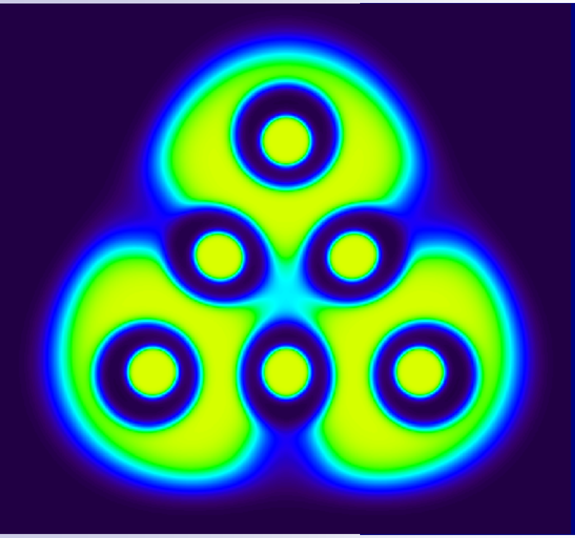


Predicting reactive sites

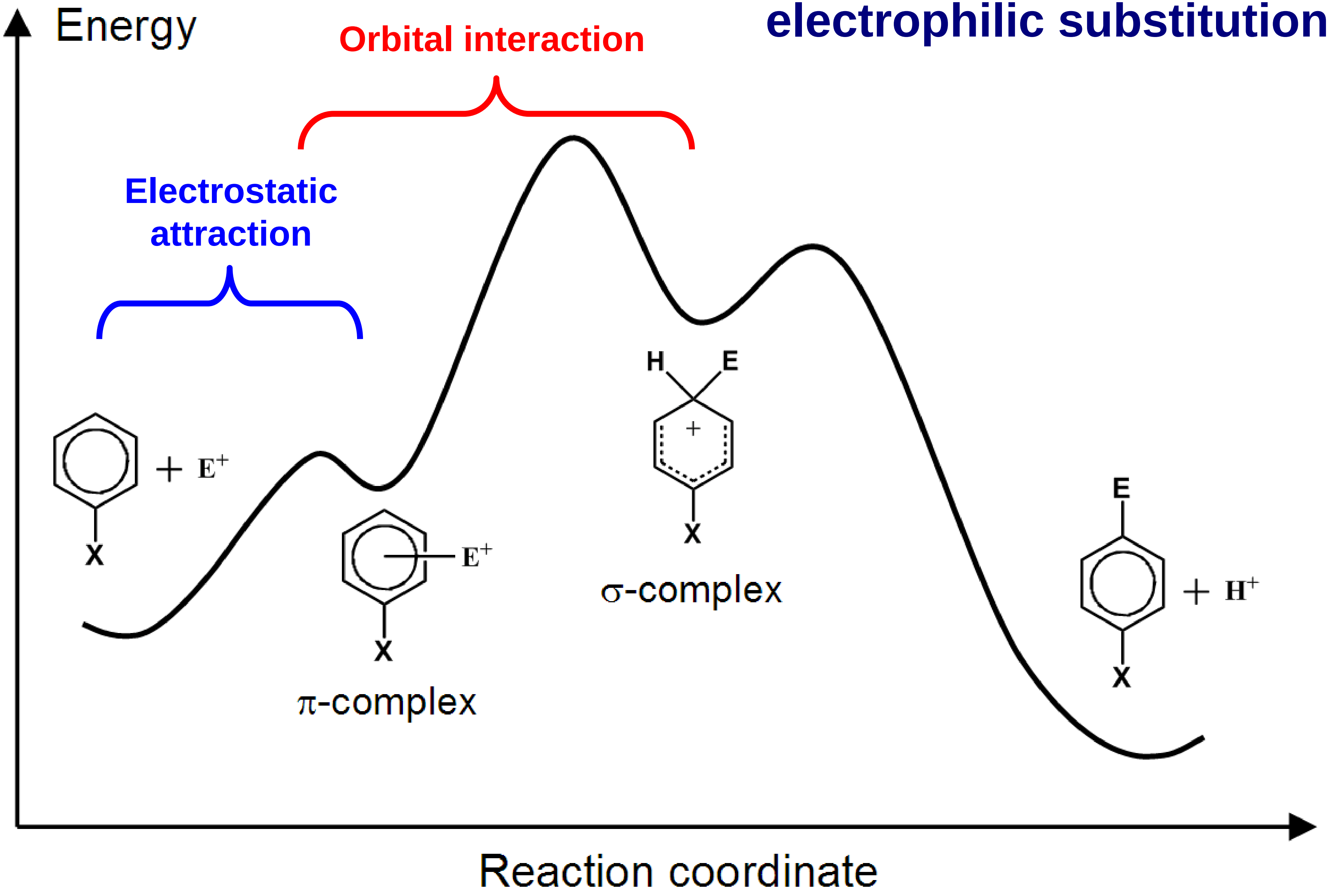
Tian Lu

sobereva@sina.com

Beijing Kein Research Center for Natural Sciences

2014-May-17

Basic mechanism of electrophilic substitution



Possible ways of predicting reactive sites for electrophilic reaction

- Electrostatic potential (ESP)
- Atomic charge
- Population number of p_z atomic orbital
- Average local ionization energy (ALIE)
- Frontier molecular orbital theory
- Fukui function
- Dual descriptor

Electrostatic potential (ESP)

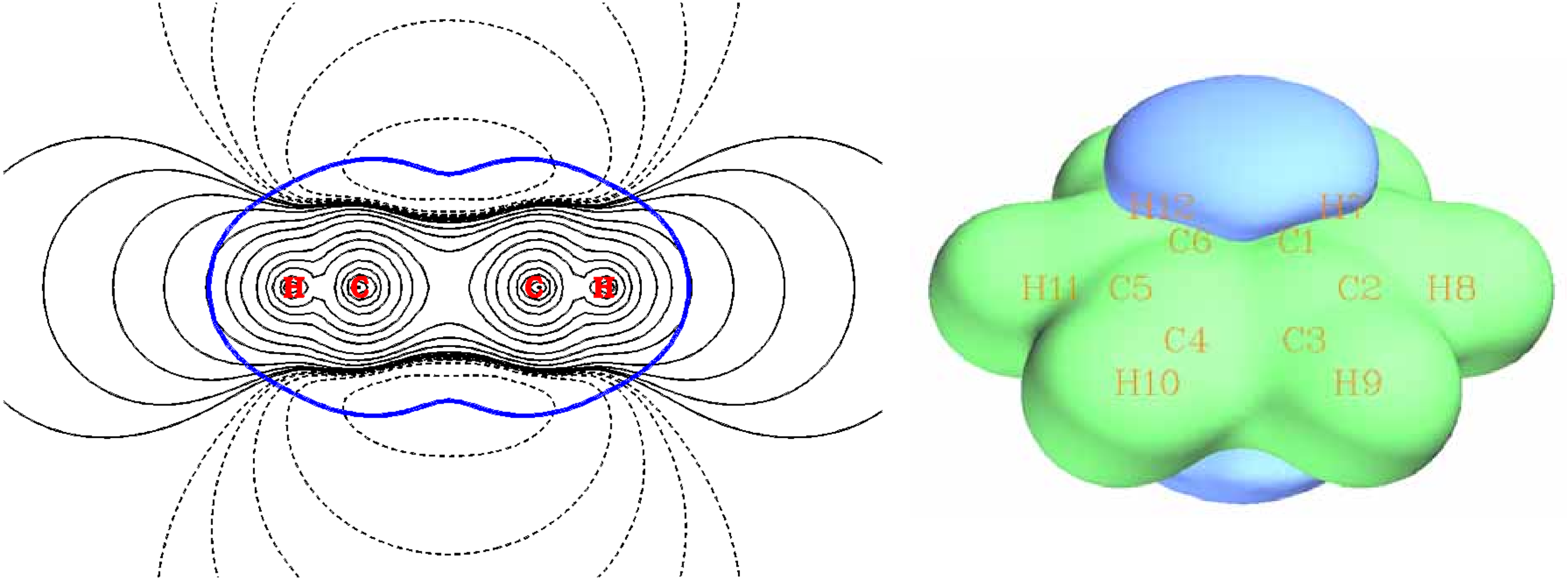
$$V_{tot}(\mathbf{r}) = V_{nu}(\mathbf{r}) + V_{ele}(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'$$

ESP measures the electrostatic interaction between an unit point charge placed at **r** and the system of interest

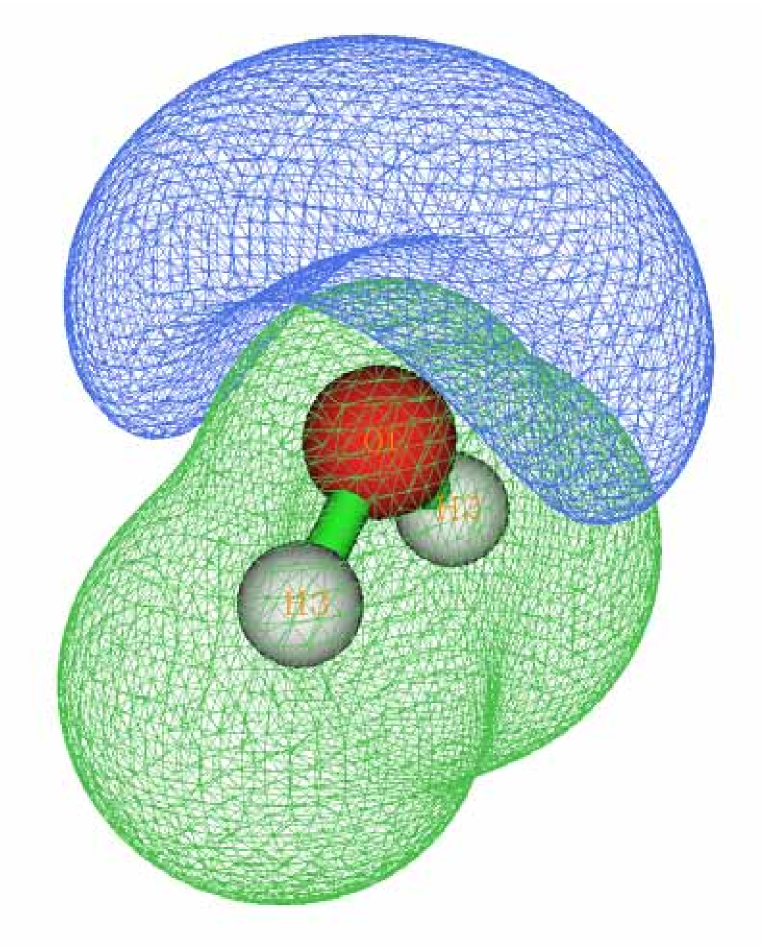
A positive (negative) value implies that current position is dominated by nuclear (electronic) charges

After formation of bonds, if electron density distribution is remarkably enriched somewhere, then negative ESP could occur (but must far from nuclei)

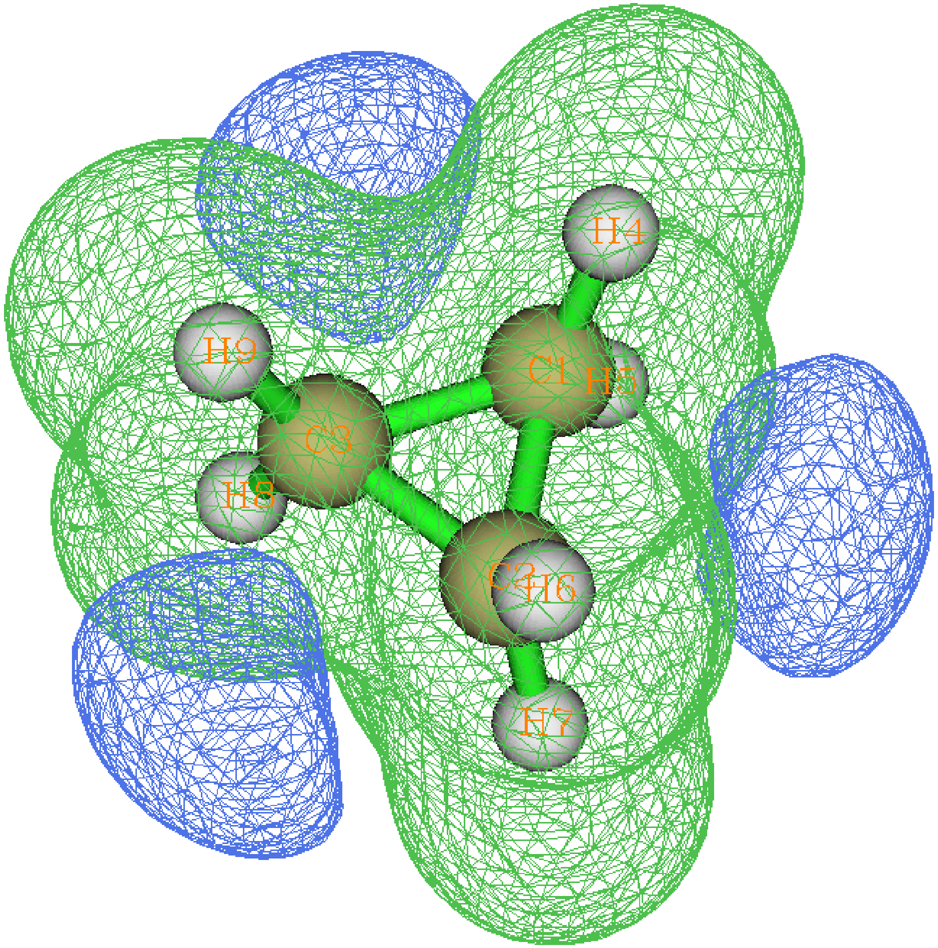
Negative ESP usually occurs at π-cloud region and lone pair region



Benzene



Water

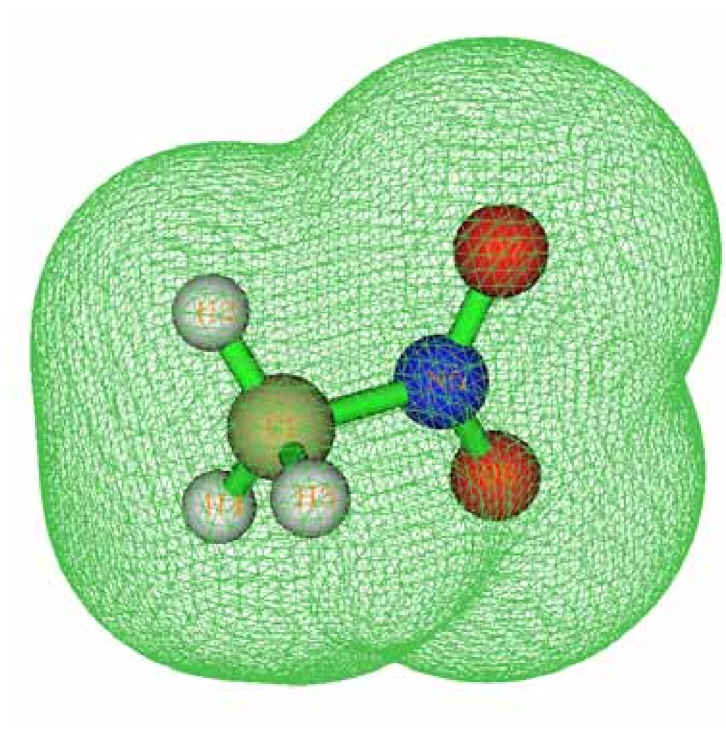


Cyclopropane

ESP is commonly investigated on vdW surface. The most commonly used definition of vdW surface is the one proposed by Bader

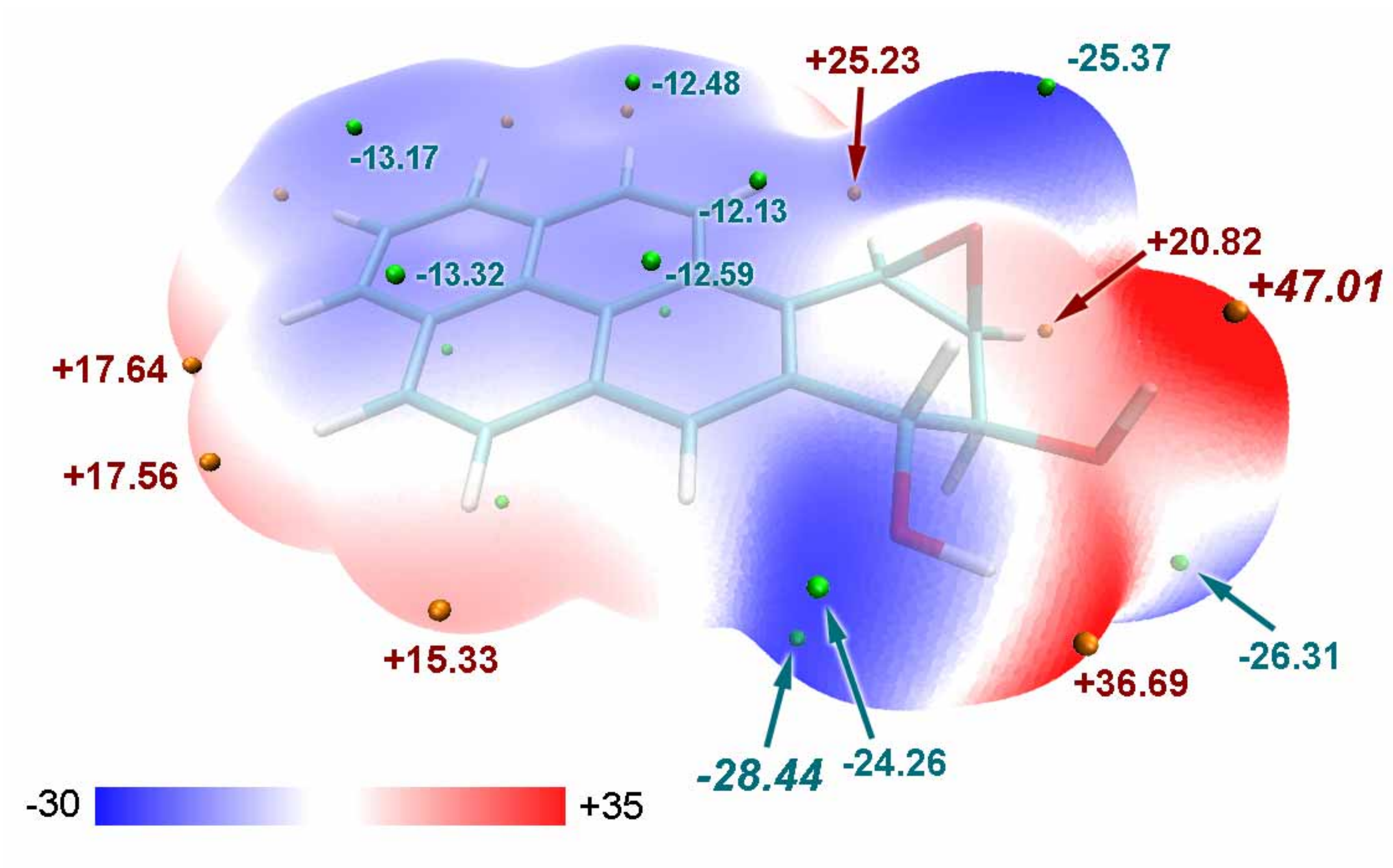
Isosurface of ρ=0.001 a.u. for gas phase
Isosurface of ρ=0.002 a.u. for condensed phase

a.u. ≡ e/Bohr3



This definition reflects specific electron structure features of a molecule, such as lone pairs, π electrons, etc.

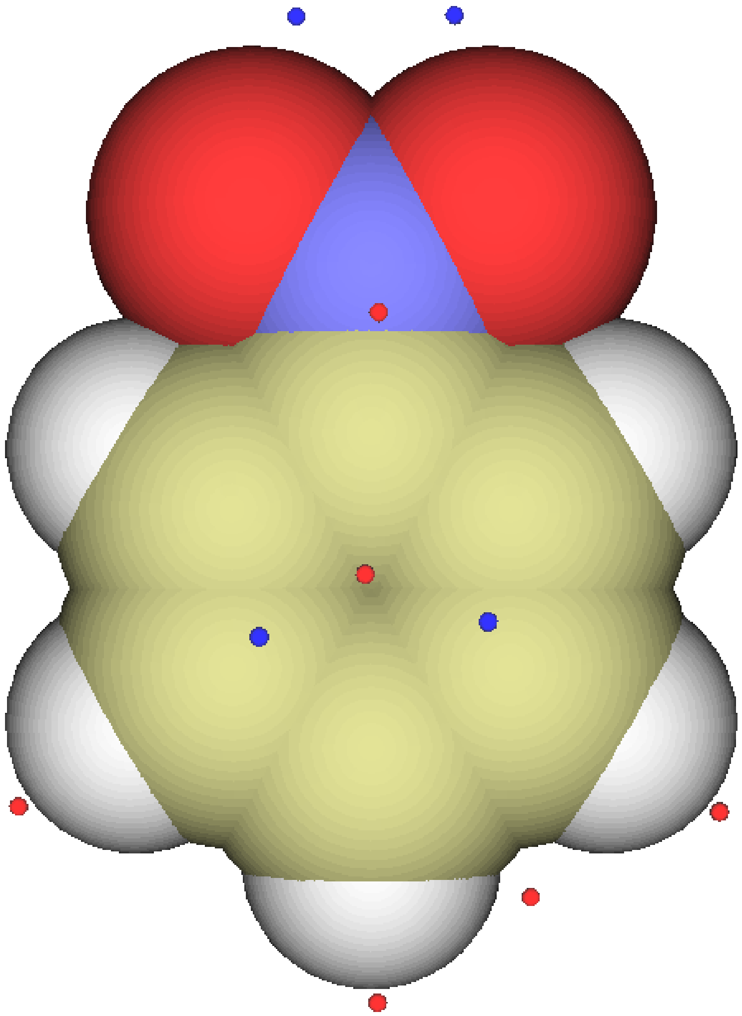
Quantitative molecular surface analysis of ESP



ESP mapped vdW surface of benzo[a]pyrene diol epoxide

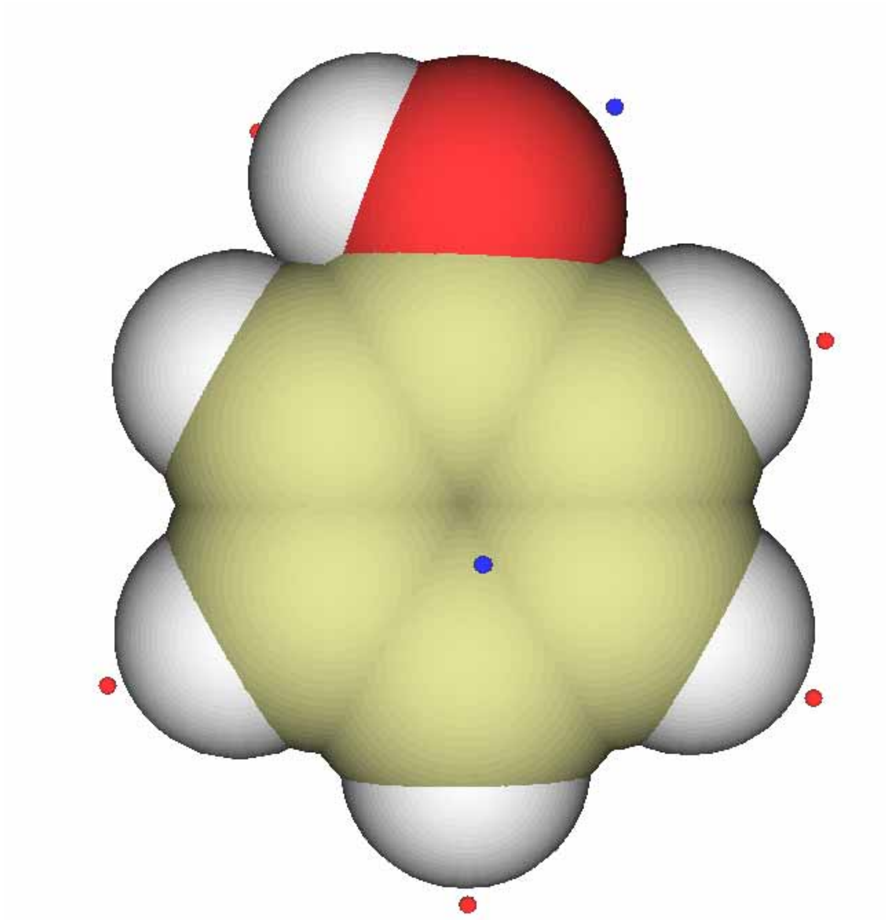
DOI: 10.1007/s11224-014-0430-6

The atom closer to the global ESP minimum (maximum) on vdW surface is more likely to be the favorable site of electrophilic (nucleophilic) reaction



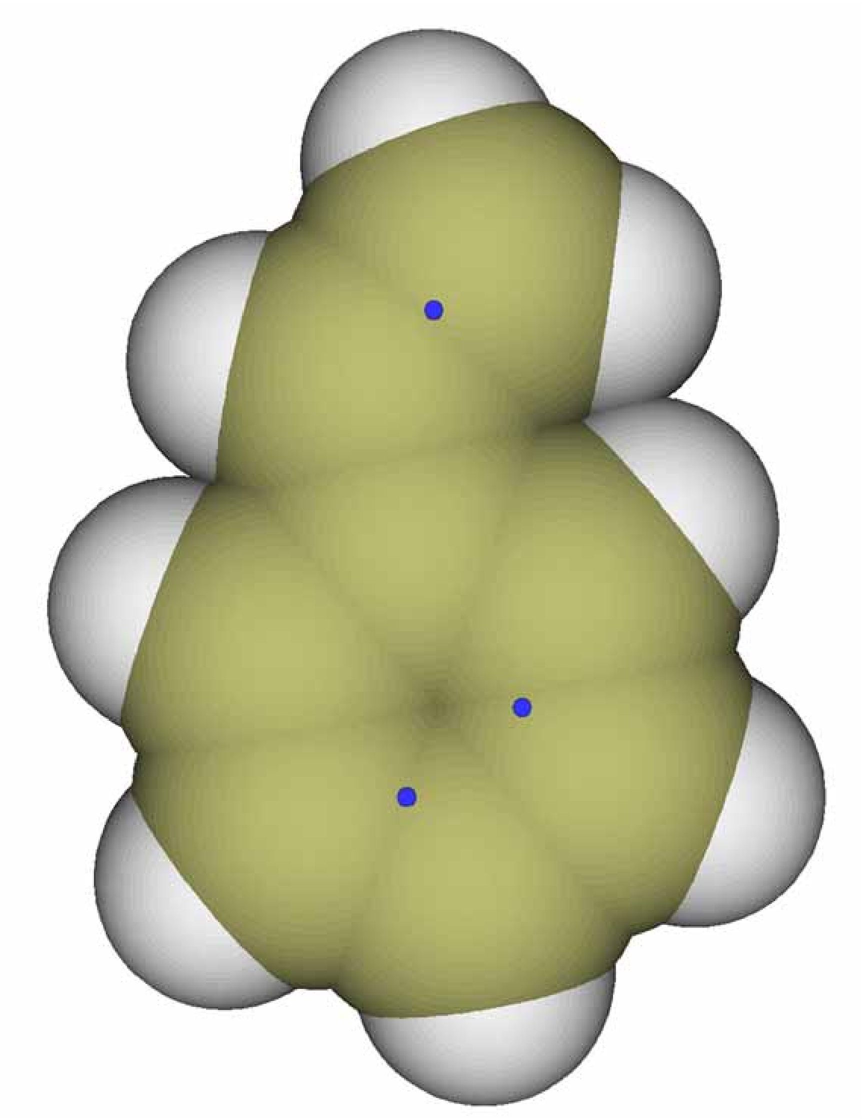
Nitrobenzene

The role of ortho-para director of hydroxyl group is not fully revealed
Impossible to compare relative reactivity of ortho and meta carbons



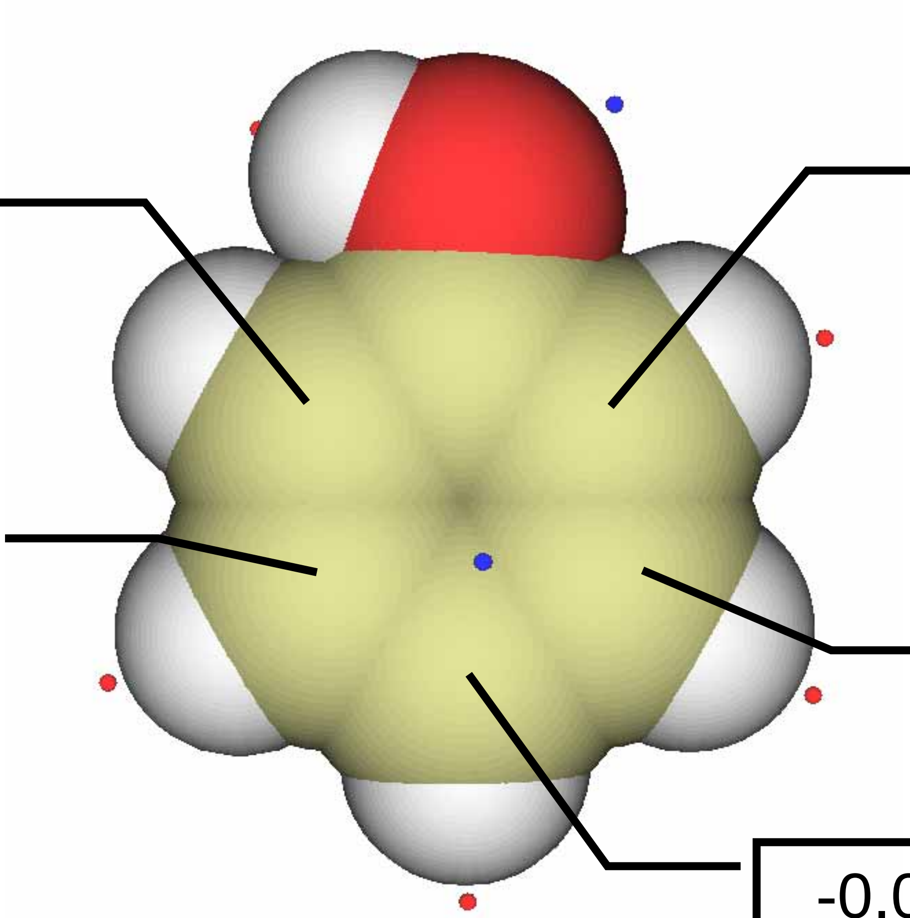
Phenol

Ambiguity of quantitative molecular surface analysis of ESP



Styrene

Examing ESP above 1.6 Å of molecular plane



-0.0181

-0.0225

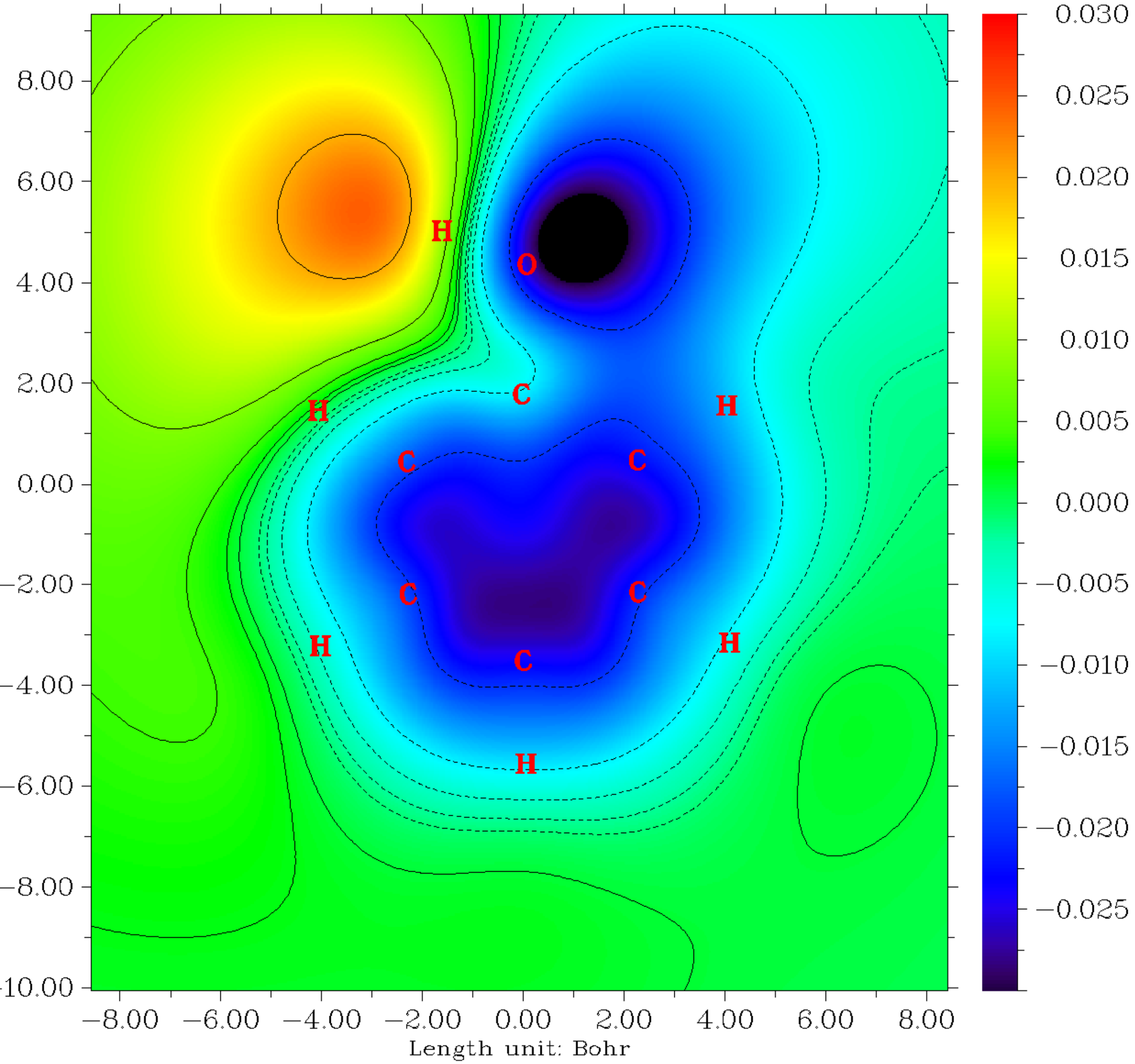
-0.0192

-0.0208

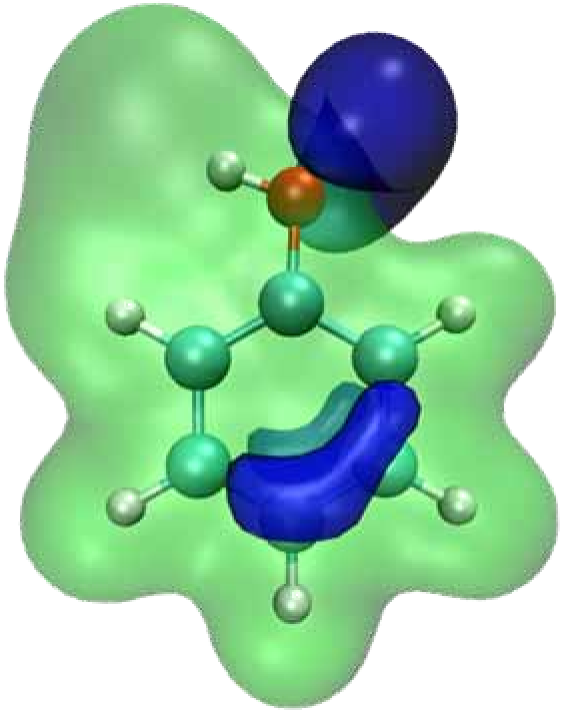
-0.0236

Avg. *ortho* position: -0.0203
Avg. *meta* position: -0.0200

ESP map above 1.6 Å of molecular plane



ESP isosurface (±0.027)

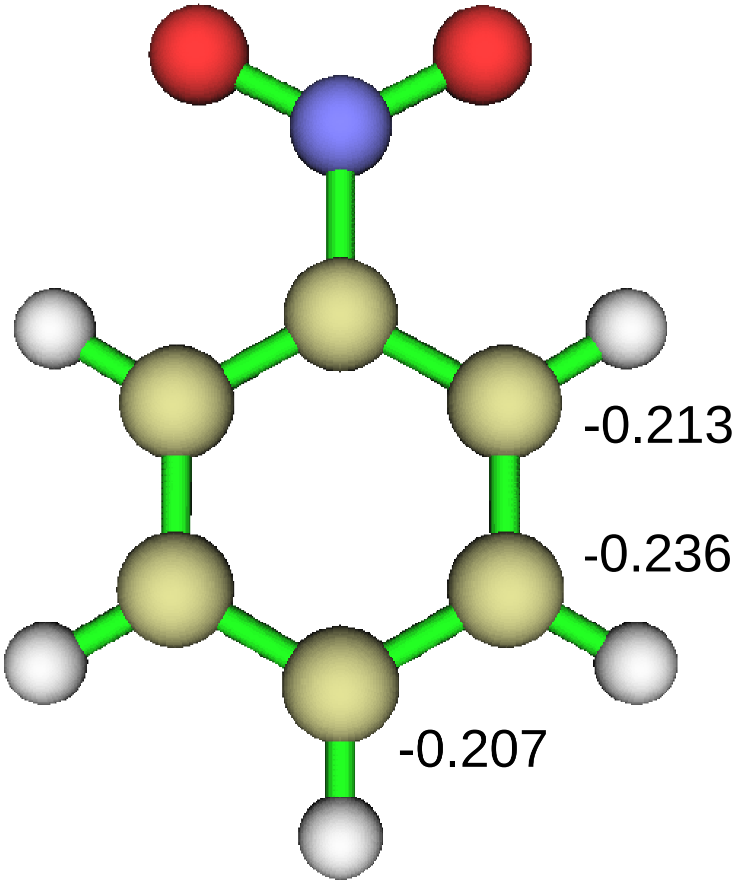


Atomic charges

- Point charge at nuclear position
- Non-observable
- No unique definition
- The most straightforward and simple model to exhibit charge distribution in chemical systems

Tian Lu, Feiwu Chen, Comparison of Computational Methods for Atomic Charges, *Acta Phys.-Chim. Sinica*, **28**, 1-18 (2012) (in Chinese)

An atom with more negative (positive) atomic charge is more likely to be the favorable site of electrophilic (nucleophilic) attack



Experimental yield：meta > ortho > para

Category of atomic charges

- Based on wavefunction

Mulliken, MMPA, Lowdin, RLPA, NPA, AOIM, Davidson, Hall, Sato, Thole, EMPA...

- Fitting ESP

Cox-Williams, Singh-Kollman, Besler-Merz-Kollman, CHELP, CHELPG, RESP...

- Based on partition of electron density in real space

Hirshfeld, Hirshfeld-I, **ADCH**, ISA, AIM, IPP, modified-Voronoi, VDD, Politzer...

- Based on electronegativity equilibrium principle

EEM, QEq, PEOE, PEPE, MPEOE...

- Others

CBAC, GDAC, GAPT, CM1/2/3, AM1-BCC, Del-Re...

ADCH: Tian Lu and Feiwu Chen, *J. Theor. Comput. Chem.*, **11**, 163 (2012)

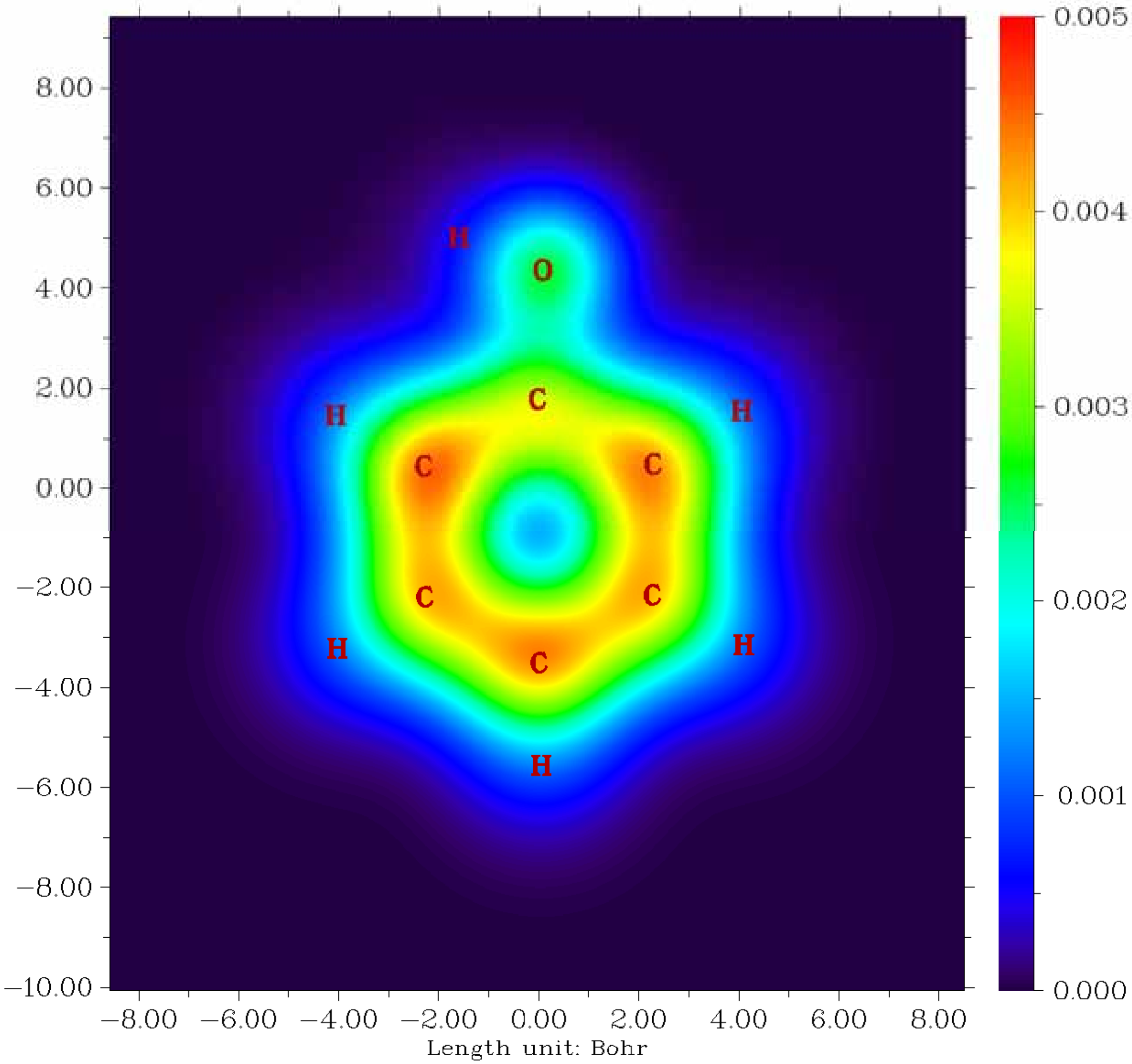
Population number of p_z orbital

Make the molecular plane parallel to XY plane, then the atom having larger population number on its p_z atomic orbital will have more tendency to be the favorable site of electrophilic attack

Unlike atomic charges, the influence of σ electrons are removed

The concept of atomic orbital must be recovered from *ab initio* wavefunction first. Natural population analysis (NPA) is recommended for this purpose

Electron density above 1.6Å of phenol plane



Not as sensitive to conformation as ESP

p$_z$ population

Avg. ortho=1.060
Avg. meta=0.985
Avg. para=1.036

Unable to faithfully show that reactivity of para is larger than ortho!

Average local ionization energy (ALIE)

$$\bar{I}(\mathbf{r}) = \frac{\sum_i \rho_i(\mathbf{r}) |\varepsilon_i|}{\rho(\mathbf{r})}$$

Lower value of ALIE indicates that the electrons at the point are more weakly bounded, and thus more active

Works well for predicting reactive site of electrophilic and radical reaction

Useless for study of nucleophilic attack

Politzer *et al.*, *J. Mol. Model*, **16**, 1731

Chapter 8 of the book *Theoretical Aspects of Chemical Reactivity (2007)*

J. Mol. Model, **9**, 342 (2003)

Local electron affinity

$$EA_{\mathrm{L}}(\mathbf{r}) = \frac{-\sum_{i \in vir} |\varphi_i(\mathbf{r})|^2 \varepsilon_i}{\sum_{i \in vir} |\varphi_i(\mathbf{r})|^2}$$

Similar to ALIE, but aims for studying nucleophilic attack. Not popular

Other related quantities

Local Mulliken electronegativity

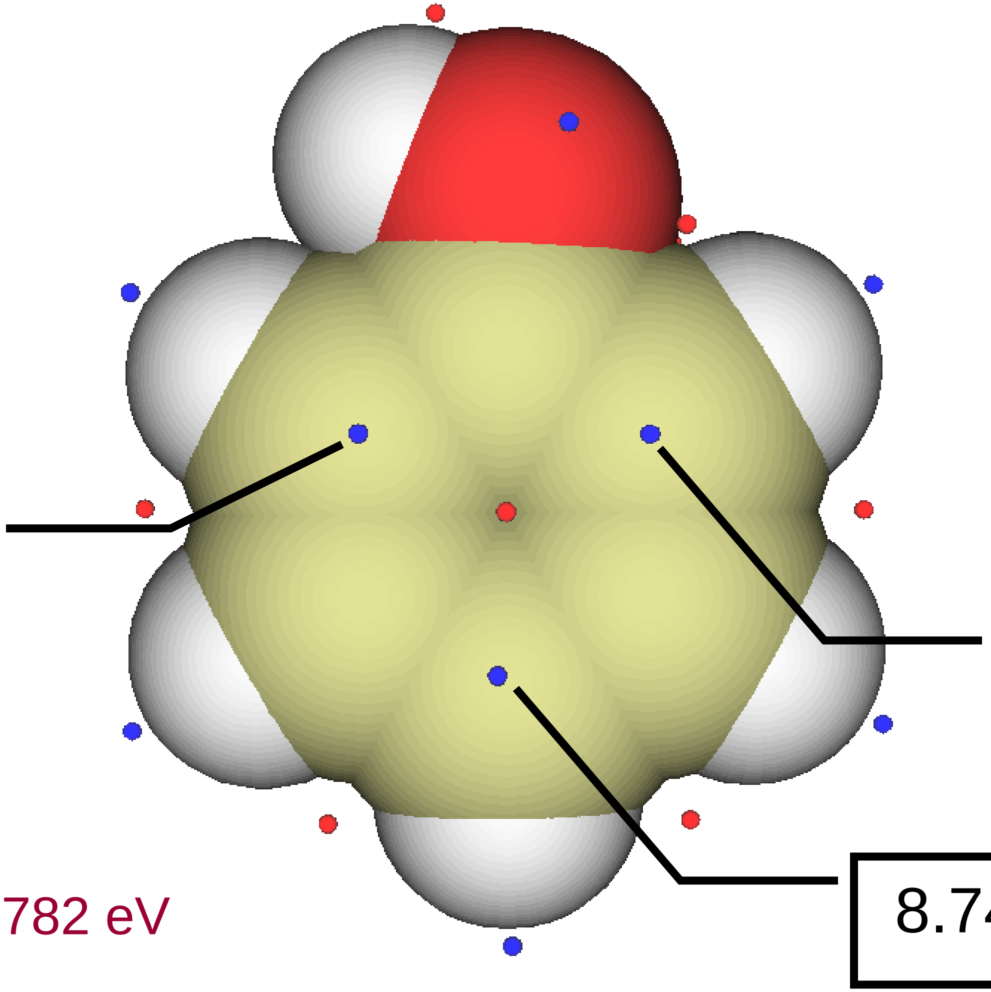
$$\chi_{\mathrm{L}}(\mathbf{r}) = [\bar{I}(\mathbf{r}) + EA_{\mathrm{L}}(\mathbf{r})]/2$$

Local hardness

$$\eta_{\mathrm{L}}(\mathbf{r}) = [\bar{I}(\mathbf{r}) - EA_{\mathrm{L}}(\mathbf{r})]/2$$

These quantities are meaningful only when minimal basis-set is used

Reactive site should be accompanied by a minimum of ALIE on vdW surface, and the atom closest to the global minimum can be predicted to be primary site



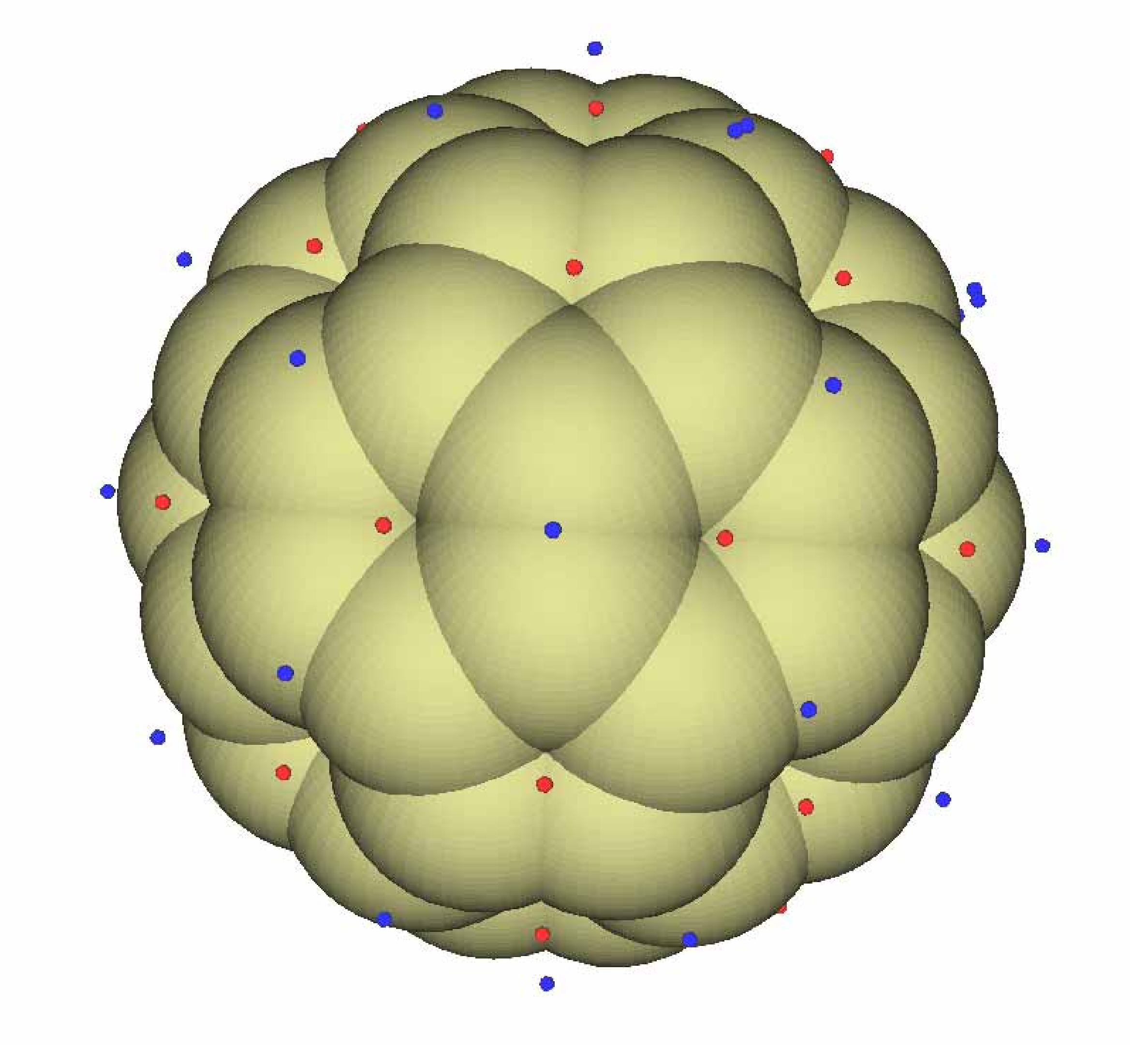
8.815eV

8.748eV

8.748eV

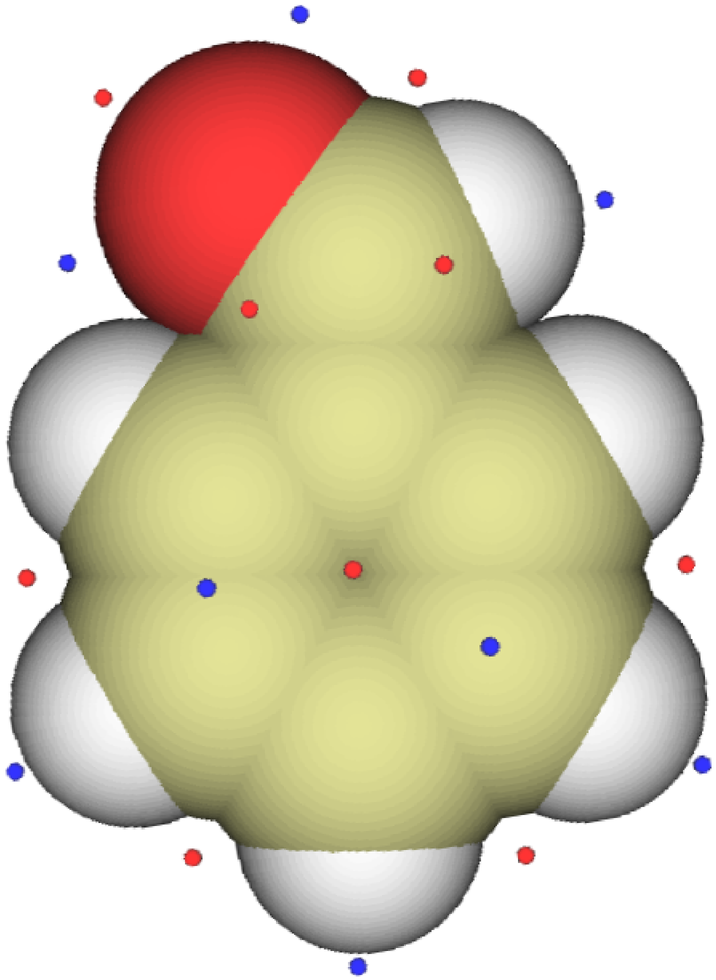
Avg. of ortho: 8.782 eV

Experimental yield：para > ortho, no meta



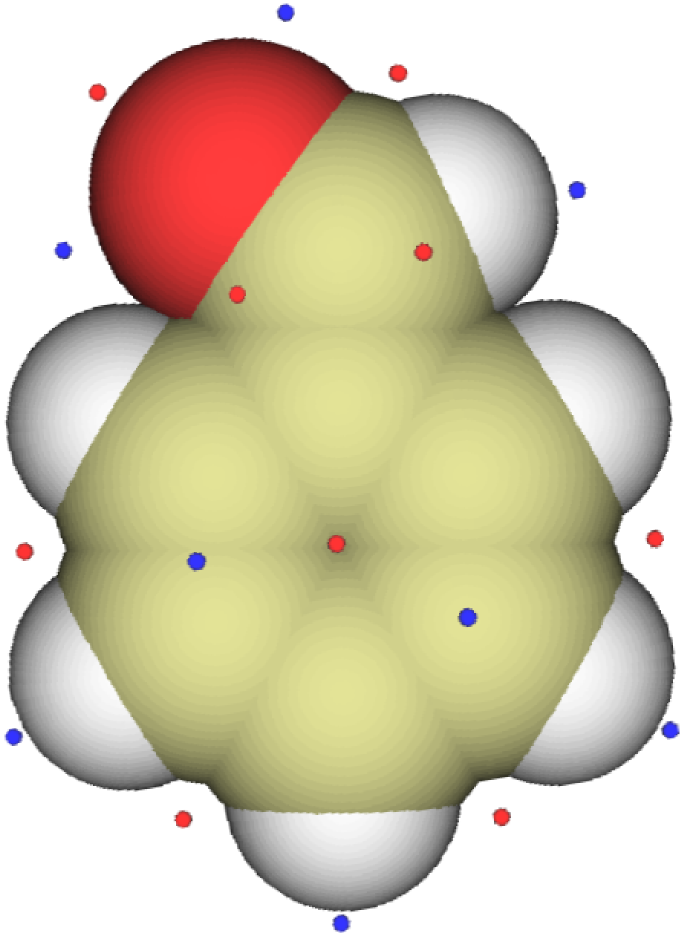
Fullerene, [6,6] bond is more reactive than [5,6]

Ambiguity of quantitative molecular surface analysis of ALIE



Benzaldehyde

ALIE above 1.6 Å of molecular plane

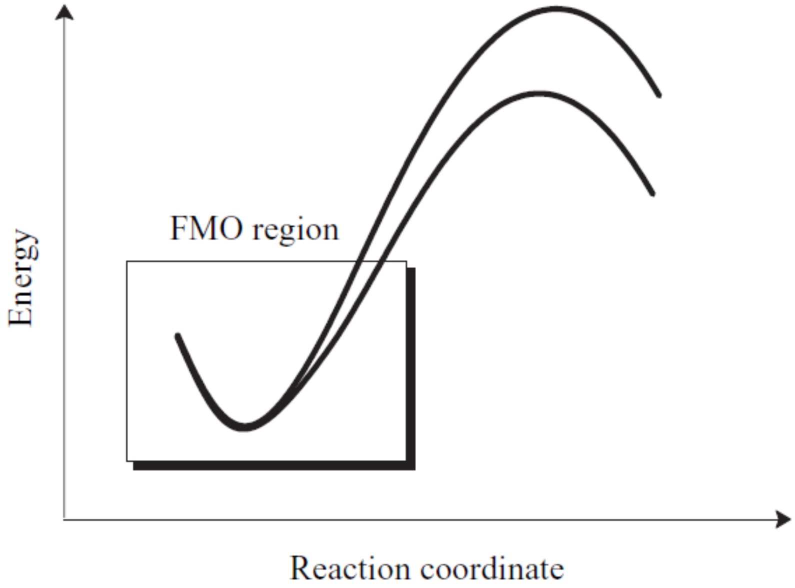


Avg. ortho: 0.3789

Avg. meta: 0.3752

Avg. para: 0.3795

Experimental yield：m >> o ≥ p



Frontier orbital molecular theory

The change in the energy can be derived from second-order perturbation theory

$$
\begin{aligned}
\Delta E = {} & \overbrace{-\sum_{\mathrm{A,B}}^{\text{atoms}} (\rho_{\mathrm{A}}+\rho_{\mathrm{B}})\langle \chi_{\mathrm{A}}|\mathbf{V}|\chi_{\mathrm{B}}\rangle\langle \chi_{\mathrm{A}}|\chi_{\mathrm{B}}\rangle}^{\text{Steric effect due to occupied MOs}} + \overbrace{\sum_{\mathrm{A,B}}^{\text{atoms}} \frac{Q_{\mathrm{A}}Q_{\mathrm{B}}}{R_{\mathrm{AB}}}}^{\text{Electrostatic interaction}} + \\
& \left. \left(\sum_{i\in \mathrm{A}}^{\substack{\text{occ.}\\ \text{MO}}} \sum_{a\in \mathrm{B}}^{\substack{\text{vir.}\\ \text{MO}}} + \sum_{i\in \mathrm{B}}^{\substack{\text{occ.}\\ \text{MO}}} \sum_{a\in \mathrm{A}}^{\substack{\text{vir.}\\ \text{MO}}} \right) \frac{2\left(\sum_{\alpha}^{\text{AO}} c_{\alpha i} c_{\alpha a} \langle \chi_{\alpha i}|\mathbf{V}|\chi_{\alpha a}\rangle\right)^2}{\varepsilon_i - \varepsilon_a} \right\} \text{Stablizing interaction due to mix of occupied and virtual MOs}
\end{aligned}
$$

Only HOMO-LUMO mix is assumed to be crucial

Atom with larger contribution to HOMO (LUMO) is more likely to be the preferential site of electrophilic (nucleophilic) attack

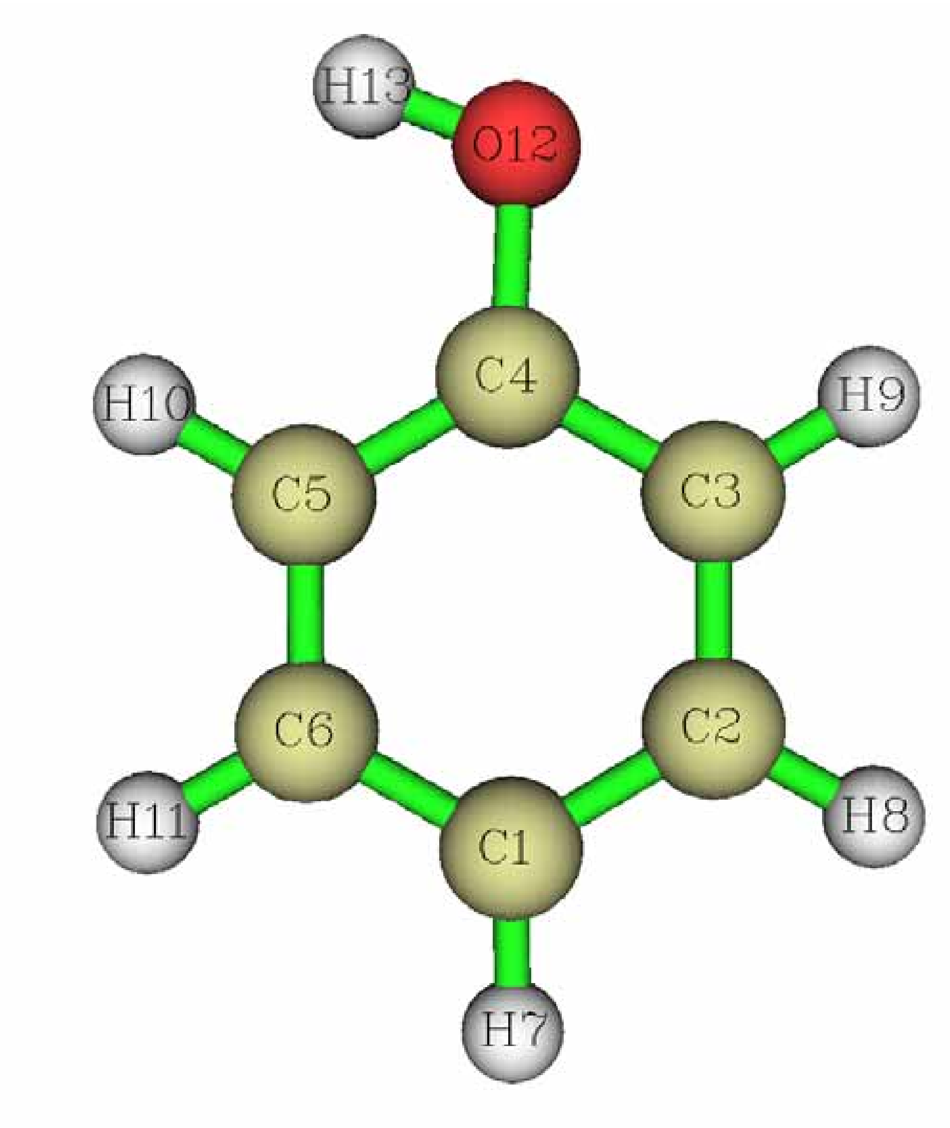
The ways to evaluate MO composition

- Mulliken analysis
- Modified Mulliken analysis (SCPA, Stout-politzer...)
- Natural atomic orbital (NAO) method
- Hirshfeld
- Becke

Tian Lu, Feiwu Chen, Calculation of Molecular Orbital Composition, *Acta Chim. Sinica*, **69**, 2393-2406 (2011) (in Chinese)

MO analysis (Becke)

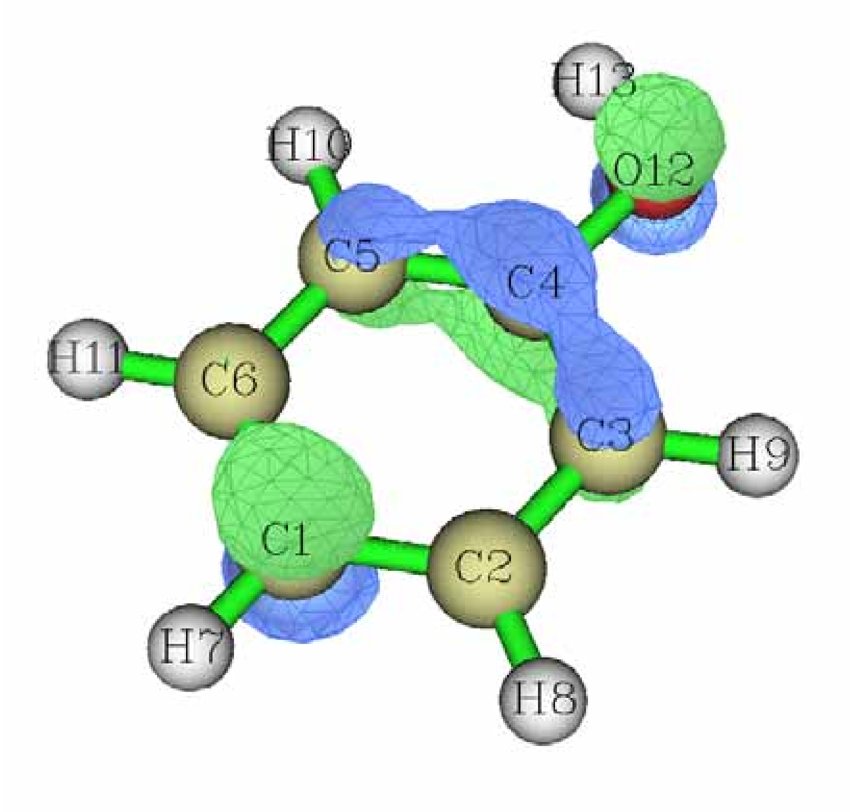
1(C) : 24.56%
2(C) : 4.85%
3(C) : 12.96%
4(C) : 18.44%
5(C) : 11.23%
6(C) : 6.77%

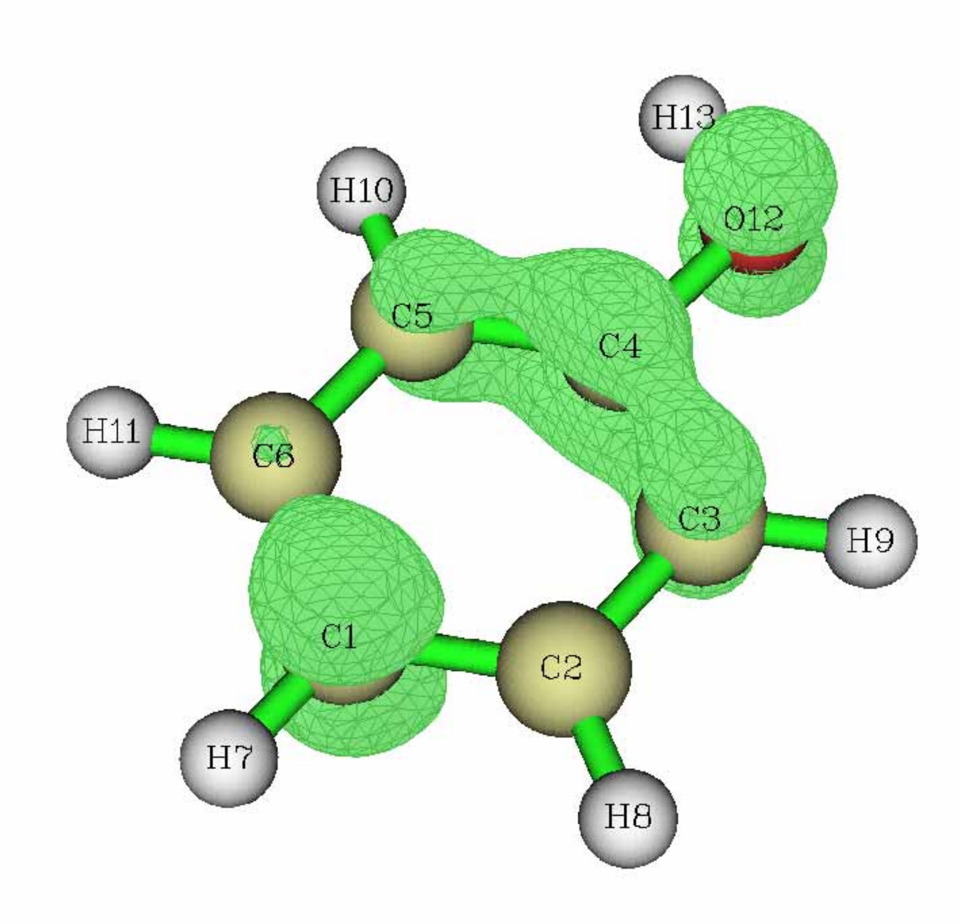


HOMO of phenol

HOMO density of phenol

$$\rho^{\mathrm{HOMO}} = 2\times|\varphi^{\mathrm{HOMO}}|^2$$

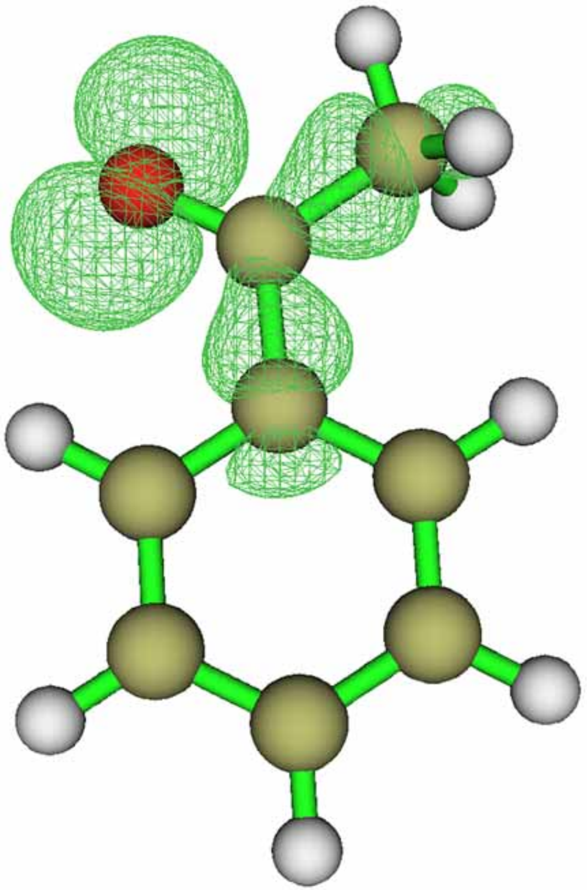




Frontier molecular orbital theory completely fails in two cases

1. **HOMO or LUMO is not located in interesting region**

2. **Occurrence of (near)degeneracy of frontier MOs**



J. Am. Chem. Soc., **106**, 4049 (1984)

Fukui function

$$f(\mathbf{r}) = \left[\frac{\partial \rho(\mathbf{r})}{\partial N}\right]_{\nu}$$

Constant nuclear coordinate

Nucleophilic attack $f^{+}(\mathbf{r}) = \rho_{N+1}(\mathbf{r}) - \rho_{N}(\mathbf{r}) \approx \rho^{\mathrm{LUMO}}(\mathbf{r})$

Electrophilic attack $f^{-}(\mathbf{r}) = \rho_{N}(\mathbf{r}) - \rho_{N-1}(\mathbf{r}) \approx \rho^{\mathrm{HOMO}}(\mathbf{r})$

Radical attack (May be ill defined due to the assumption of zero charge transfer)

$$f^{0}(\mathbf{r}) = \frac{f^{+}(\mathbf{r}) + f^{-}(\mathbf{r})}{2} = \frac{\rho_{N+1}(\mathbf{r}) - \rho_{N-1}(\mathbf{r})}{2} \approx \frac{\rho^{\mathrm{HOMO}}(\mathbf{r}) + \rho^{\mathrm{LUMO}}(\mathbf{r})}{2}$$

Condensed Fukui function

$$f_{\mathrm{A}}^{+} = p_{N+1}^{\mathrm{A}} - p_{N}^{\mathrm{A}}$$

$$q^{A} = Z^{\mathrm{A}} - p^{\mathrm{A}}$$

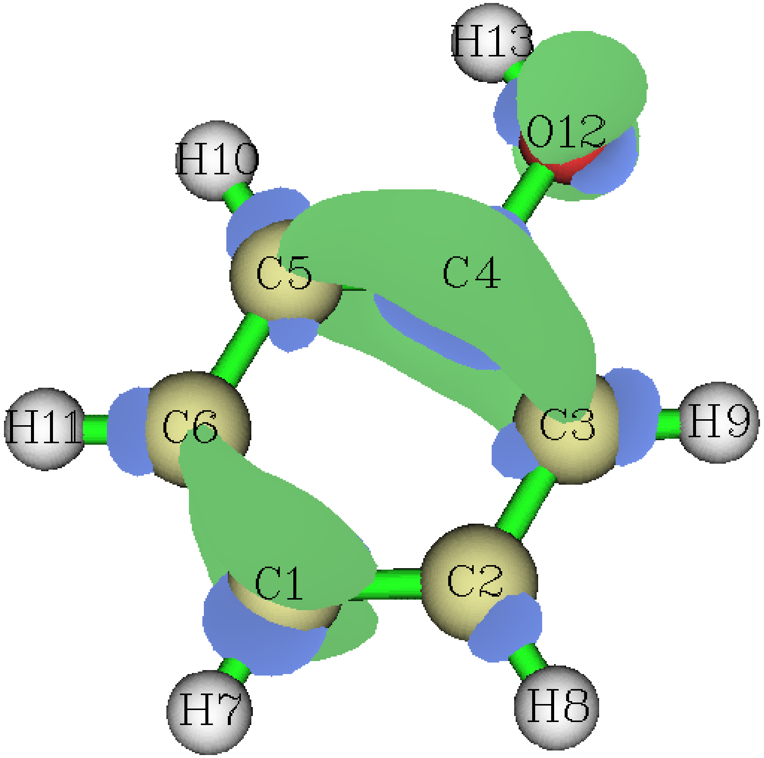
$$\text{Nucleophilic attack}: f_{\mathrm{A}}^{+} = q_{N}^{\mathrm{A}} - q_{N+1}^{\mathrm{A}}$$

$$\text{Electrophilic attack}: f_{\mathrm{A}}^{-} = q_{N-1}^{\mathrm{A}} - q_{N}^{\mathrm{A}}$$

$$\text{Radical attick}: f_{\mathrm{A}}^{0} = (q_{N-1}^{\mathrm{A}} - q_{N+1}^{\mathrm{A}})/2$$

Hirshfeld is one of the most ideal atomic charge model to evaluate cFF

	N	*N*-1	*N*+1	f^-	f^+
C1	-0.055	0.078	-0.092	**0.132**	0.038
C2	-0.020	0.033	-0.156	**0.053**	0.136
C3	-0.056	0.041	-0.200	**0.097**	0.144
C4	0.099	0.251	0.061	**0.152**	0.038
C5	-0.071	0.018	-0.208	**0.089**	0.137
C6	-0.022	0.046	-0.159	**0.068**	0.137



$$f^- = 0.007$$

To compare reactivity of sites in different molecules, global softness must be taken into account, which measure electron reactivity of the whole molecule and closely relates to polarizability

Global softness

$$S = \left(\frac{\partial N}{\partial \mu}\right)_{\nu} = \frac{1}{\eta}$$

Hardness η = IP - EA

Local softness

$$s(\mathbf{r}) = \left[\frac{\partial \rho(\mathbf{r})}{\partial \mu}\right]_{\nu} = \left(\frac{\partial N}{\partial \mu}\right)_{\nu}\left[\frac{\partial \rho(\mathbf{r})}{\partial N}\right]_{\nu} = S \cdot f(\mathbf{r})$$

Condensed local softness

$$s_{\mathrm{A}} = S \cdot f_{\mathrm{A}}$$

J. Phys. Chem. A, **109**, 205

Dual descriptor

$$\begin{aligned}\Delta f(\mathbf{r}) &= f^{+}(\mathbf{r}) - f^{-}(\mathbf{r}) \\ &= \left[\rho_{N+1}(\mathbf{r}) - \rho_{N}(\mathbf{r})\right] - \left[\rho_{N}(\mathbf{r}) - \rho_{N-1}(\mathbf{r})\right] \\ &= \rho_{N+1}(\mathbf{r}) - 2\rho_{N}(\mathbf{r}) + \rho_{N-1}(\mathbf{r})\end{aligned}$$

Approximation form based on spin density

$$\rho_{N+1}^{s} \approx \rho_{N+1} - \rho_{N}$$

$$\rho_{N-1}^{s} \approx \rho_{N} - \rho_{N-1}$$

$$\Delta f(\mathbf{r}) \approx \rho_{N+1}^{s}(\mathbf{r}) - \rho_{N-1}^{s}(\mathbf{r})$$

Condensed dual descriptor

$$\Delta f_{\mathrm{A}} = f_{\mathrm{A}}^{+} - f_{\mathrm{A}}^{-} = 2q_N^{\mathrm{A}} - q_{N+1}^{\mathrm{A}} - q_{N-1}^{\mathrm{A}}$$

Approximation form based on spin population p^s

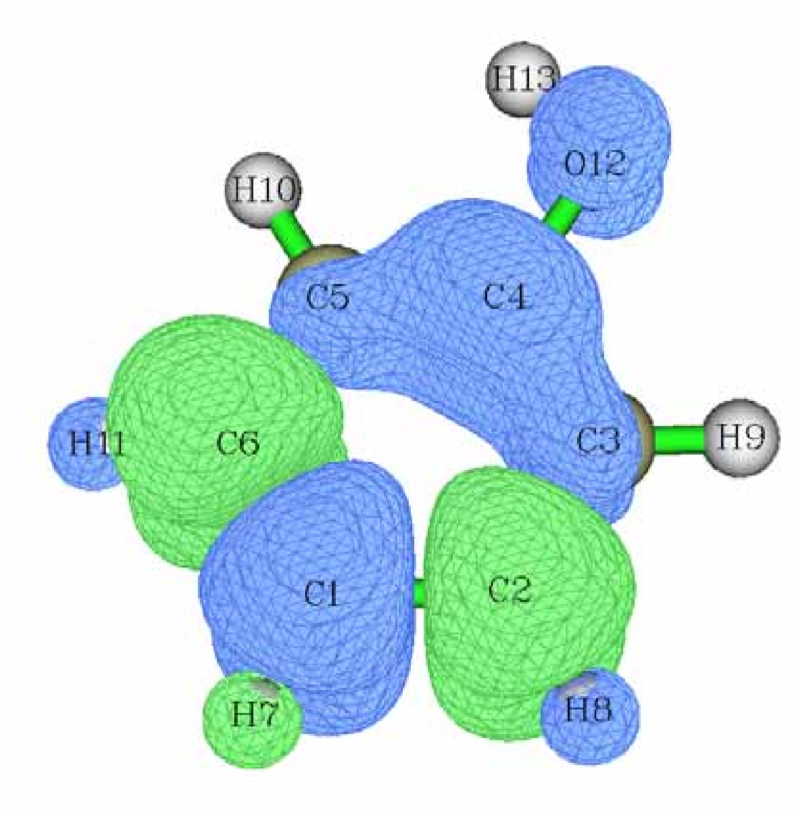
$$\Delta f_{\mathrm{A}} = p_{N+1}^{s} - p_{N-1}^{s}$$

Both types of reactive sites can be revealed simultaneously

$\Delta f > 0$: Favorable for nucleophilic attack
$\Delta f < 0$: Favorable for electrophilic attack

In practice, the sign is not critical. The more positive (negative) the dual descriptor at a site, the more tendency that the site reacts with a nucleophile (electrophile)

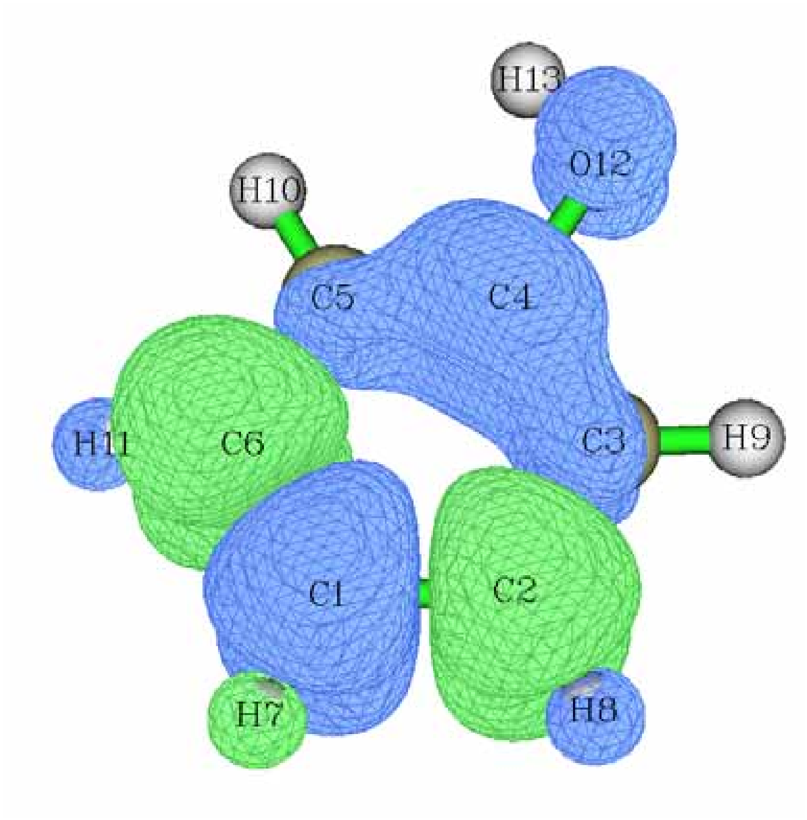
	N	*N*-1	*N*+1	Δf
C1	-0.055	0.078	-0.092	**-0.095**
C2	-0.020	0.033	-0.156	**0.083**
C3	-0.056	0.041	-0.200	**0.048**
C4	0.099	0.251	0.061	**-0.114**
C5	-0.071	0.018	-0.208	**0.048**
C6	-0.022	0.046	-0.159	**0.069**



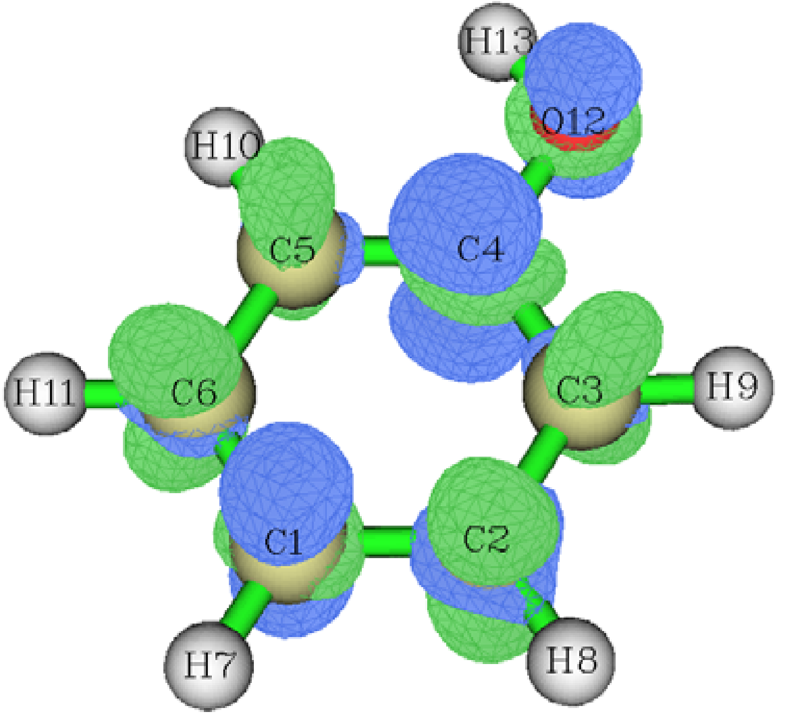
$$\Delta f = 0.005$$

Evaluated based on spin density

Exact form of dual descriptor is not necessarily better than the approximation form



$$\rho^{s}_{N+1}(\mathbf{r}) - \rho^{s}_{N-1}(\mathbf{r})$$



$$\rho_{N+1}(\mathbf{r}) - 2\rho_{N}(\mathbf{r}) + \rho_{N-1}(\mathbf{r})$$

Relative electrophilicity s_k^+ / s_k^-

Relative nucleophilicity s_k^- / s_k^+

The site having the highest s_k^+/s_k^- is the most probable site to be attacked by a nucleophile, and the site having the highest s_k^-/s_k^+ ratio is the most probable site to be attacked by an electrophile

Not a good idea, rarely used!

Rong Fu, Tian Lu, Feiwu Chen, Comparison of the Methods for Predicting the Reactive Site of Electrophilic Substitution Reaction, *Acta Phys.-Chim. Sinica*, **30**, 628-639 (2014) (in Chinese)

Considered methods

The methods focusing on electron softness

1. Fukui function isosurface
2. Condensed Fukui function derived by Hirshfeld charge
3. Dual descriptor isosurface
4. Condensed dual descriptor derived by Hirshfeld charge
5. HOMO composition derived by NAO method
6. Extrema of ALIE on vdW surface
7. ALIE above 1.6 Angstrom of molecular plane

The methods focusing on electrostatic effect

1. Extrema of ESP on vdW surface
2. ESP above 1.6 Angstrom of molecular plane
3. Population number of p_z atomic orbital
4. Hirshfeld, CHELPG, Mulliken and NPA charges

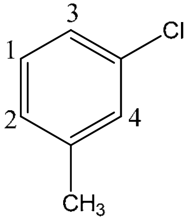
Test systems

Monosubstituted

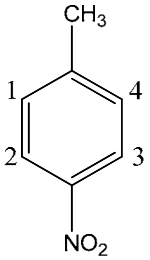
ortho-para directors:
O^-, OH, NH_2, $CHCH_2$, $NHCOCH_3$, F, Cl, CH_3
meta directors:
N, CHO, NO_2, SO_3H, $COCH_3$, CF_3

Disubstituted

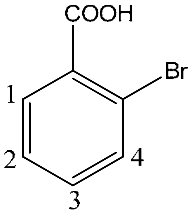
Synergistic directing effect



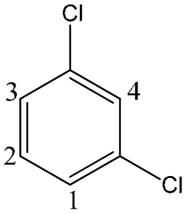
S1



S2

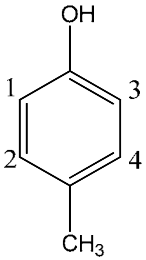


S3

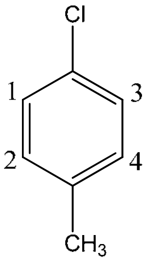


S4

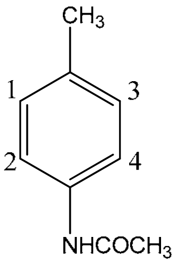
Conflicting directing effect



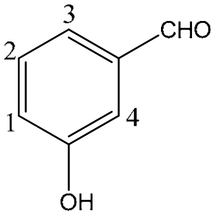
C1



C2



C3



C4

Main conclusions after a thorough comparsion

Dual descriptor has the best overall performance

For single ortho-para director and dual directors cases, the methods focusing on electron softness work well. However, the ones representing electrostatic effects do not show satisfactory prediction ability

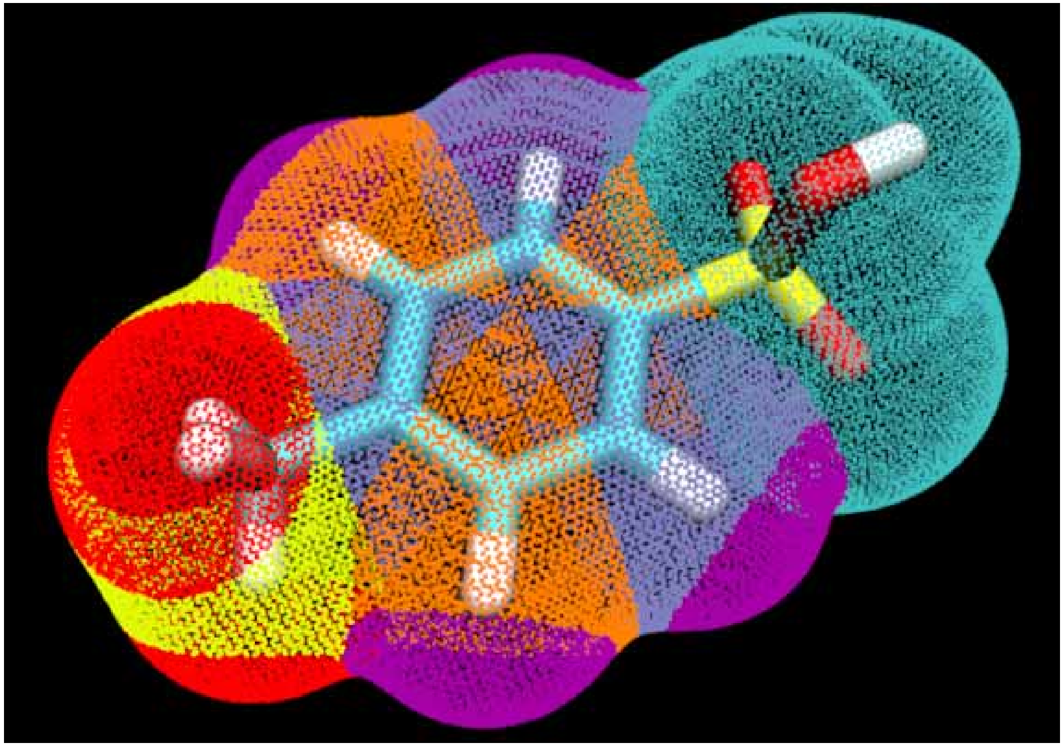
For single meta director cases, the situation is inverted. Though the prediction ability of the methods representing electrostatic effects are still unsatisfactory, they are even slightly better than the other methods

p_z-population performs better than all kinds of atomic charges (in which Mulliken is the worse one)

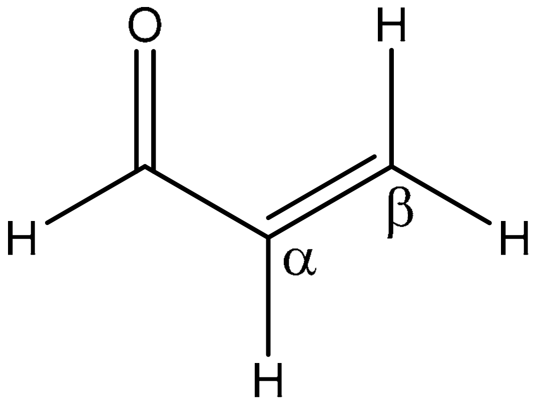
My novel ways to characterize real space function distribution

Novel way 1: Local molecular surface analysis

A novel way to study distribution of ESP, ALIE, Fukui function, dual descriptor and HOMO/LUMO on each site

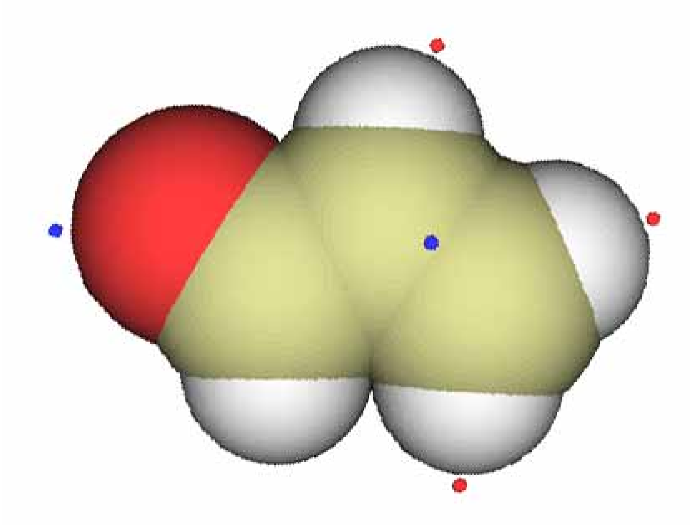


Case study: Acrolein



Preferential sequence of nucleophilic attack

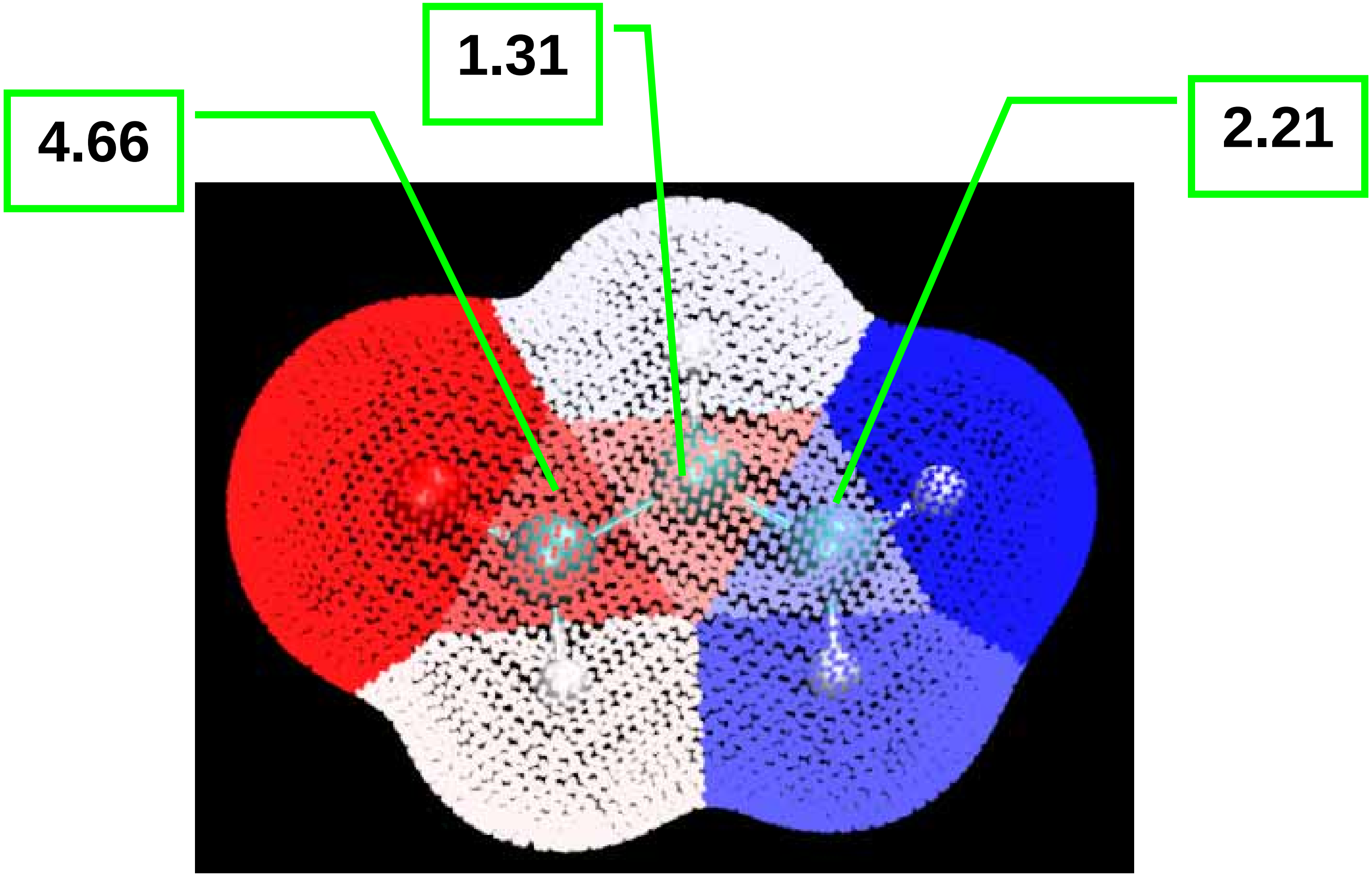
Carboxyl carbon > β > α



ESP extrama on molecular surface

Completely useless!

Average of ESP on local molecular surface of each carbon (a.u.)



4.66

1.31

2.21

In complete agreement with experimental observation!

Novel way 2: Fitting vdW surface distribution

Very similar to deriving atomic charges by fitting ESP, but the function is not limited to ESP

Exact value

$$F(p_1, p_2 \ldots p_N) = \sum_i [V(\mathbf{r}_i) - V'(\mathbf{r}_i)]^2$$

Atomic values

$$V'(\mathbf{r}_i) = \sum_A \frac{p_A}{r_{A,i}}$$

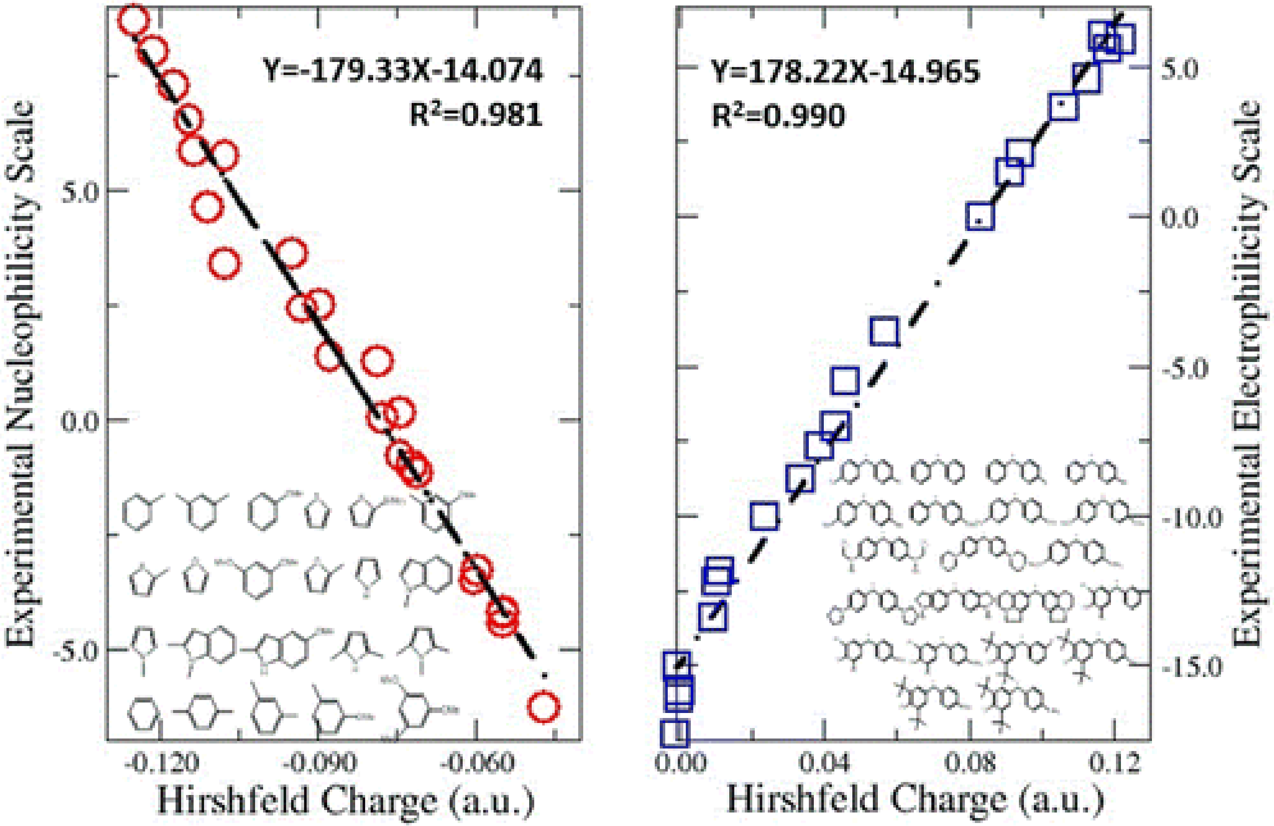
Approximate value reproduced by atomic values

Minimizing error function *F* to obtain atomic values

The fitting points are distributed over vdW surface

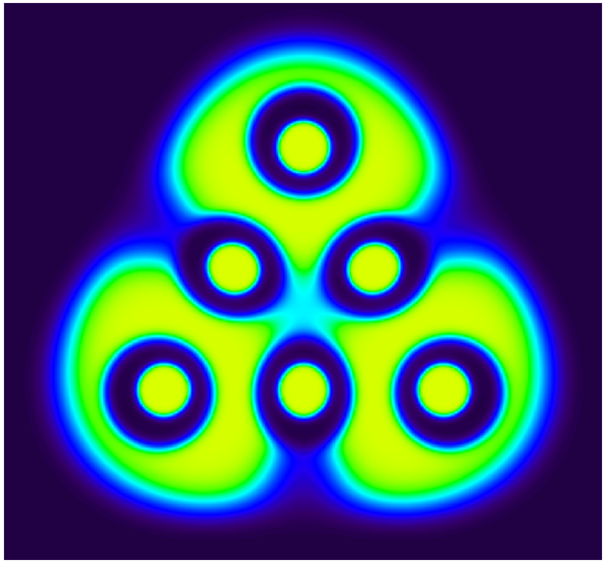
	Qualitative		Quantitative		
	Isosurface	**Map onto vdW surface**	**Surface extrama analysis**	**Fit vdW surface distribution**	**Obtain atomic value**
HOMO density	Important	?	?	?	Orbital composition analysis
Fukui function	Important	?	*Not good*	*Not good*	Condensed Fukui function
Dual descriptor	Important	?	?	?	Condensed dual descriptor
ALIE		Useful	Important	Possible	
Electrostatic	*Not good*	Important	Important	CHELPG, MK	Atomic charges

Experimental nucleophilicity scale and electrophilicity scale are suprisingly closely related to Hirshfeld charge!



Shubin Liu, Chunying Rong, Tian Lu, Information Conservation Principle Determines Electrophilicity, Nucleophilicity, and Regioselectivity, *J. Phys. Chem. A*, **118**, 3698-3704 (2014)

All of the analyses can be easily finished by Multiwfn! see Section 4.A.4 of the manual for summary



Multiwfn official site:

http://sobereva.com/multiwfn

The significance, functions and uses of multifunctional wavefunction analysis program Multiwfn" (in Chinese)
http://sobereva.com/184
Tips for getting start with Multiwfn" (in Chinese)
http://sobereva.com/167