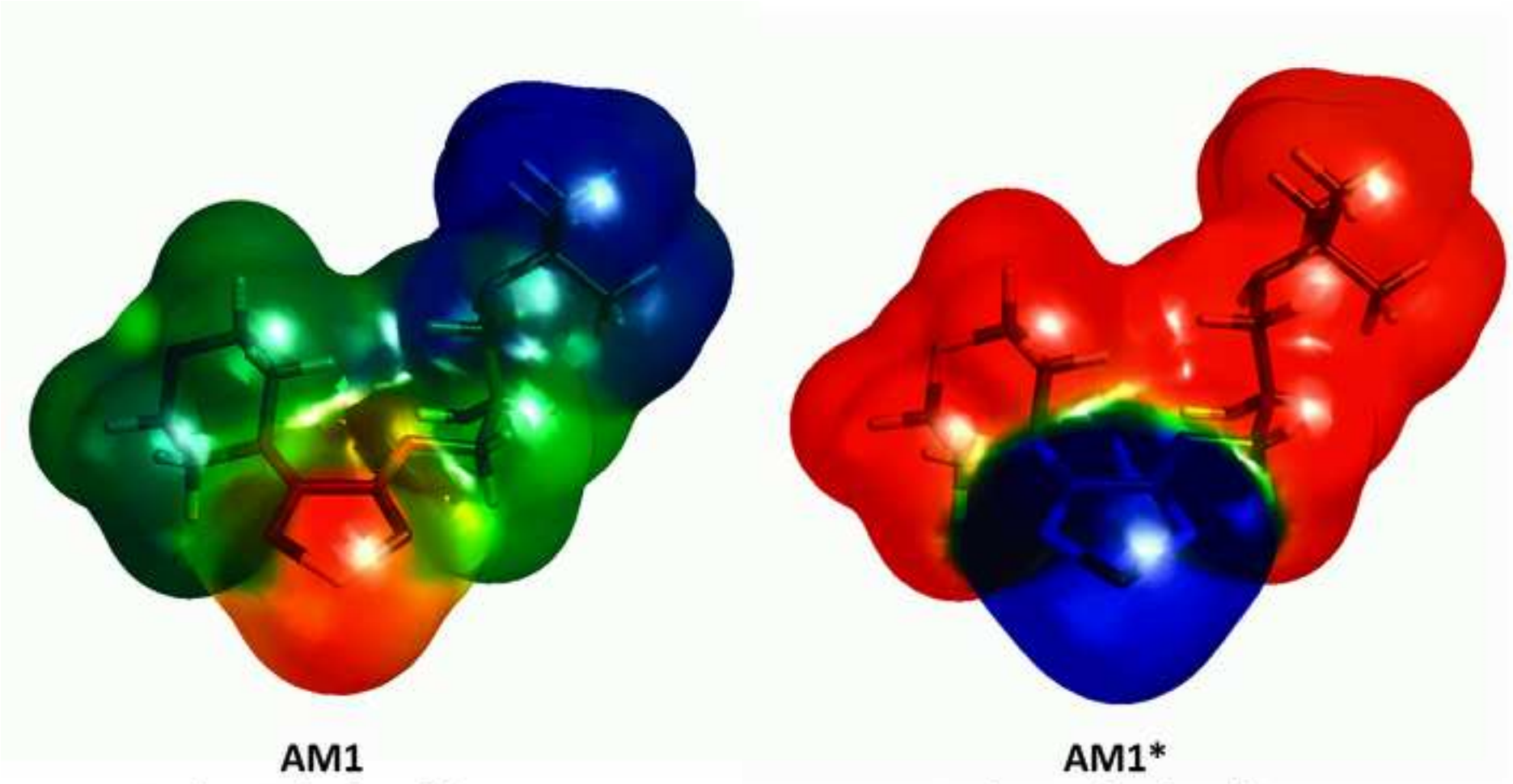


figure 2
Click here to download high resolution image

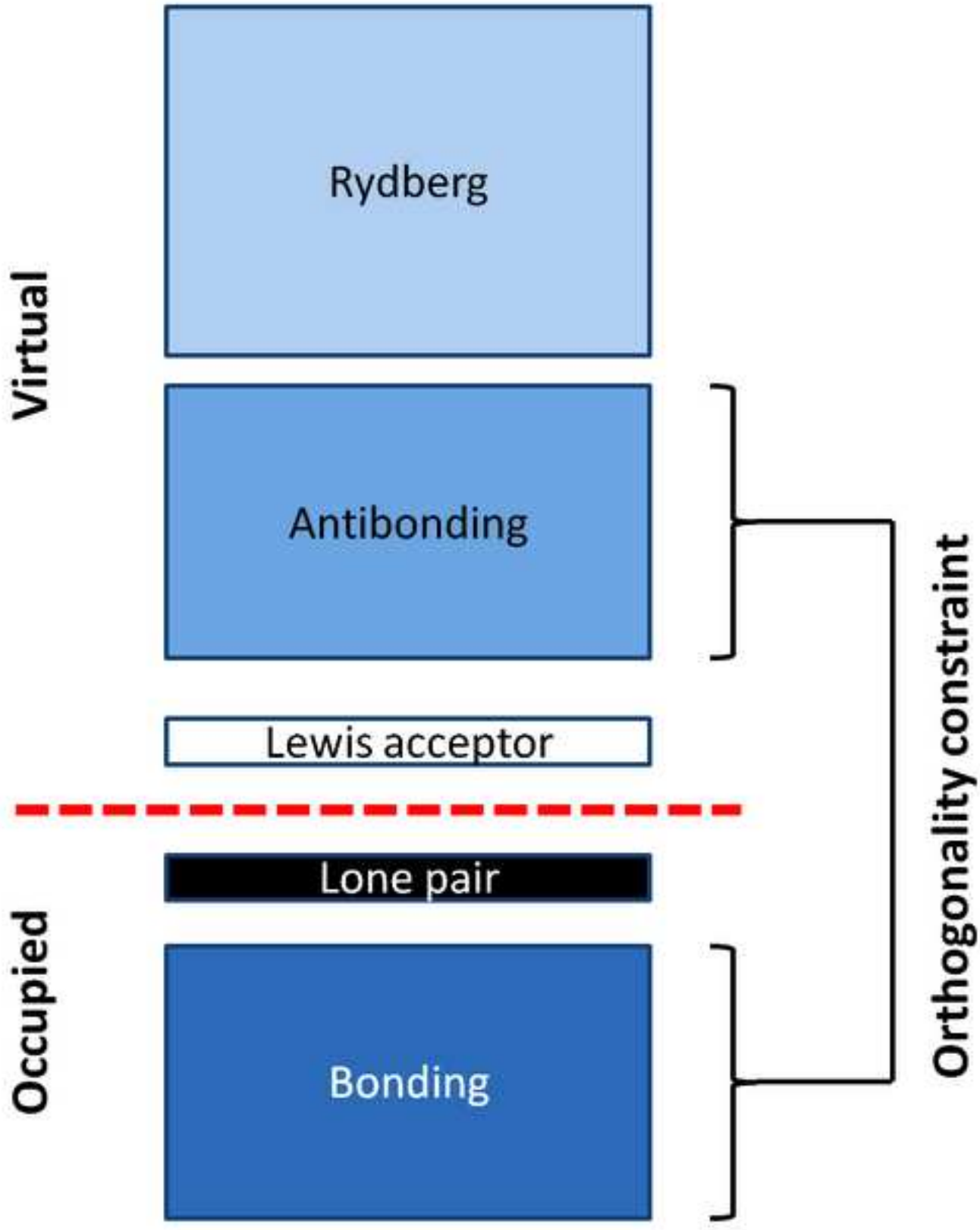
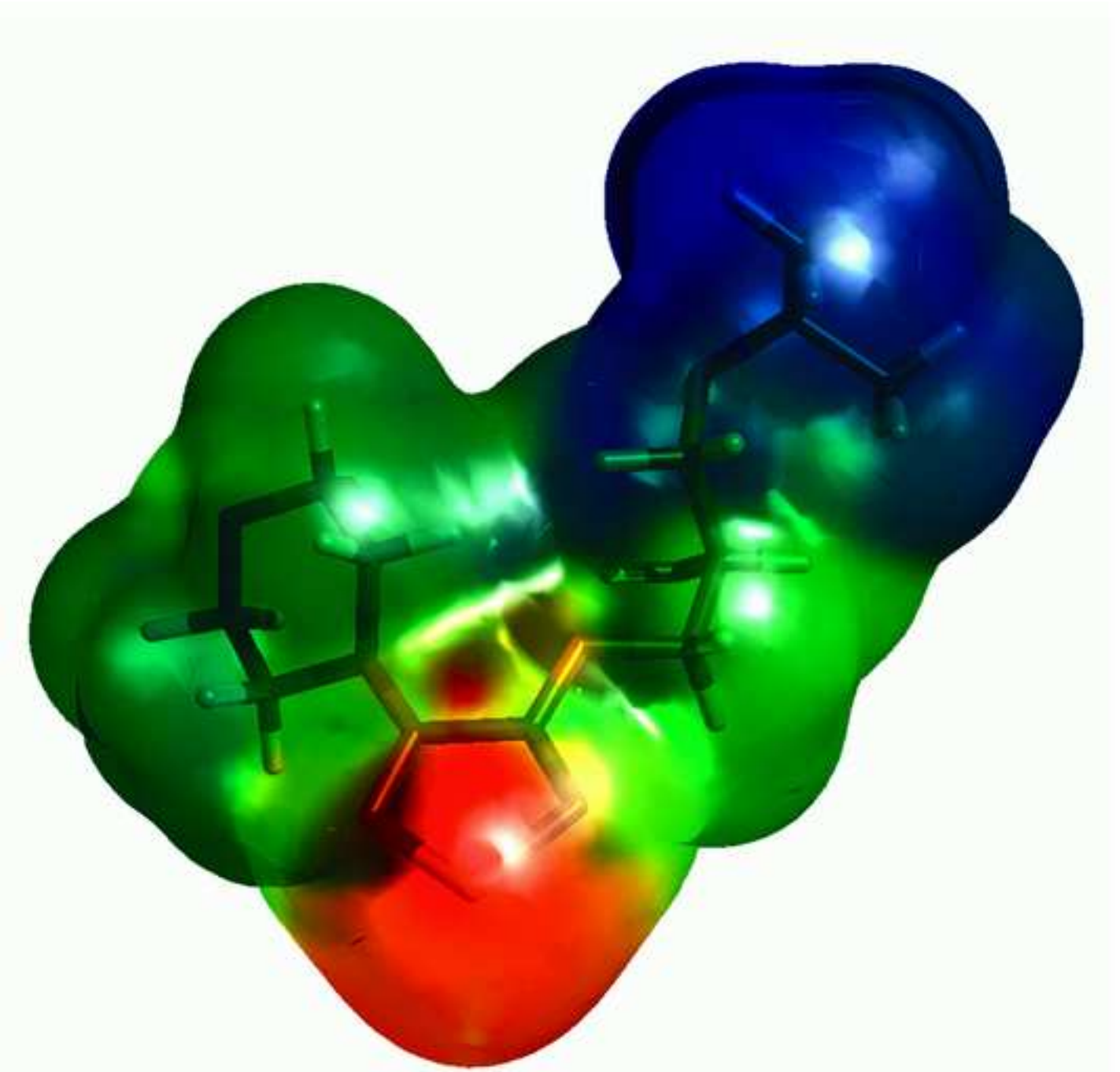


figure 3
Click here to download high resolution image



AM1*

Red = 0 kcal mol^{-1}

Blue = -120 kcal mol^{-1}

figure 4
Click here to download high resolution image

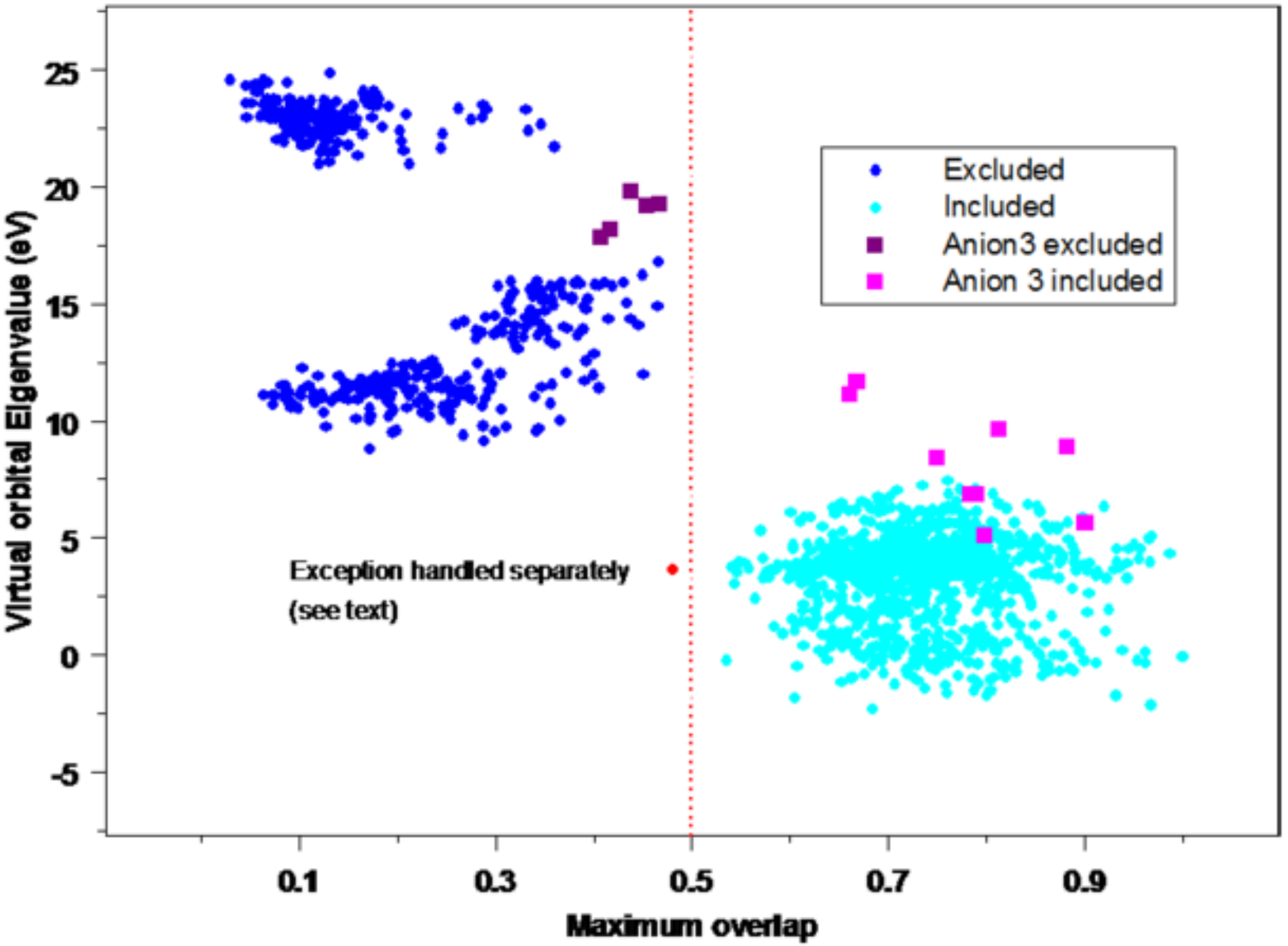


figure 5
Click here to download high resolution image