ParaSurf ™ and ParaFit™

Tim Clark

Cepos InSilico Ltd
The Old Vicarage
Bedford Road
Kempston
Bedford, MK42 8BQ
United Kingdom

Cepos InSilico GmbH
Computer-Chemie-Centrum
Nägelsbachtraße 25
91052 Erlangen
Germany

Cepos InSilico

- Cepos InSilico is a UK limited company located in Kempston, Beds
  - Company Secretary Paul Hillier
- Close links to the Universities of Erlangen and Portsmouth
- 100% owned German subsidiary (Cepos InSilico GmbH) located in Erlangen formed in December 2009
  - CEO Dr. Ute Seidel

News

From March 1st 2010, Cepos products will be distributed in North America by Cache Research LLC, Beaverton, Oregon.

http://cacheresearch.com/home.html

More News

Cepos InSilico and InhibOx have agreed a joint distribution and technical cooperation effective immediately

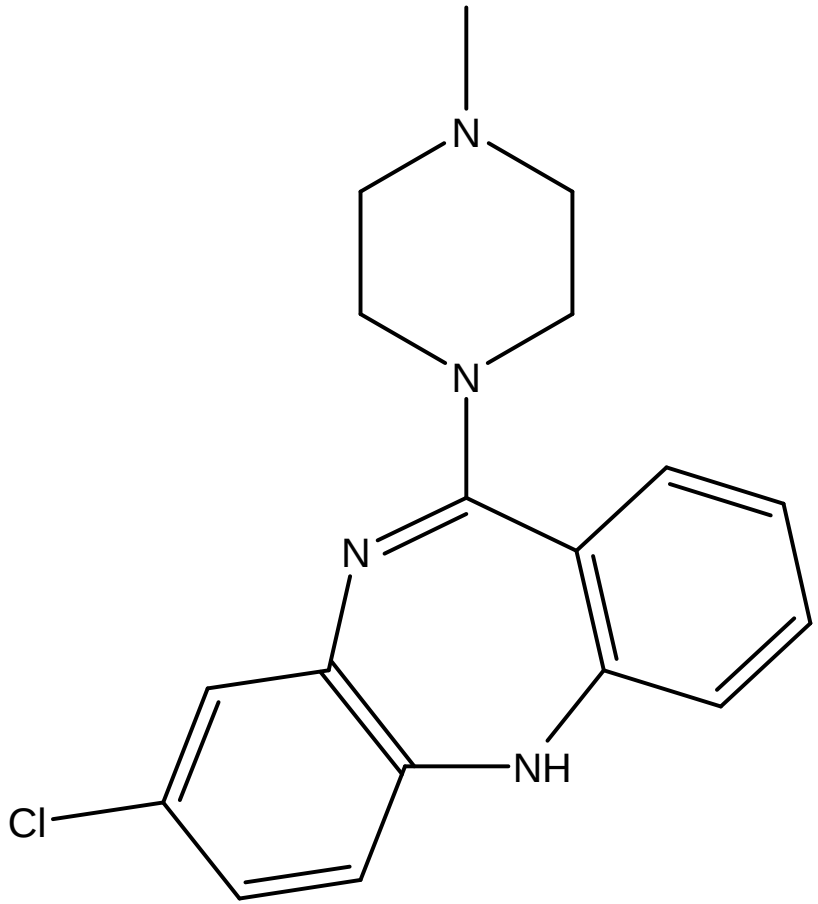
http://www.inhibox.com/index.html

Even More News

Cepos InSilico and Molcad GmbH have agreed a joint distribution and technical cooperation effective December 1st 2010.

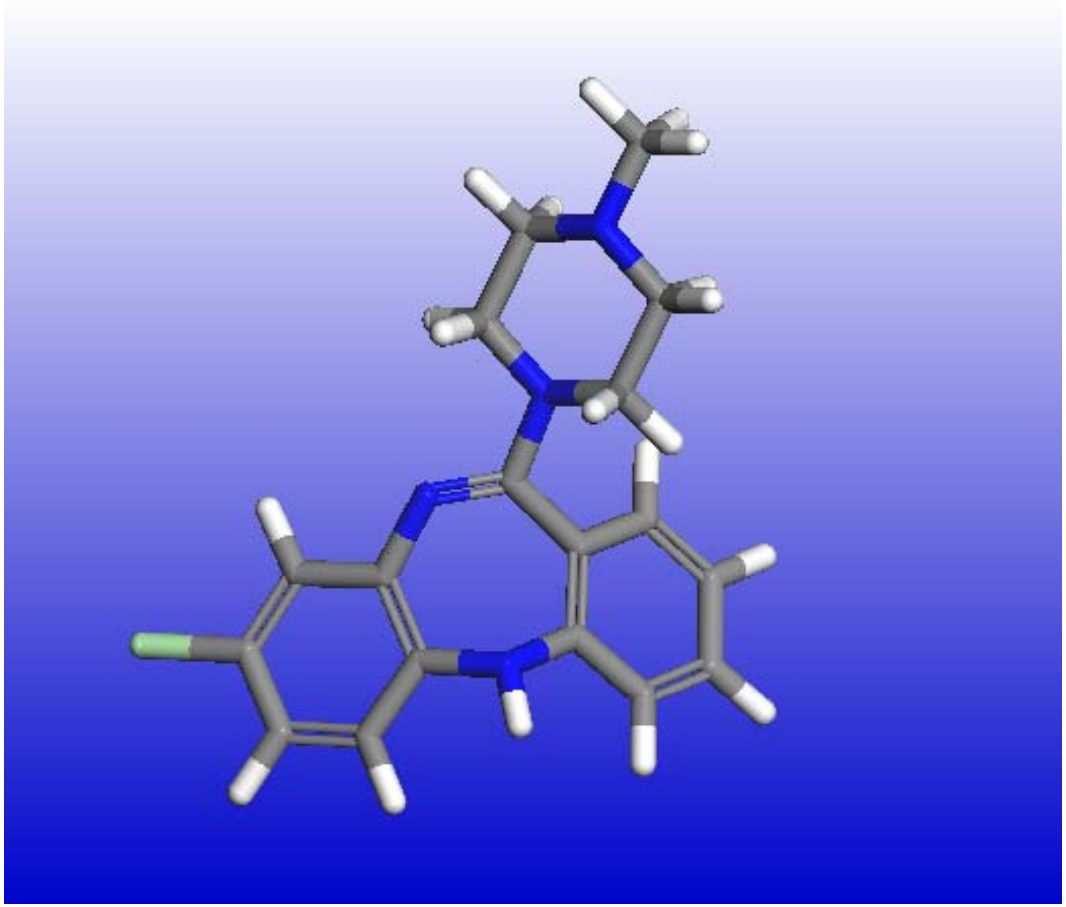
http://www.molcad.de

Descriptions of Molecules



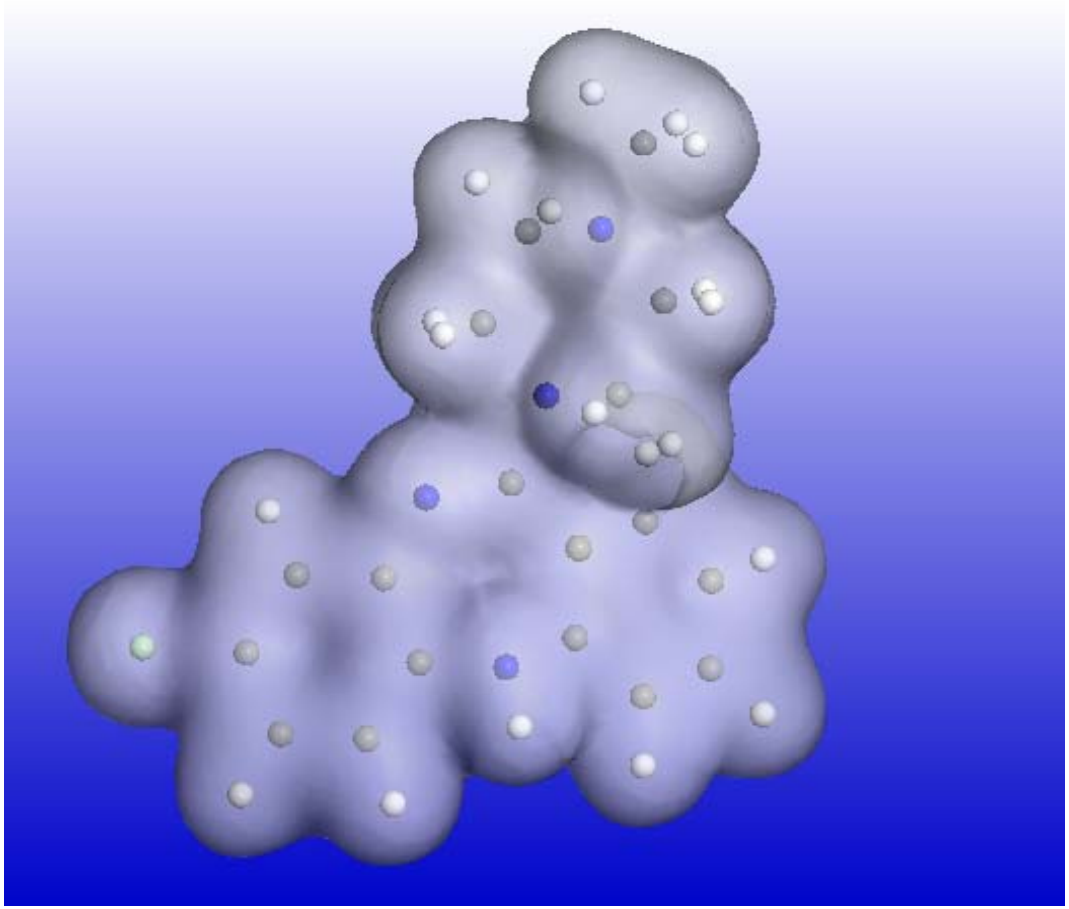
CEPOS

Descriptions of Molecules



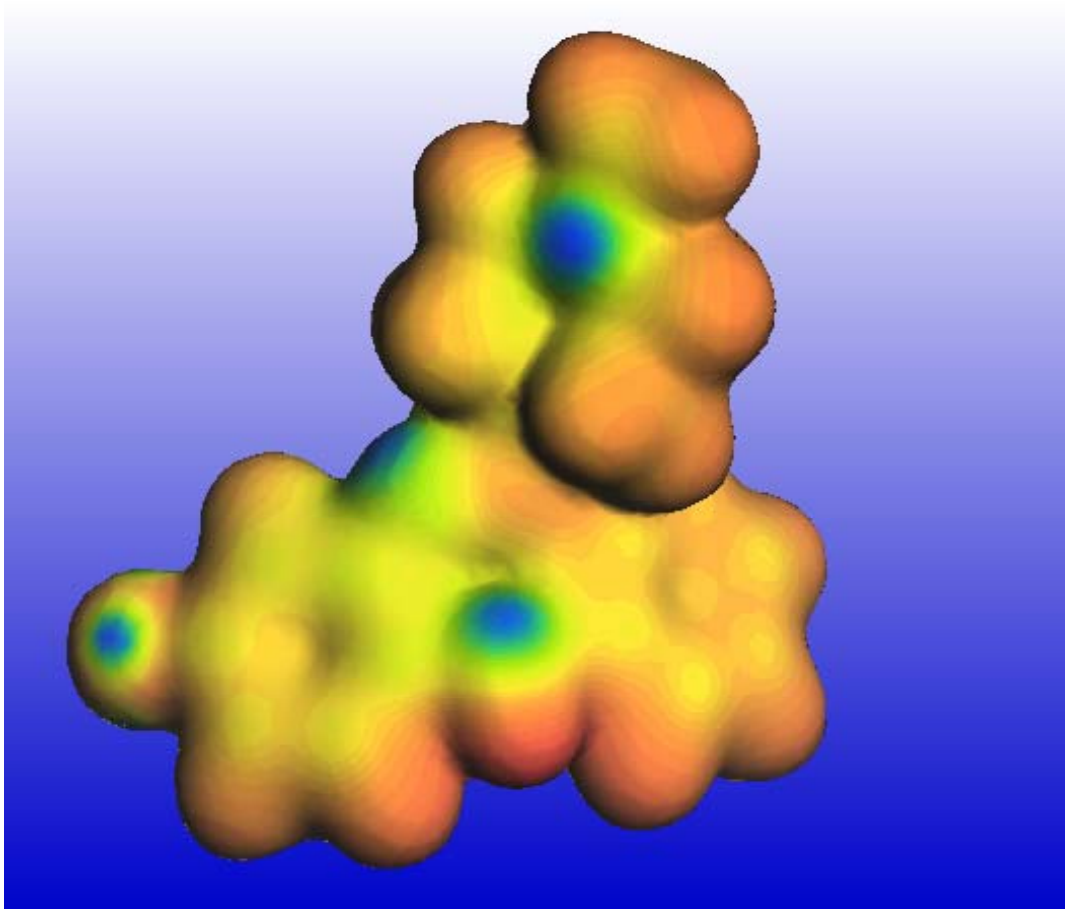
CEPOS

Descriptions of Molecules



CEPOS

Descriptions of Molecules



Surface Description

- **ParaSurf™ describes intermolecular interactions completely, including**
  - The shape (repulsions, electron density)
  - Intermolecular interactions
    - Coulomb
    - Donor/Acceptor
    - Acceptor/Donor
    - Dispersion

CEPOS

Local Properties

- To do this, it uses a series of local properties derived from semiempirical MO theory (usually AM1).
- These are:
  - The electron density
  - The Molecular Electrostatic Potential and the electrostatic field normal to the surface
  - The local ionization energy (Politzer et al.)
  - Local electron affinity
  - The local polarizability
  - Local hardness and local electronegativity (derived from the ionization energy and electron affinity)

Local Ionization Energy

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

Sjoberg, P.; Murray, J. S.; Brinck, T.; Politzer, P. A., *Can. J. Chem.* 1990, **68**, 1440;

Murray, J. S.; Abu-Awwad, F.; Politzer, P., THEOCHEM 2000, **501-502**, 241;

Hussein, W.; Walker, C. G.; Peralta-Inga, Z.; Murray, J. S., *Int. J. Quant. Chem.* 2001, **82**, 160;

Politzer, P.; Murray, J. S.; Concha, M. C., *Int. J. Quant. Chem.* 2002, **88**,19.

Other Local Properties

- Local Electron affinity:

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

- Local Hardness:

$$\eta_L(\mathbf{r}) = \frac{\left(IP_L(\mathbf{r}) - EA_L(\mathbf{r})\right)}{2}$$

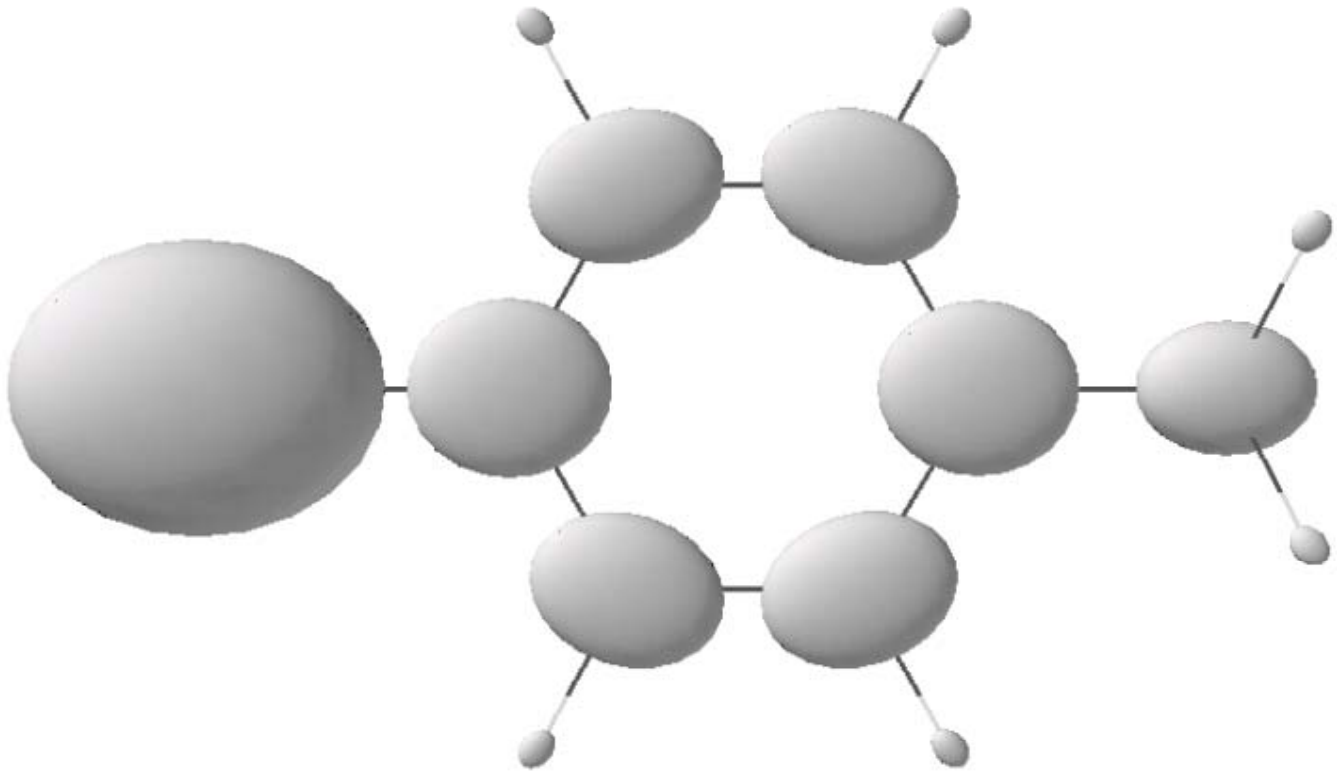
- Local electronegativity

$$\chi_L(\mathbf{r}) = \frac{\left(IP_L(\mathbf{r}) + EA_L(\mathbf{r})\right)}{2}$$

B. Ehresmann, B. Martin, A. H. C. Horn and T. Clark, *J. Mol. Model.* **2003**, *9*, 342-347.

B. Ehresmann, M. J. de Groot, A. Alex and T. Clark, *J. Chem. Inf. Comp. Sci.* **2004**, *44*, 658-668.

Local Polarizability



G. Schürer, P. Gedeck, M. Gottschalk, T. Clark, *Int. J.Quant. Chem.*, **1999**, *75,* 17-31; B. Martin, P. Gedeck, T. Clark, *Int. J. Quant. Chem.*, **2000**, *77*, 473.

Local Polarizability

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

$\rho_j^1(\mathbf{r})$ Density due to a singly occupied atomic orbital *j*

q_j Coulson population of atomic orbital *j*

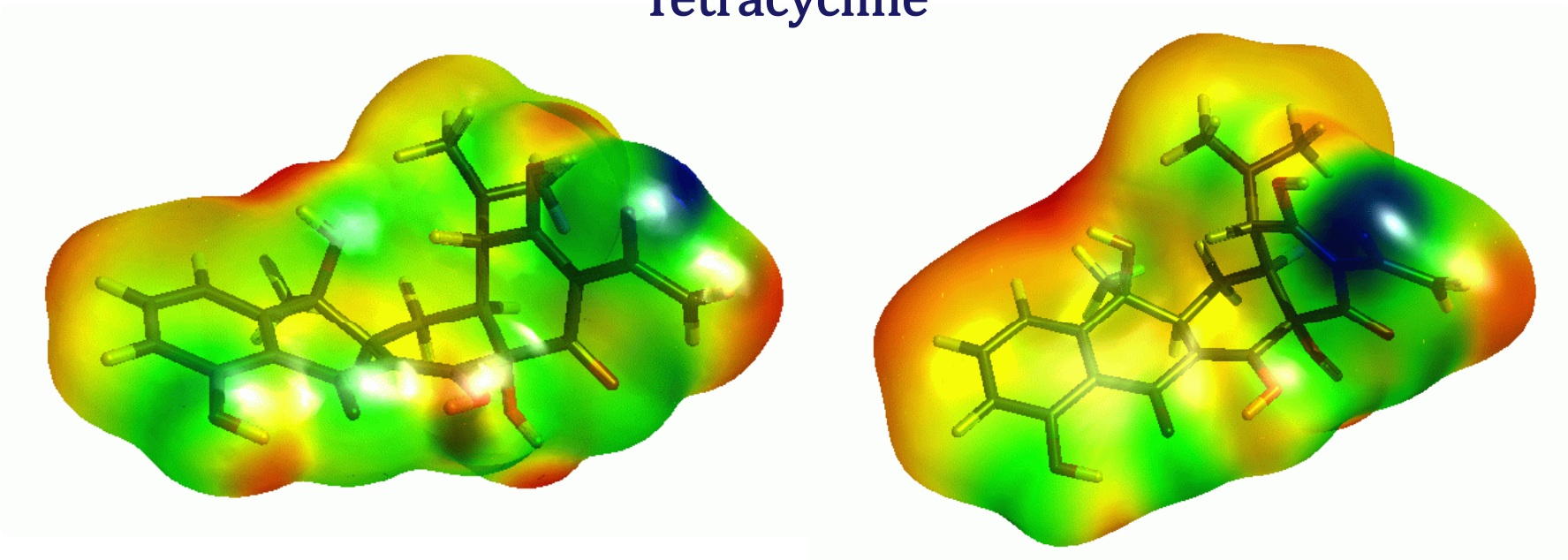
$\bar{\alpha}_j$ Mean polarizability calculated for atomic orbital *j*

B. Ehresmann, B. Martin, A. H. C. Horn and T. Clark, *J. Mol. Model.* **2003**, *9*, 342-347.

B. Ehresmann, M. J. de Groot, A. Alex and T. Clark, *J. Chem. Inf. Comp. Sci.* **2004**, *44*, 658-668.

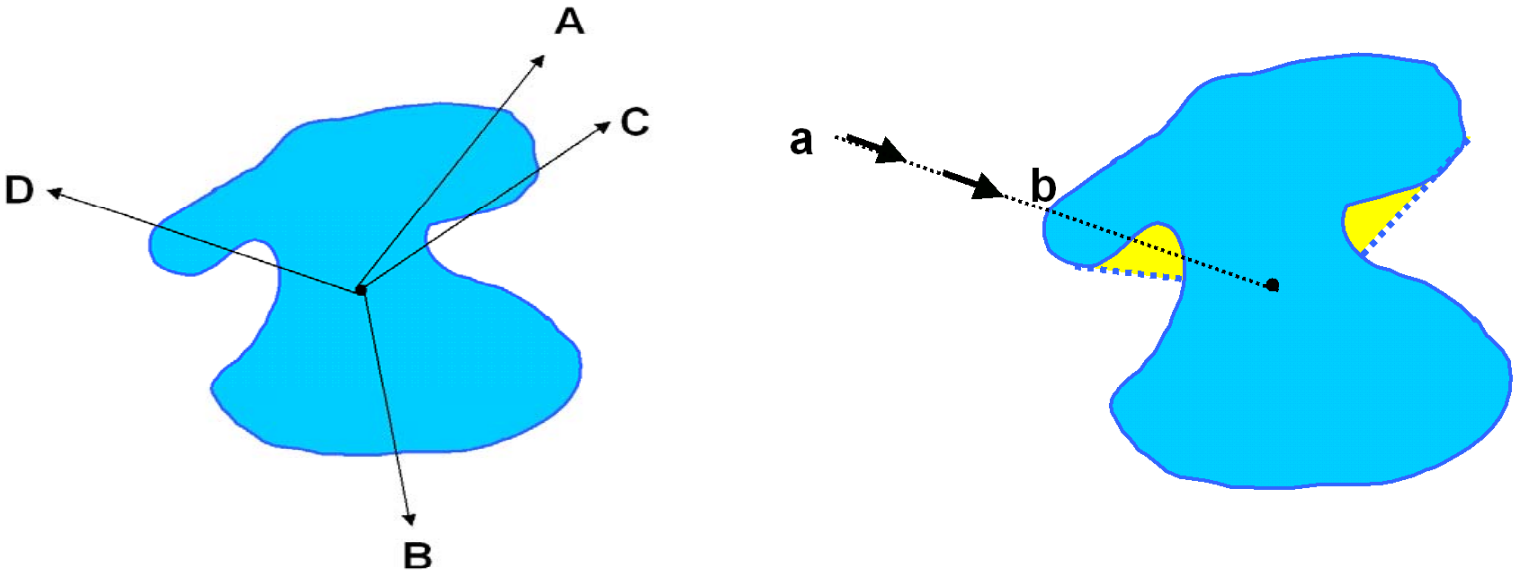
Alternative Surfaces

Tetracycline



Marching-cube surface. Can be based on the electron density (isodensity surface) or on the solvent-excluded surface using van der Waals´ radii

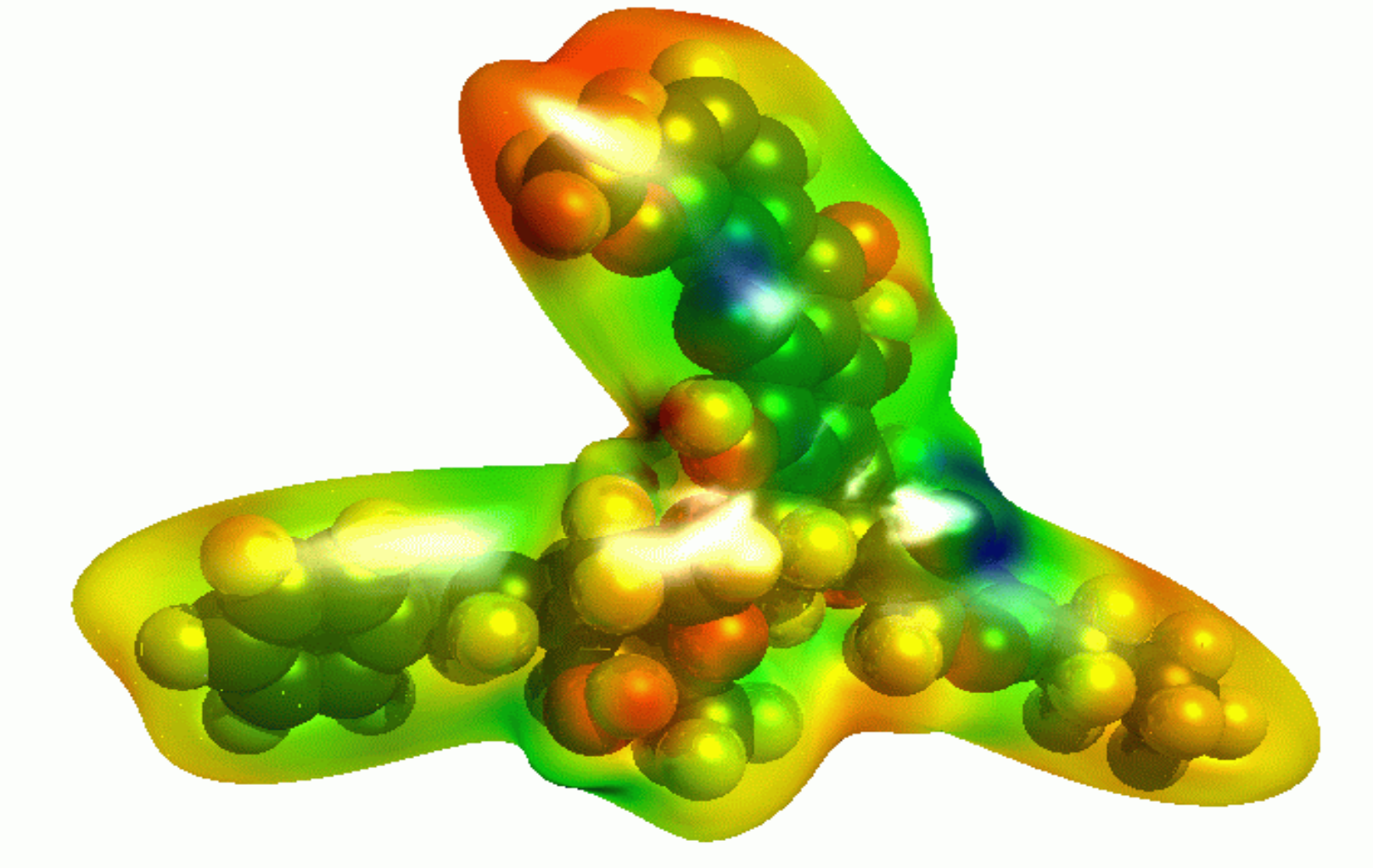
Alternative Surfaces



Shrink-wrap surface. Can also be based on the electron density (isodensity surface) or on the solvent-excluded surface using van der Waals´ radii. These surfaces are single-valued along any given radial vector to the center, so may not “fit” as well as marching cube (yellow areas). However, they can be represented analytically.

CEPOS

A Shrink-Wrap Surface

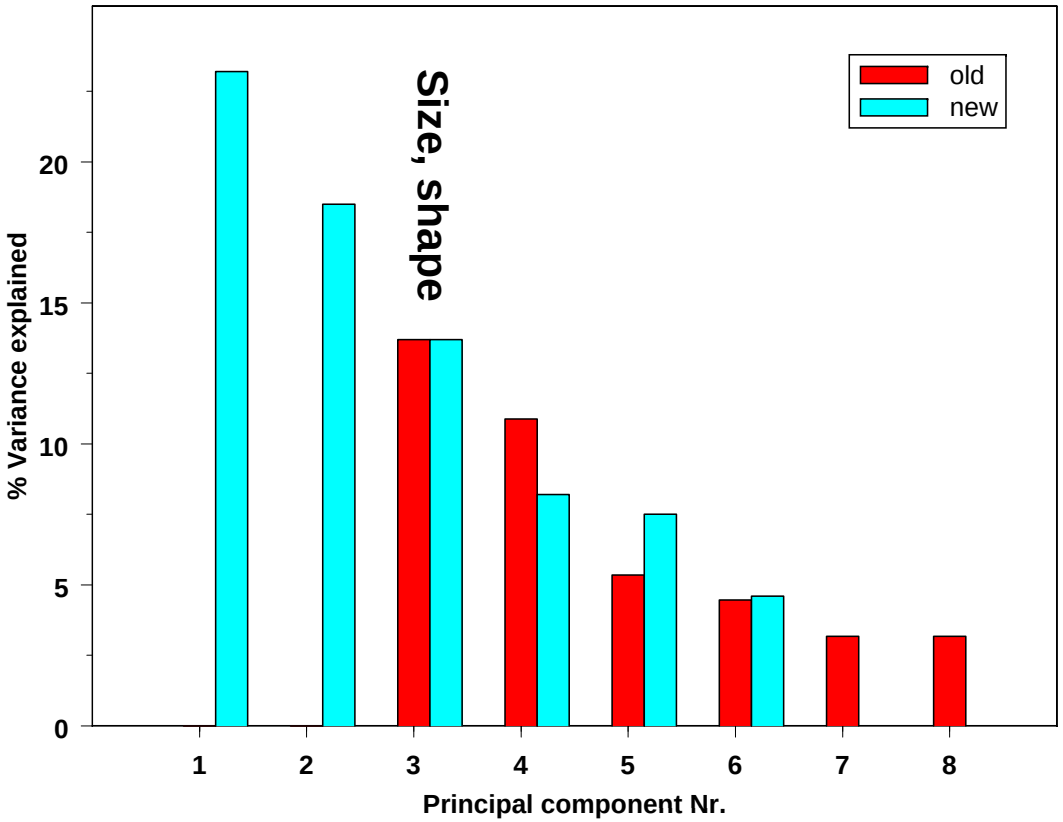


CEPOS

Descriptors

- How can we use the local properties?
  - Statistical descriptors based on the distribution of the properties on the surface:
    - Mean, maximum, minimum, range
    - Variance, skew, kurtosis
    - Balance parameter, ……
  - The descriptors are calculated for each of the local properties to give a comprehensive set of high information content for QSAR

Variance Explained



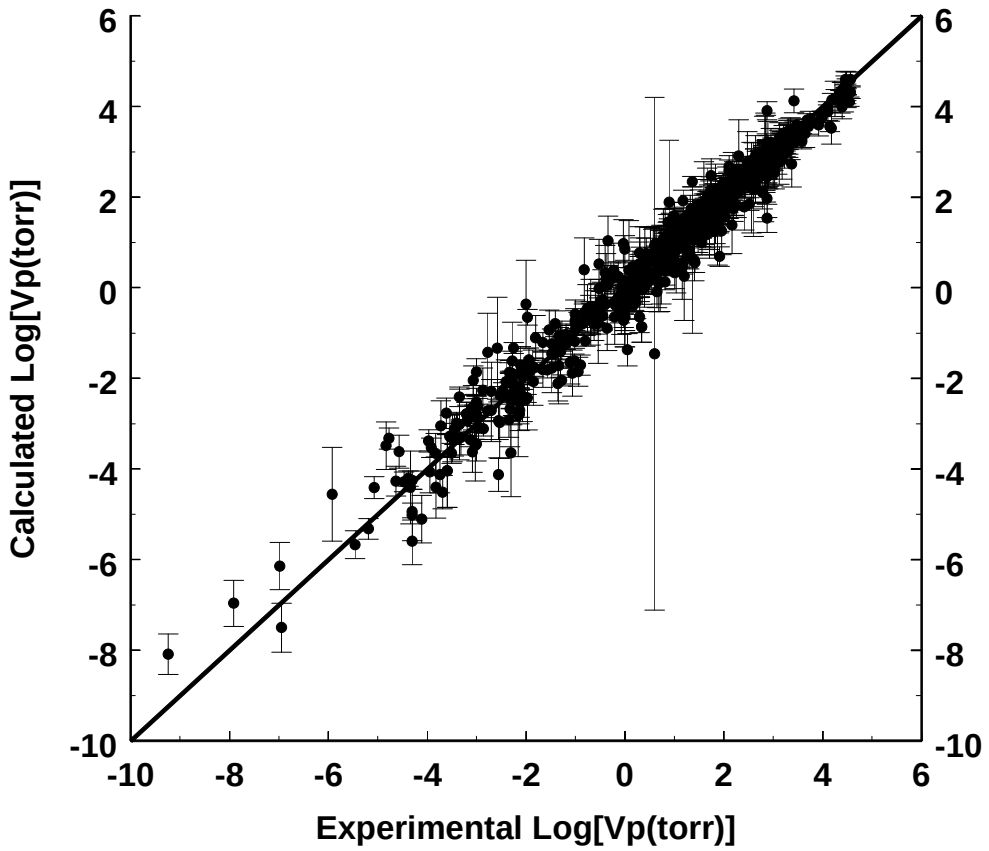
B. Ehresmann, M. J. de Groot, A. Alex and T. Clark, *J. Chem. Inf. Comp. Sci.* **2004** , *44* , 658-668.

- Maybridge database
- Principal components
- "Old" descriptors contain 2D-components
- "New" descriptors purely surface-based or global (MWt, volume etc.)
- Size, shape descriptor is almost identical
- Two most significant proncipal components are mostly local ionization energy and electron affinity

A QSPR Model for Vapor Pressure

- *8,542* measured vapor pressures at temperatures between 76 and 800 K for *2,349* compounds extracted from the Beilstein database
- Therefore include the temperature as a descriptor
- *A Temperature-dependent Quantum Mechanical/Neural Net Model for Vapor Pressure*, A. J. Chalk, B. Beck and T. Clark, *J. Chem. Inf. Comput. Sci.* 2001, *41*, 1053.

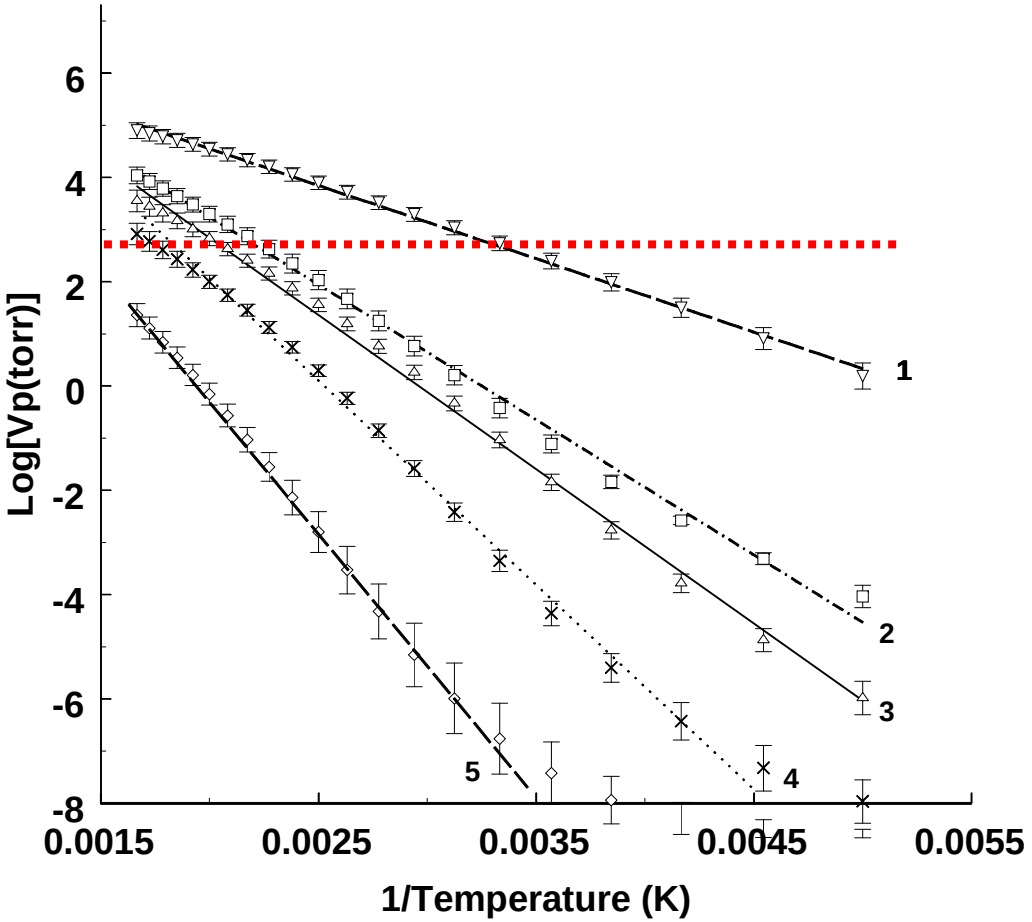
Validation Set



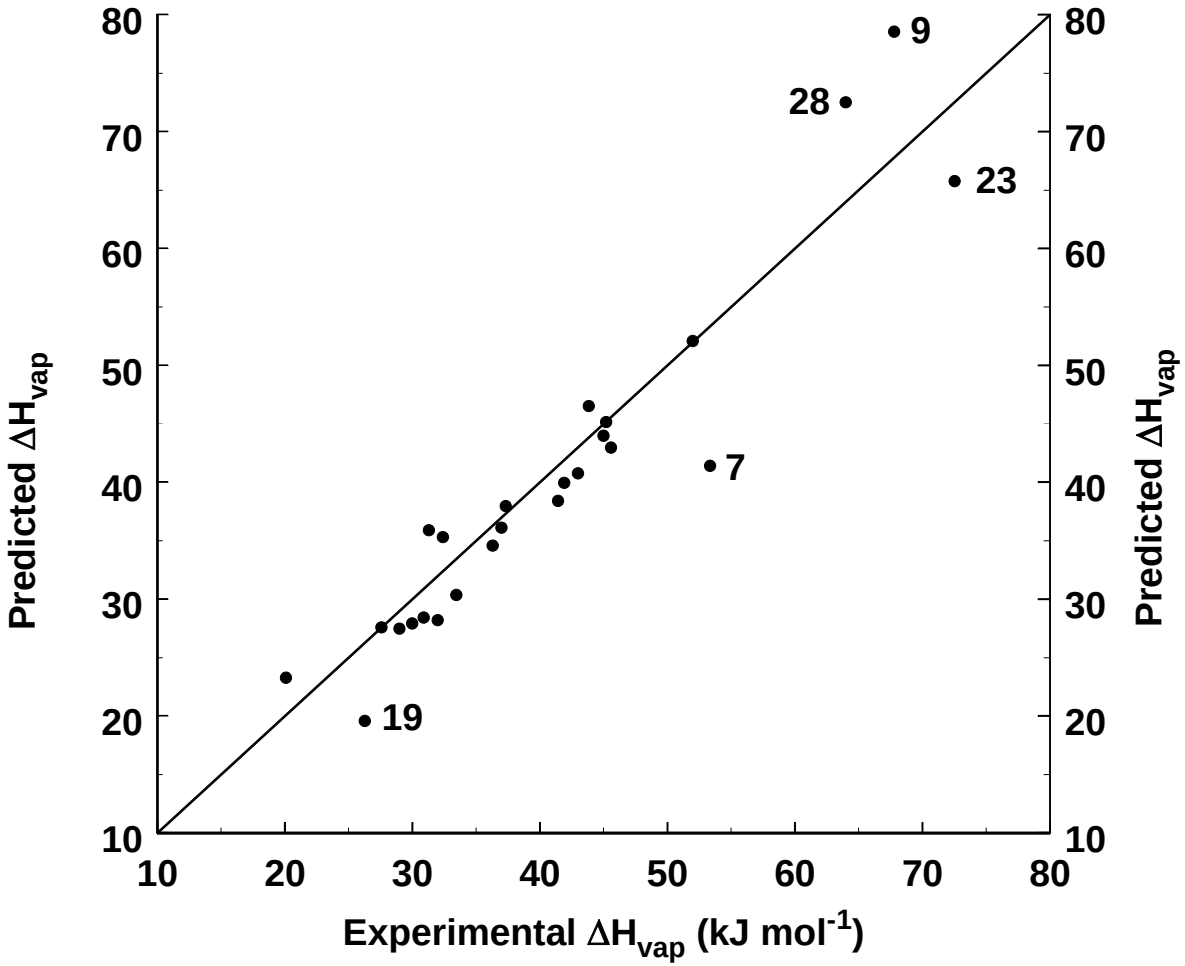
N(train) = 7,681
N(validate) = 861

	Mean	Cross-validation	Validation set
R^2	0.976	0.952	0.976
σ	0.322	0.459	0.326
MUE	0.211	0.291	0.213
LE	2.02	4.15	2.06

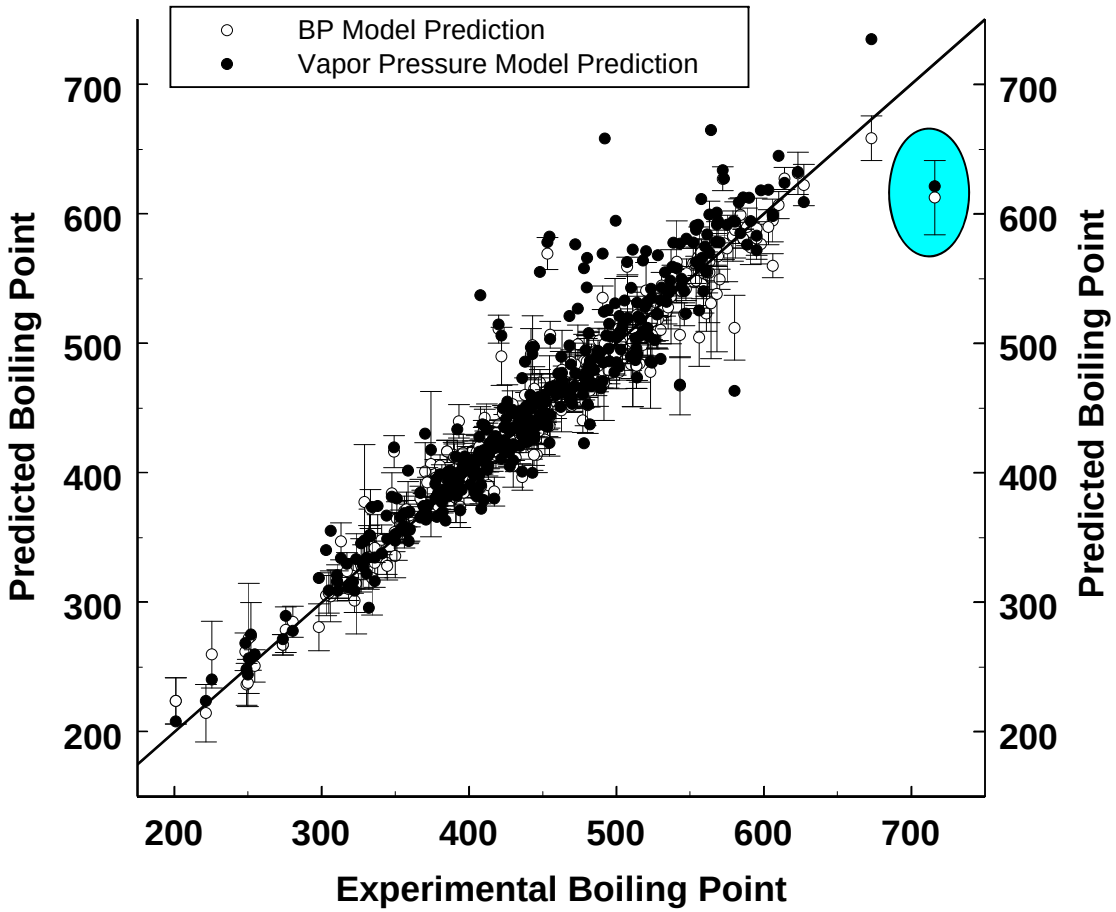
Temperature-Dependence



Heats of Vaporization



Boiling Points



	Vapor pressure model	Boiling point model[a]
R^2	0.889	0.928
σ	28.6	21.9
MUE	18.7	13.5
LE	166.1	140.0

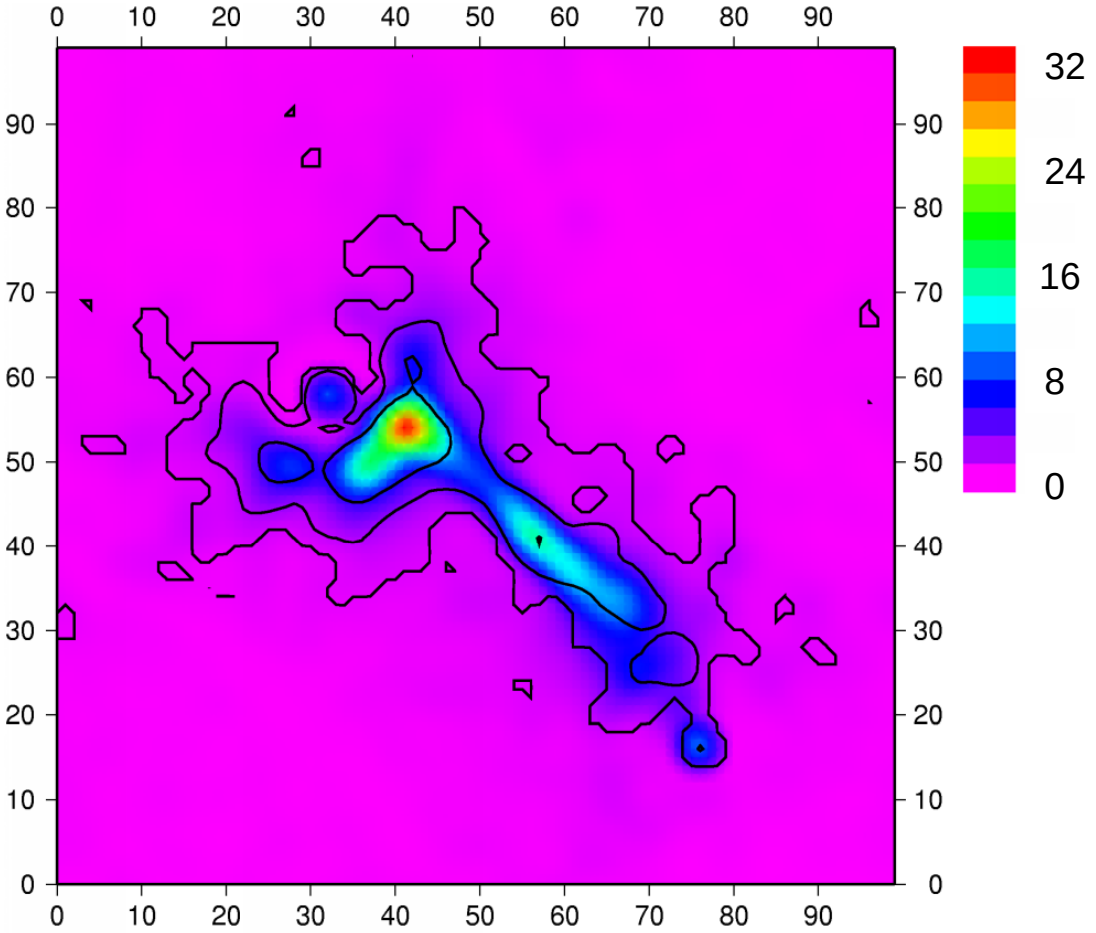
[a] *A Quantum Mechanical/Neural Net Model for Boiling Points with Error Estimation,* A. J. Chalk, B. Beck, and T. Clark, *J. Chem. Inf. Comput. Sci.* 2001, *41*, 457.

Drug-Map

- Maybridge used as the non-drug dataset
- 2,105 compounds selected from the World Drug Index as real drugs used as the drug dataset
- Select descriptors for a Kohonen map by using recursive partitioning to separate the two datasets
- *Descriptors, Physical Properties and Drug-Likeness* , M. Brüstle, B. Beck, T. Schindler, W. King, T. Mitchell and T. Clark, *J. Med. Chem.* **2002** , *45* , 3345.

World Drug Index

(42,131 compounds with 150 atoms or less)



CEPOS

Surface-integral models

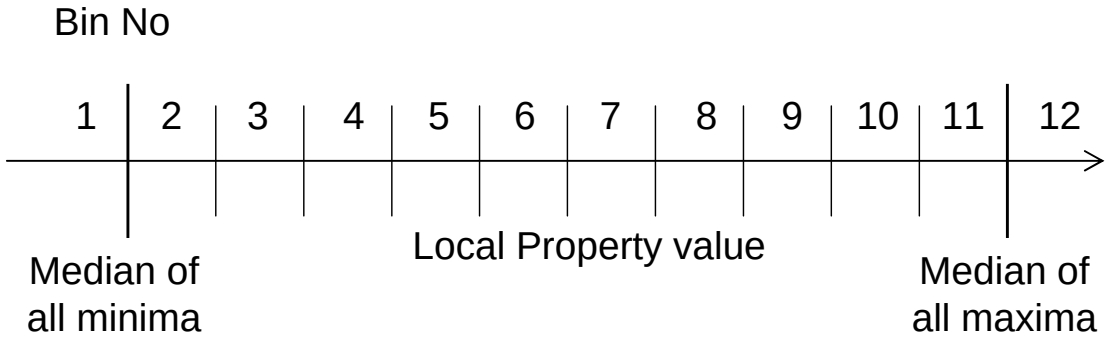
$$P = \sum_{i=1}^{ntri} f\left(V^i(\mathbf{r}_i), IE_L^i(\mathbf{r}_i), EA_L^i(\mathbf{r}_i), \alpha_L^i(\mathbf{r}_i), \eta_L^i(\mathbf{r}_i)\right) \cdot A^i$$

- P = target property
- A_i = area of triangle i
- $ntri$ = number of triangles
- $\mathbf{r}_i$ = center of triangle i

Surface-Integral QSPR Models: Local Energy Properties , B. Ehresmann, M. De Groot and T. Clark, *J. Chem. Inf. Model.* , **2005** , *45* , 1053-1060.

Binned Surface-Integral Models

Standardized binning scheme:



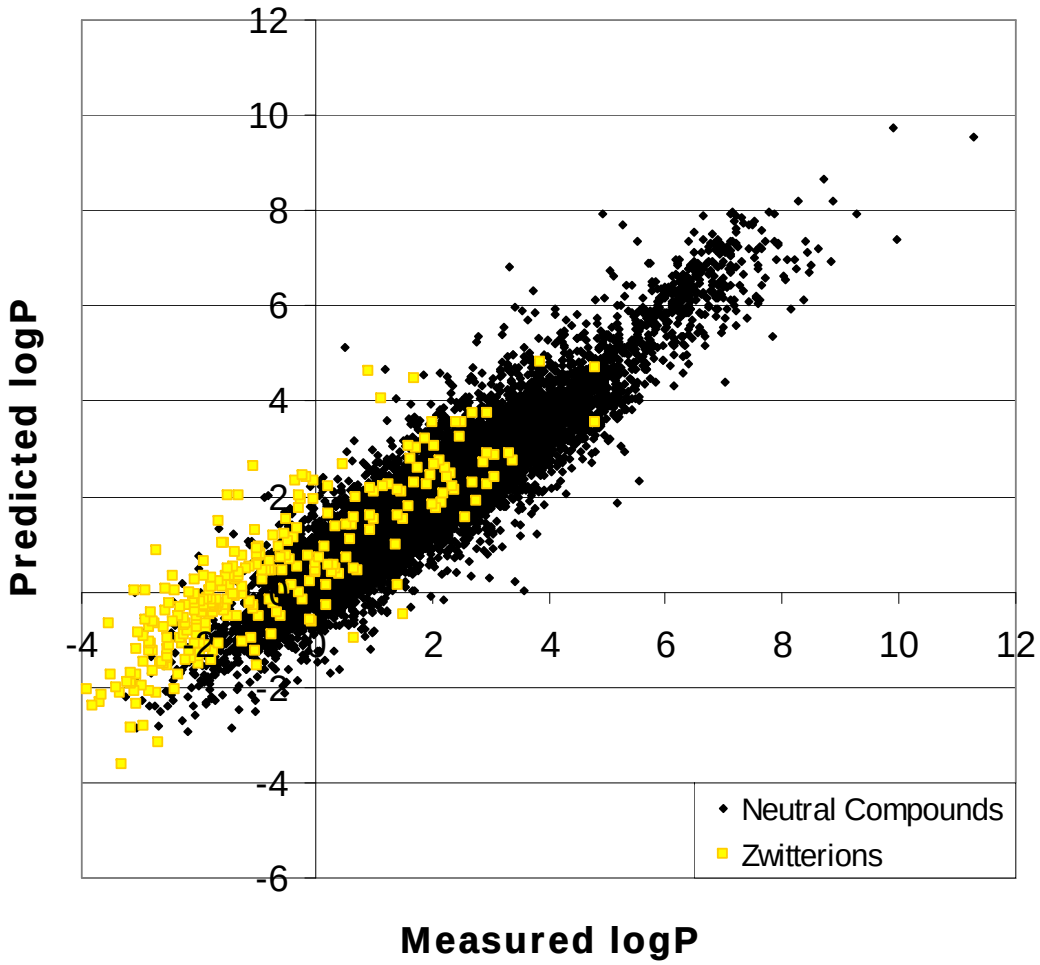
Areas as descriptors:

$$D^N = \sum_{i=1}^{Ntri} \delta_i^N A_i$$

D^N = descriptor for bin N
δ_i^N = 1 if the triangle is within the bin, 0 otherwise.

A Surface-Integral Model for logP_{OW}, C. Kramer, B. Beck and T. Clark, *J. Chem. Inf. Model.*, **2010**, *50*, ASAP (DOI: 10.1021/ci900431f).

Results (Test Set)



CEPOS

logP Model

Model	Public Validation Set	
	RMSE	
ACDlabs logP	0.45	
ClogP (BioByte)	0.52	
SIM-logP	0.64	
SlogP (MOE)	0.68	
logP_o/w (MOE)	0.77	
AlogP (TSAR)	0.86	

A Surface-Integral Model for logP_{OW}, C. Kramer, B. Beck and T. Clark, *J. Chem. Inf. Model.*, **2010**, *50*, ASAP (DOI: 10.1021/ci900431f).

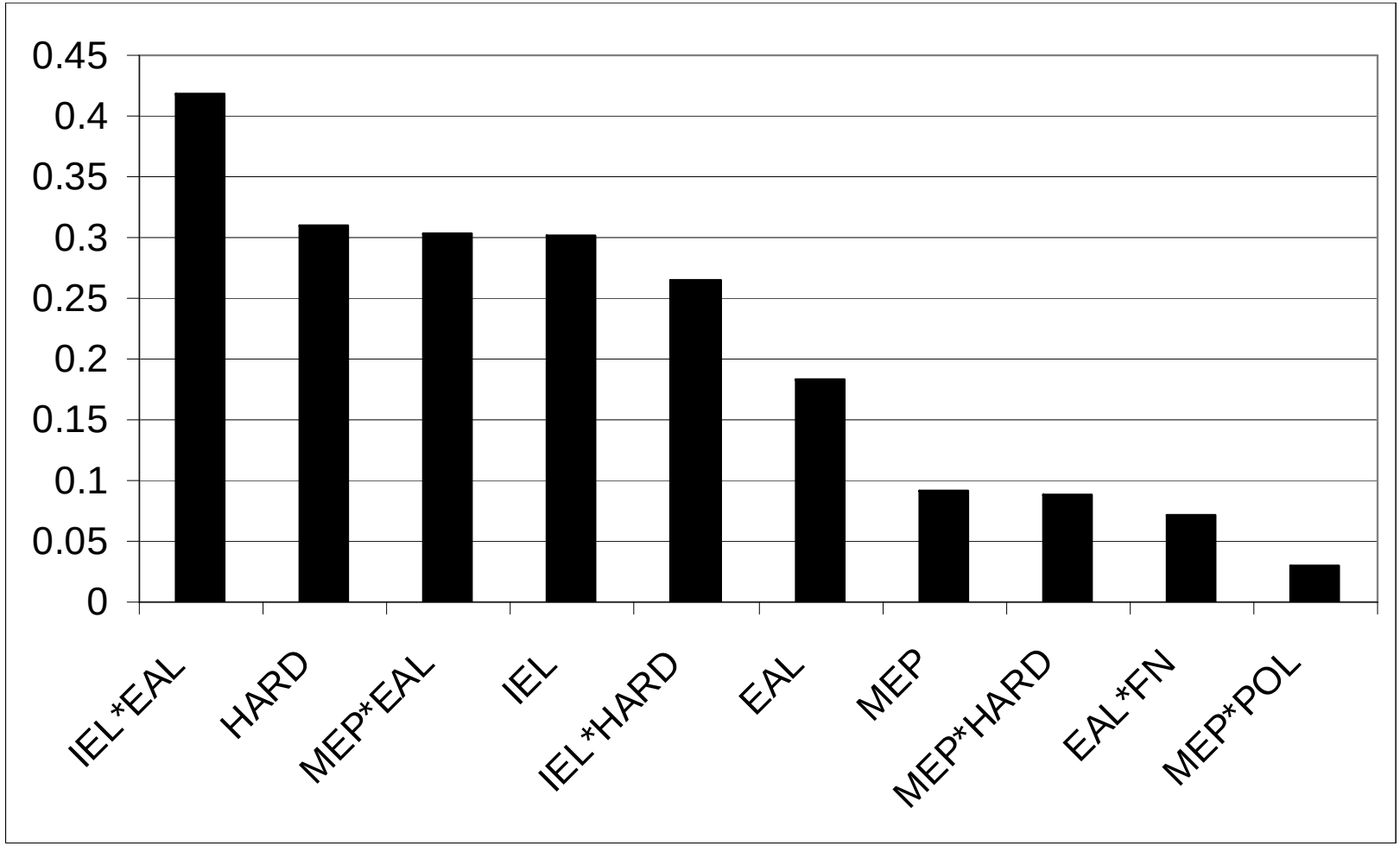
CEPOS

logP Model

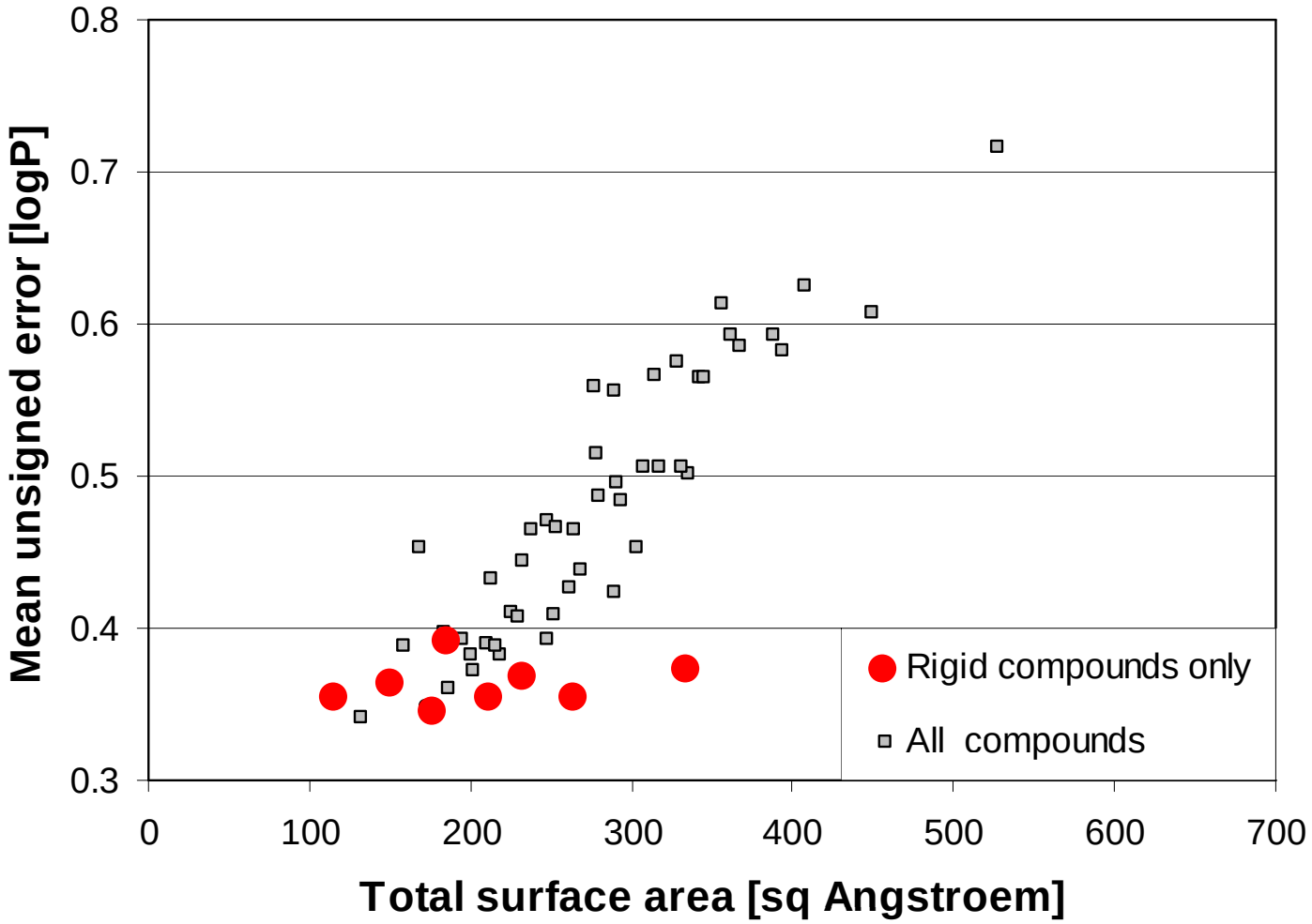
Model	Public Validation Set	In-house Validation Set
	RMSE	RMSE
SIM-logP	0.64	1.10
ClogP (BioByte)	0.52	1.14
SlogP (MOE)	0.68	1.19
AlogP (TSAR)	0.86	1.24
logP_o/w (MOE)	0.77	1.28
ACDlabs logP	0.45	1.36

A Surface-Integral Model for logP_{OW}, C. Kramer, B. Beck and T. Clark, *J. Chem. Inf. Model.*, **2010**, *50*, ASAP (DOI: 10.1021/ci900431f).

Which Properties are Important?



Effect of Conformation



Multi-Conformation logP

Compound	Experiment	Model	CORINA conformation	Boltzmann weighted	
				Gas-phase	SCRF
Omeprazole	2.23 2.38	All compounds	2.94 (0.56-0.71)	2.97	2.65
		Rigid only	2.92	2.68	2.51 (0.13-0.28) ↓
Risperidone	3.04	All compounds	2.48 (0.46)	3.20	3.14
		Rigid only	2.78	3.17	3.03 (0.01) ↑
Haloperidol	4.30	All compounds	4.52 (0.22)	4.23	4.19
		Rigid only	4.17	4.20	4.18 (0.12)→

CEPOS

Reactivity

- The local ionization energy and electron affinity are not only related to donor-acceptor interactions, but also to nucleophilic and electrophilic properties, respectively.

- They are therefore ideally suited for predicting reactivity.

- One example is the lability to oxidation by cytochrome P450 enzymes

CypScore: Quantitative Prediction of Reactivity toward Cytochromes P450 Based on Semiempirical Molecular Orbital Theory, M. Hennemann, A. Friedl, M. Lobell, J. Keldenich, A. Hillisch, T. Clark and A. H. Göller, *ChemMedChem*, **2009**, *4*, 657-669.

A Model for Metabolic Lability

- … predicts the likely sites of Cytochrome P450 metabolism in small molecules

- … is a model for a hypothetical CYP P450 super-enzyme
  - "Simulates" oxidation reactions by P450 enzymes
  - Currently no active-site accessibility restrictions

- … is based AM1 and ParaSurf
  - Describes oxidation lability of centers of chemical reactivity
  - Models available for most frequent oxidation reactions
  - All Models weighted to a consensus range
  - Restricted to Phase I reactions

Atomic Surface Descriptors (ParaSurf)

1. Triangulate surface

2. Calculate atomic contributions to the electron density at the corners

3. If two corners “belong” to the same atom, allocate the triangle to that atom

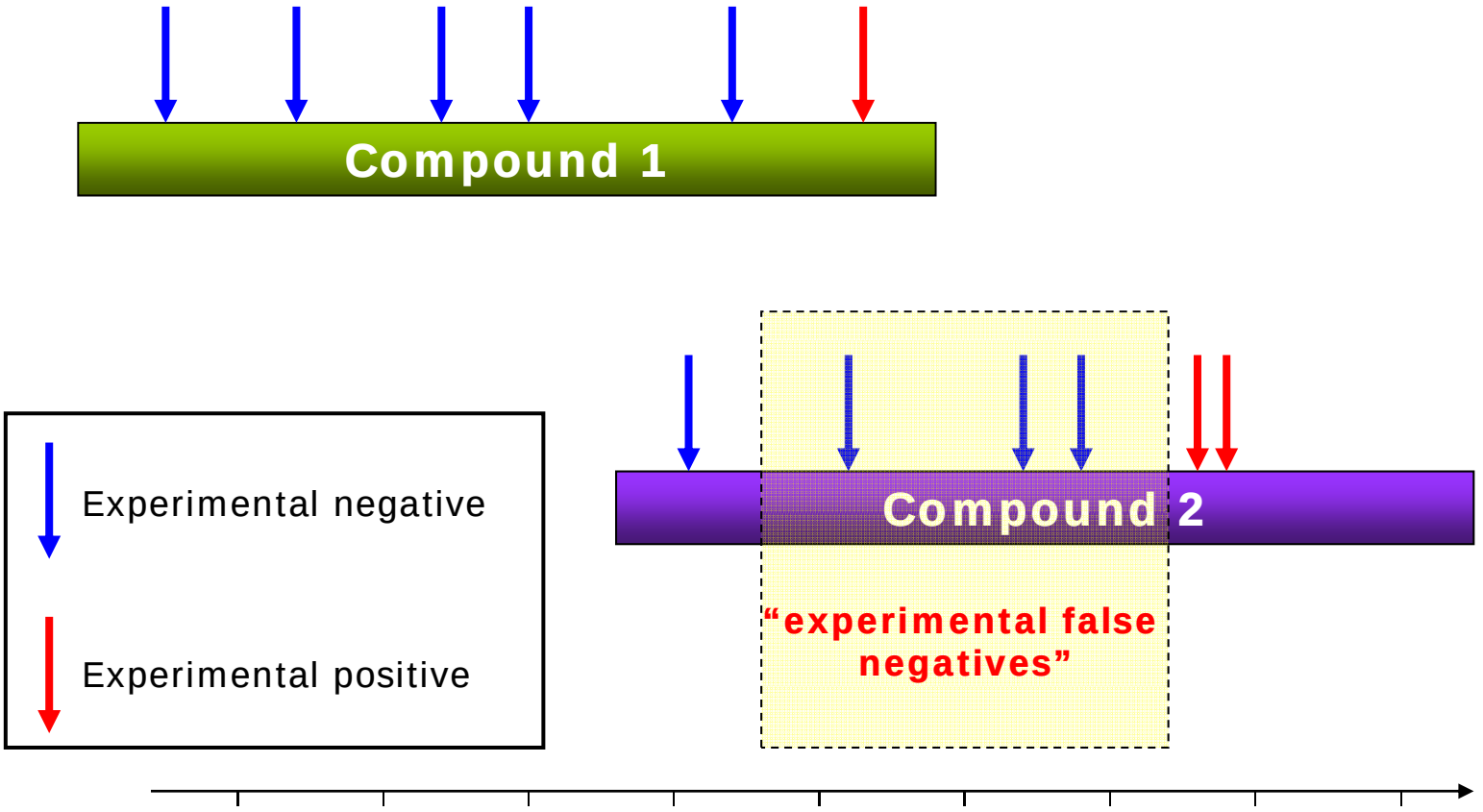
4. Atoms are described by the statistical characteristics (max, min, mean, variance etc.) of the distribution of the local properties on “their” surface.

5. Provides an excellent description of local reactivity

Models Based on ParaSurf Atomic Surface Properties

- **Dataset:**
  - MajorMetaboliteDB from Bayer with 950 compounds and 1750 reactions
- **Models:**
  - Individual models for the different types of reaction
  - All scaled to a common “metabolic strength” scale
  - All descriptor-based regression models
  - None with more than four descriptors
- **Regression philosophy:**
  - Data is binary (0,1), but we need a quantitative scale
  - Use “unsuitable” MLR and assume that the very weak positions will be detected by the descriptors
  - Better than a classification model (!) because of the structure of the data

Data


Compound 1
Compound 2
Experimental negative
Experimental positive
“experimental false negatives”
Susceptibility to metabolization

Published Models

- **Reaction centers:**
  - Csp^3
  - Csp^2 (Aryl)
  - Csp^2 (C=X)
  - N-oxidation
  - NR_2 and NR_3

- **All centered and scaled**
  - Scores range from -10 (very weak) to +10 (stable)
  - A score of zero marks equal probability of positive and negative

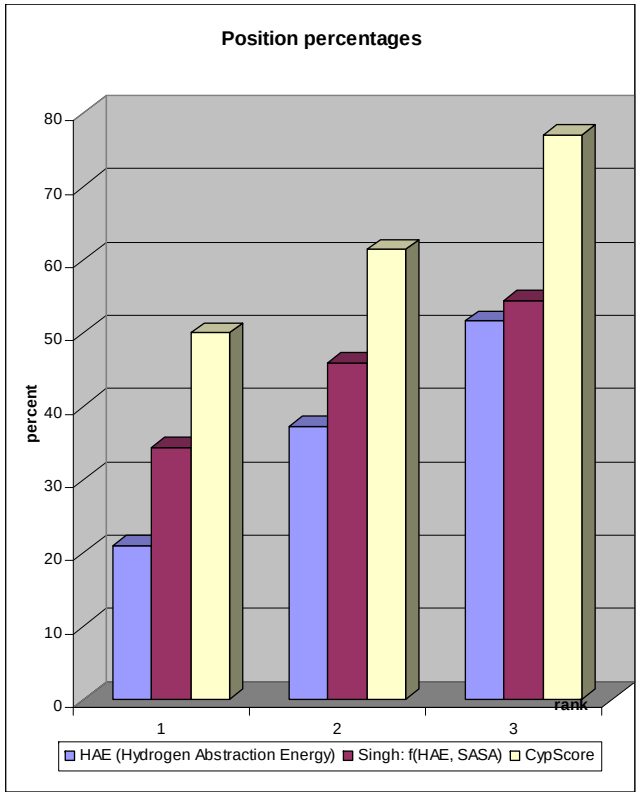
Results : "Public" validation Set

70 compounds, 41 from the Singh data set, plus 29 challenging ones

- 187 weak positions in 1797 heavy atoms (1 to 5 per molecule)
- all types of reactions
  - 124 aliphatic hydroxylations
  - 50 Double bond oxidations
  - 12 N-oxidations

- 73% of metabolic positions found

Bayer HealthCare



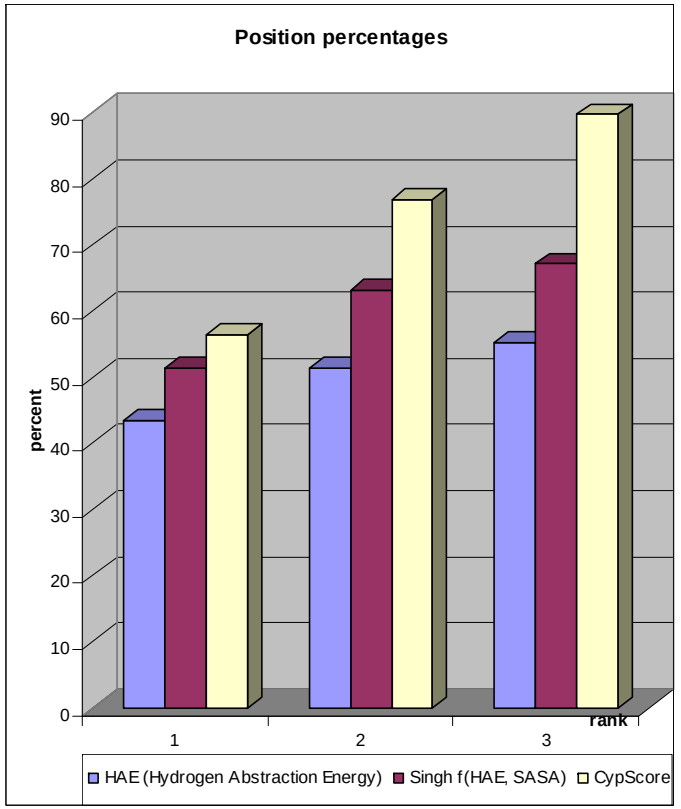
Results: in-house validation set

39 compounds from 12 in-house projects

- for these compounds the positions are clearly identified experimentally
- all types of reactions
- 90 % of metabolic positions found in 3 highest ranked positions

Bayer HealthCare

CypScore driven metabolic optimizations have been performed successfully at Bayer



CEPOS

Spherical-Harmonic Fitting

$$r(\theta,\phi) = \sum_{l=0}^{N} \sum_{m=-l}^{l} a_{lm} y_{lm}(\theta,\phi)$$

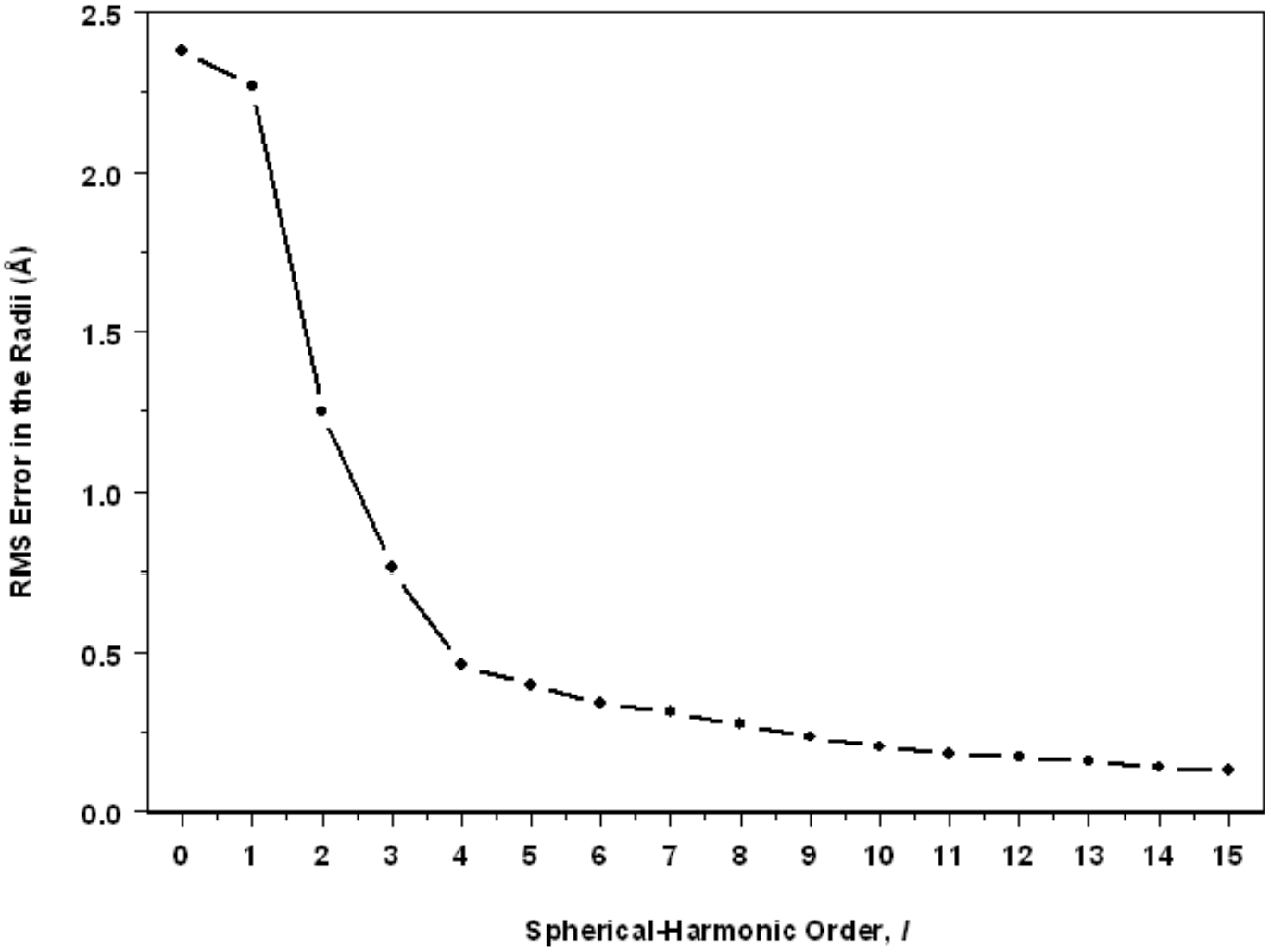
$$y_{lm}(\theta,\phi) = \{N_{lm} P_{lm}(\cos\theta) \cos|m|\phi,\ m \geq 0$$

and

$$y_{lm}(\theta,\phi) = \{N_{lm} P_{lm}(\cos\theta) \sin|m|\phi,\ m < 0$$

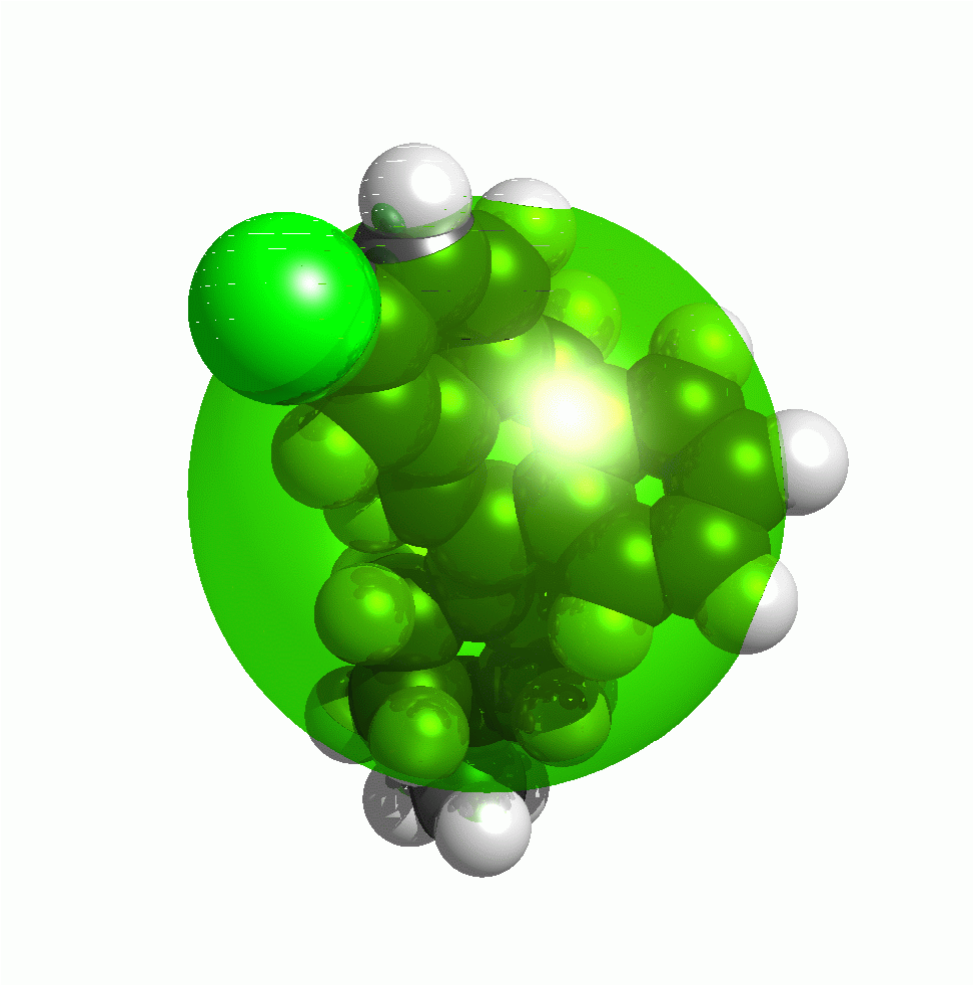
An Analytical, Variable Resolution, Complete Description of Static Molecules and Their Intermolecular Binding Properties, J.-H. Lin and T. Clark, *J. Chem. Inf. Model.*, **2005**, *45*, 1010−1016.

Radial-Distance Fit



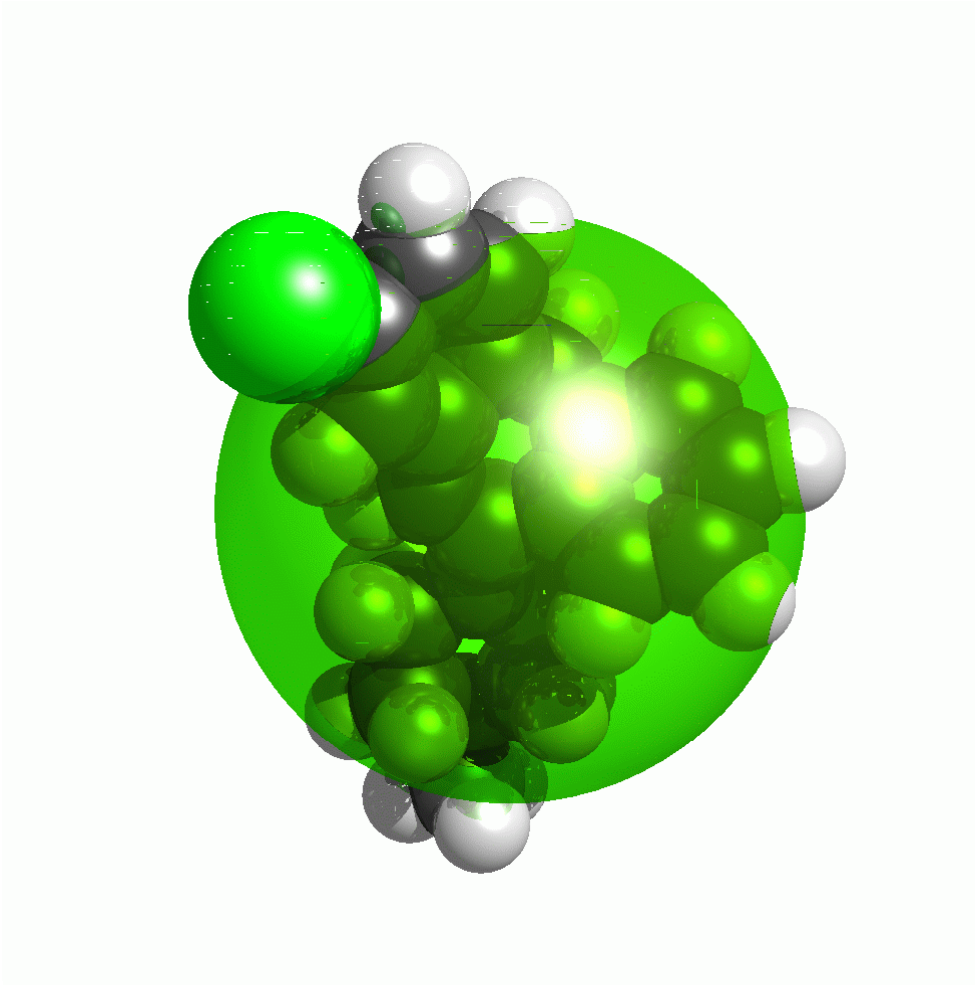
Shape Description

L=1



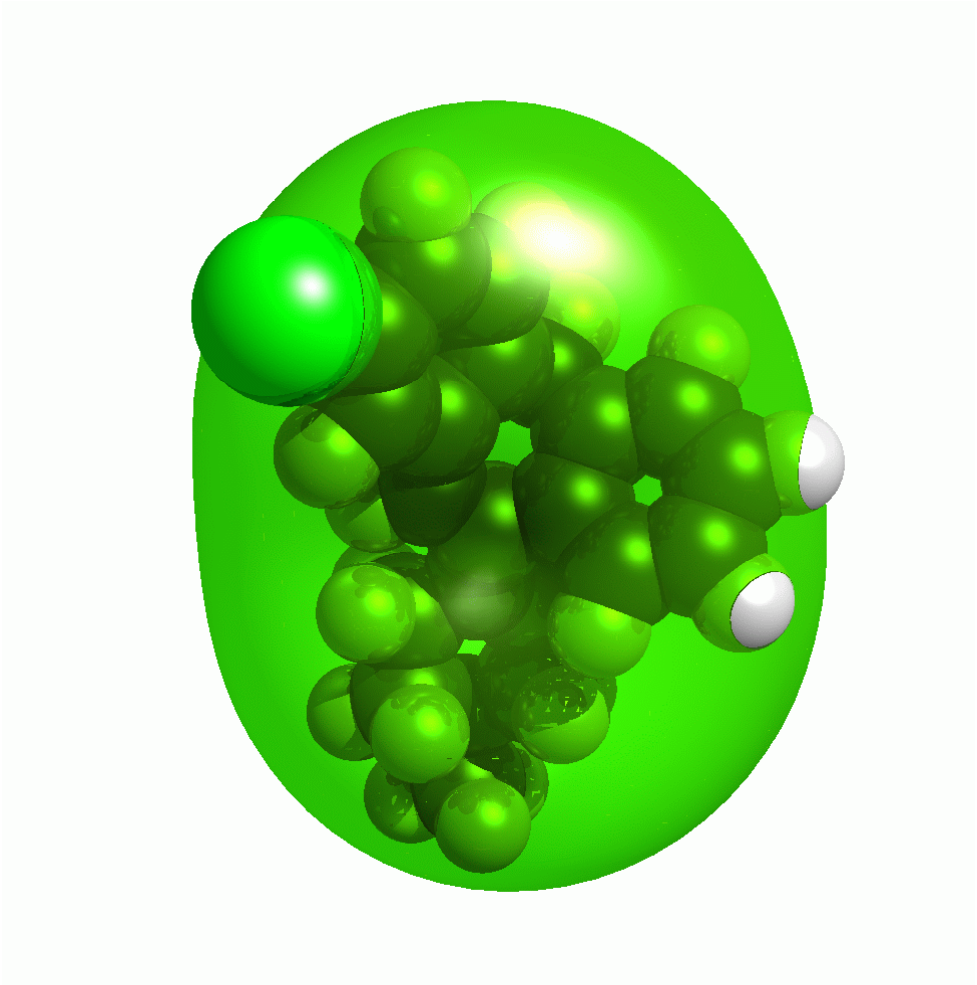
Shape Description

L=2



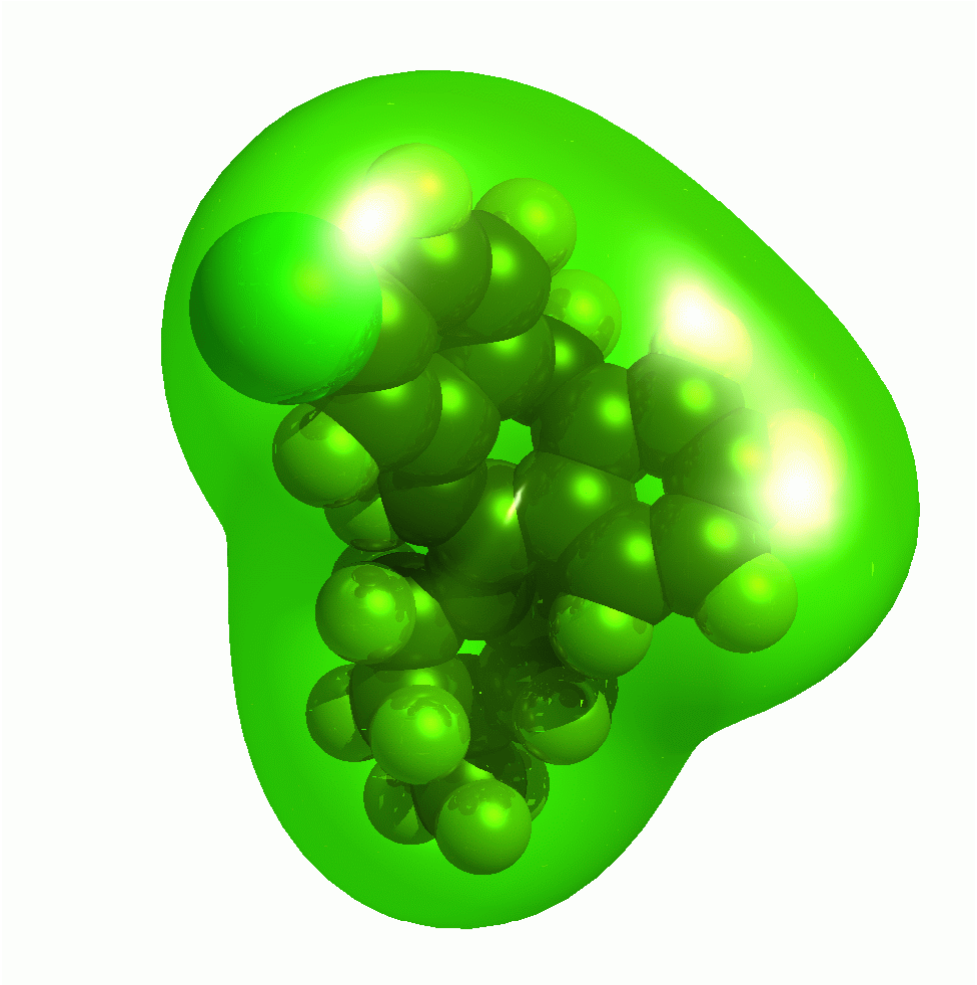
Shape Description

L=3



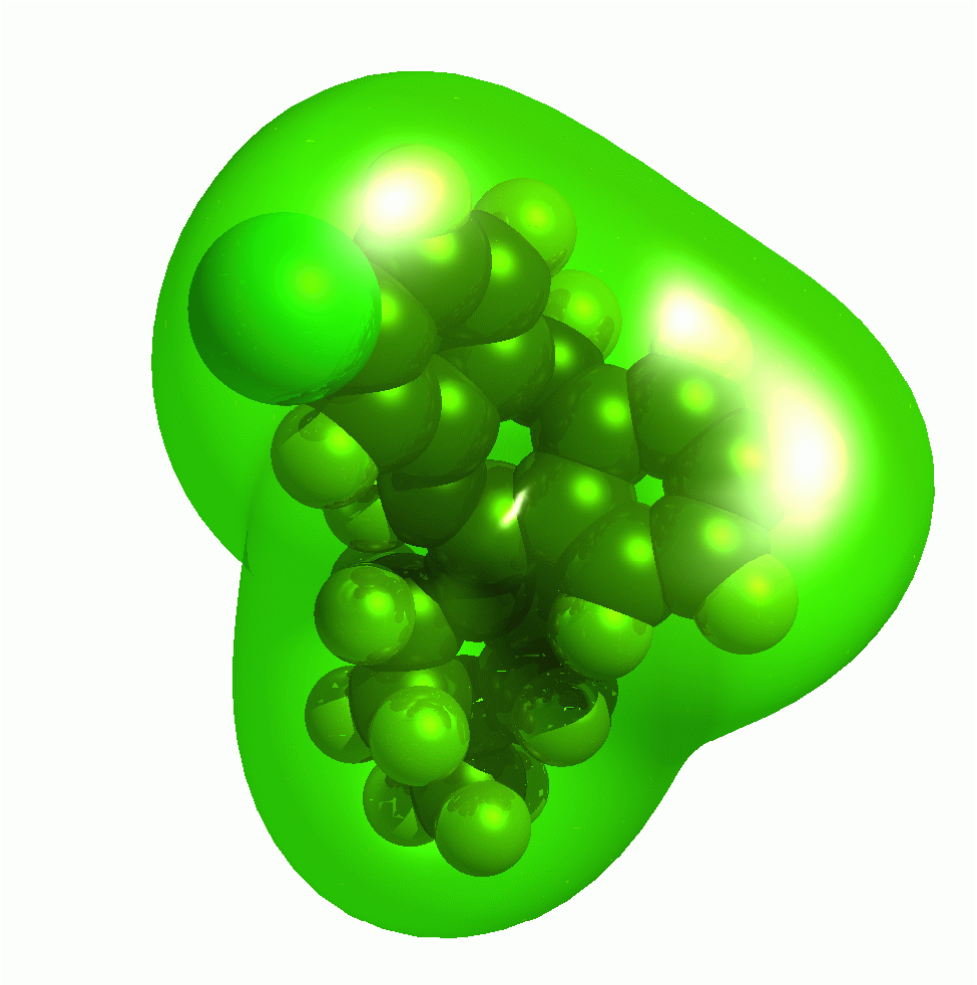
Shape Description

L=4



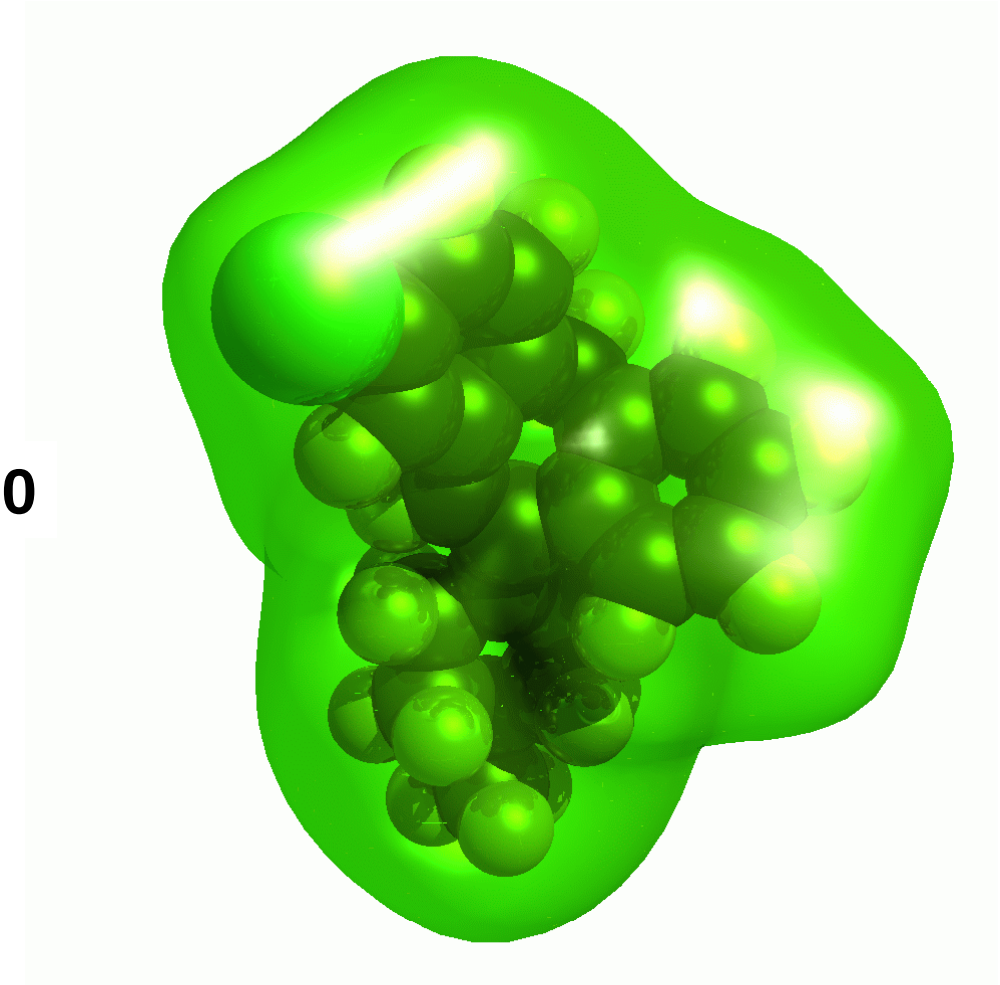
Shape Description

L=5



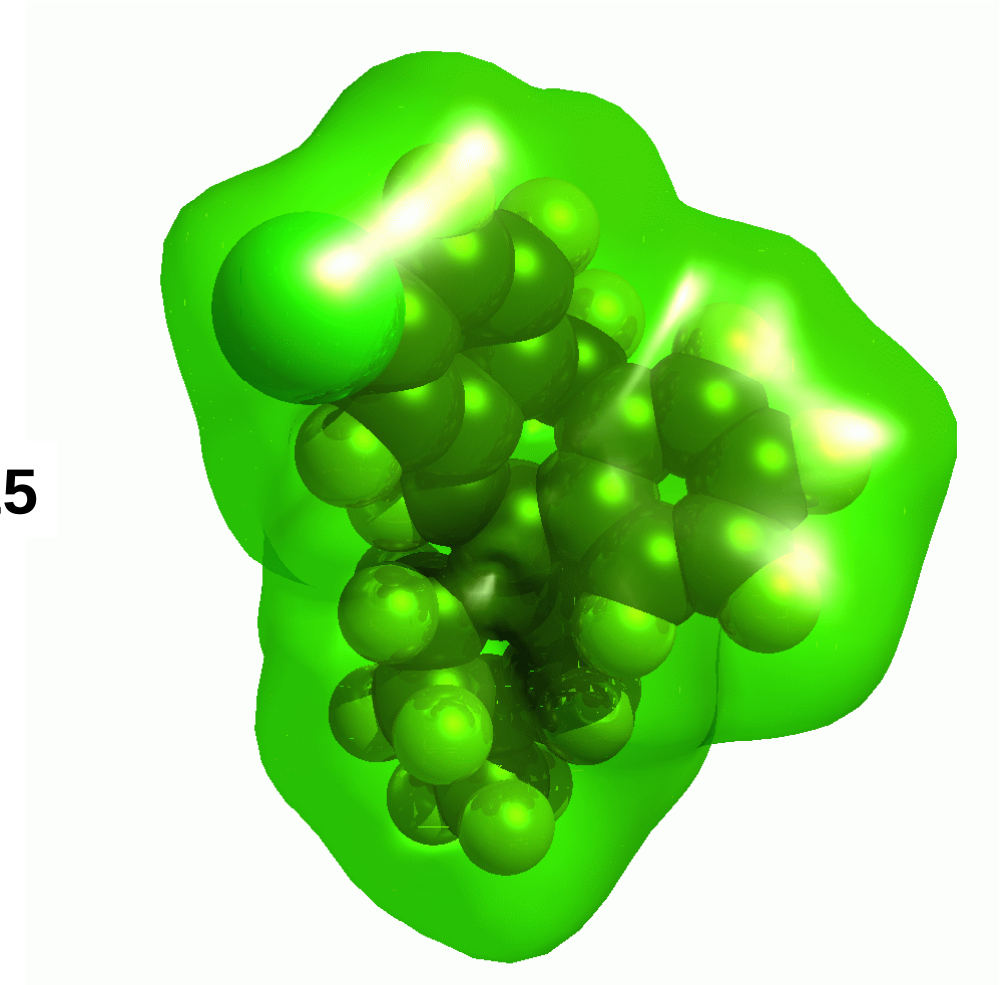
Shape Description

L=10



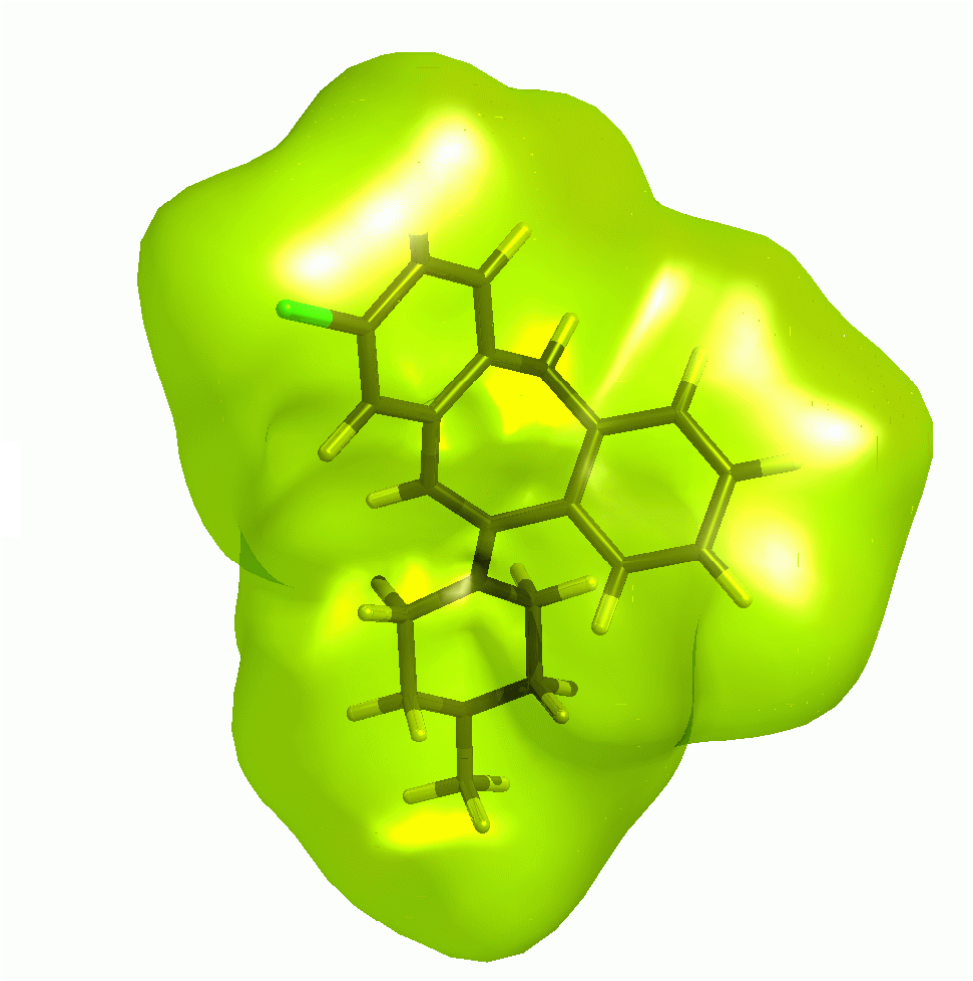
Shape Description

L=15



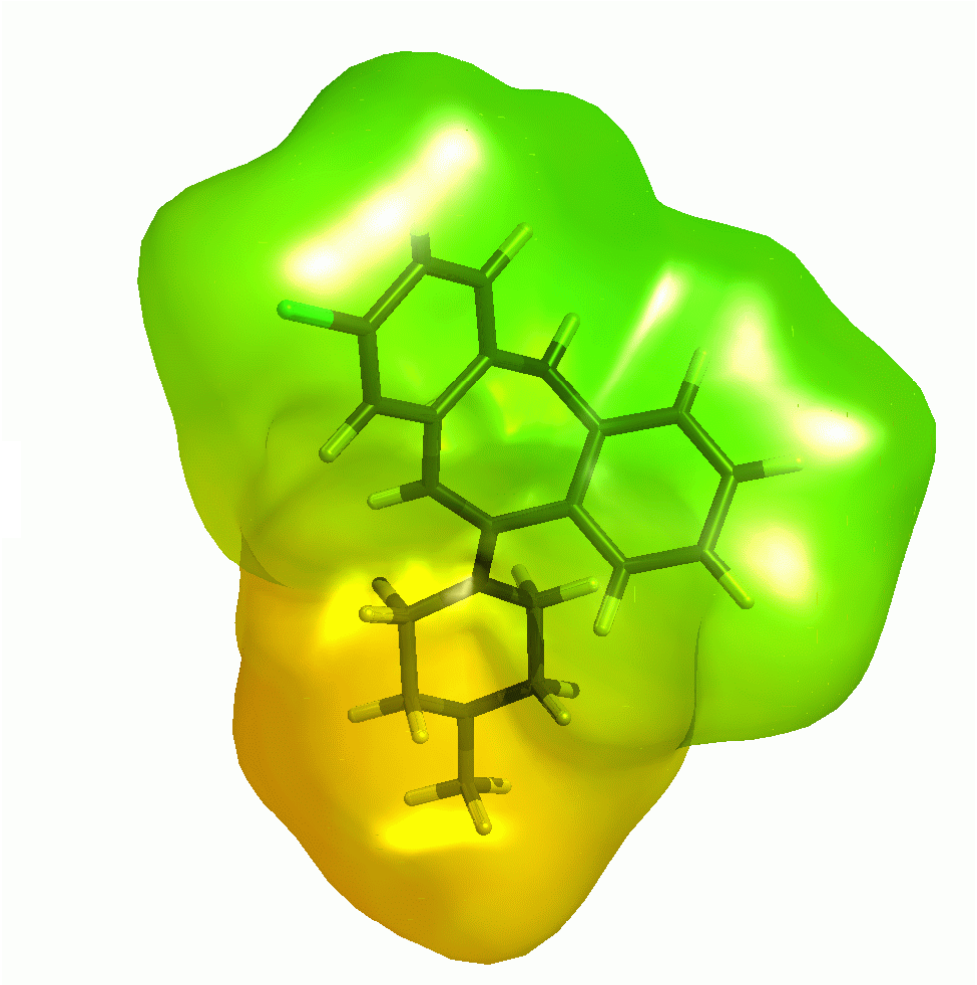
MEP

L=0



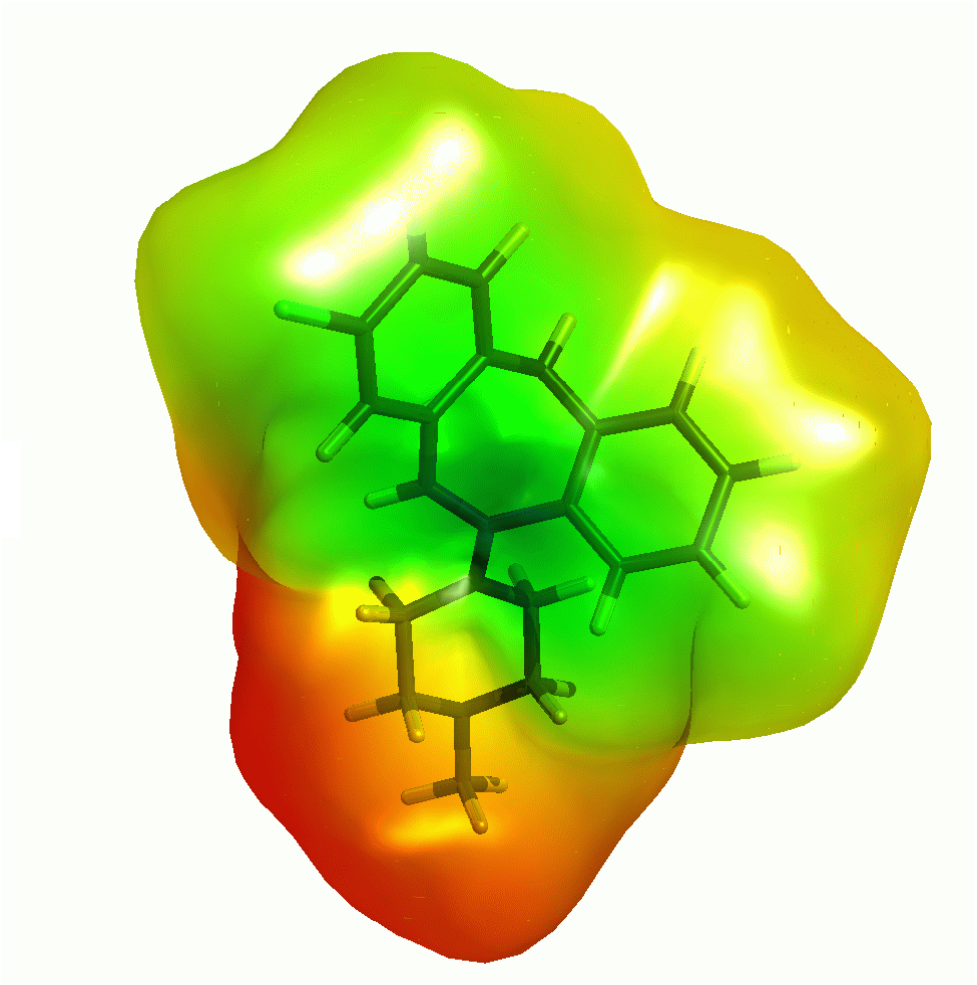
MEP

L=1



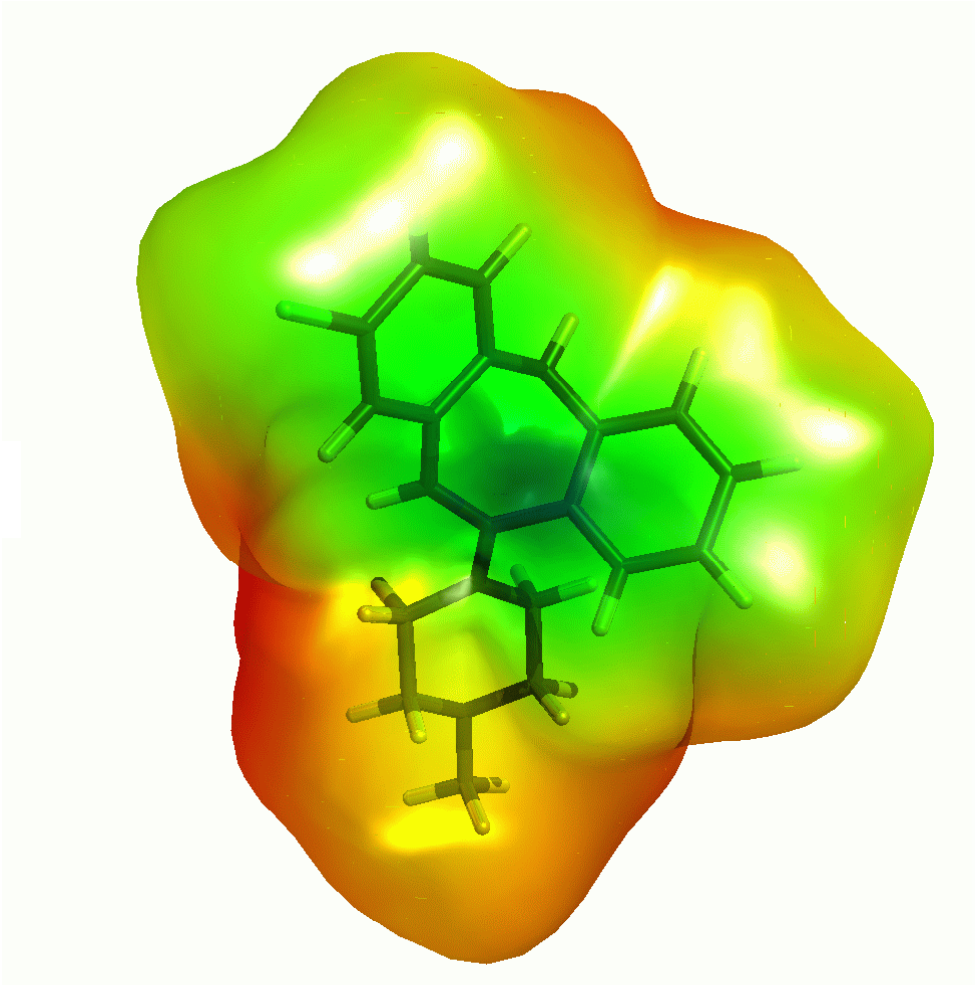
MEP

L=2



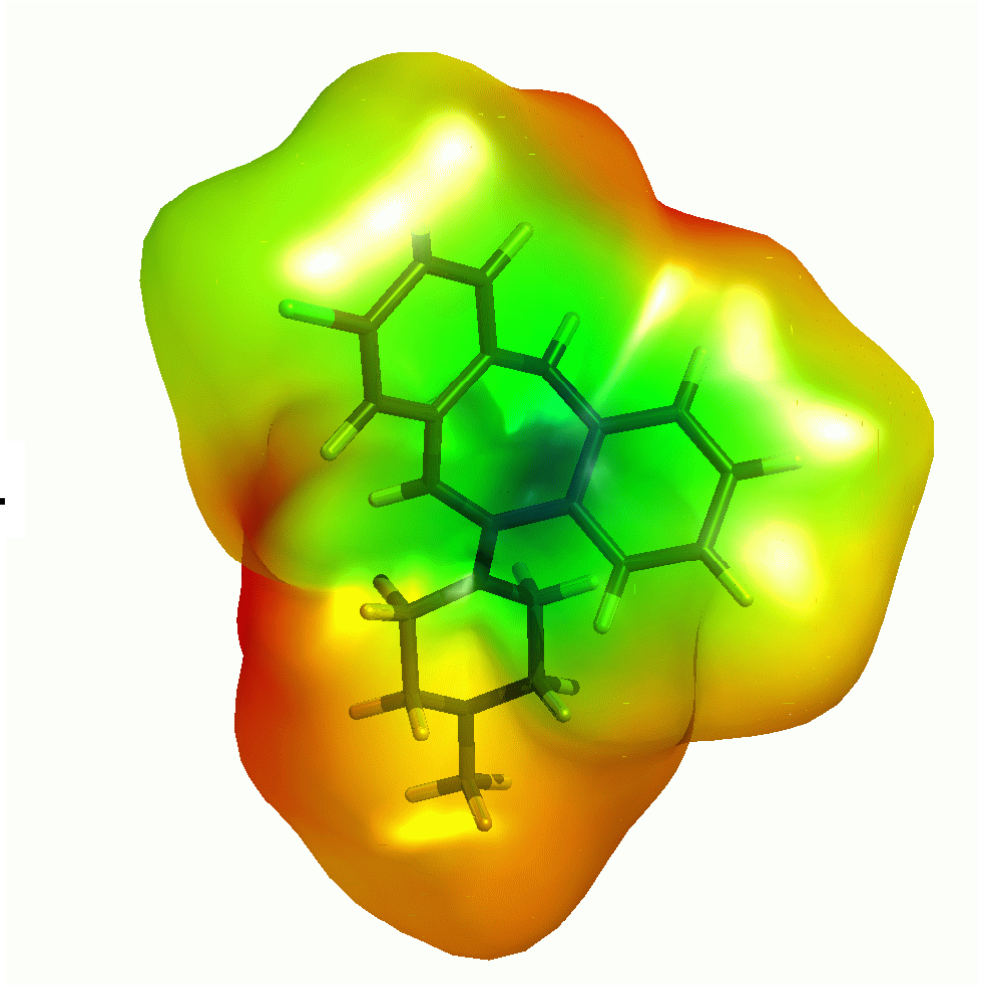
MEP

L=3



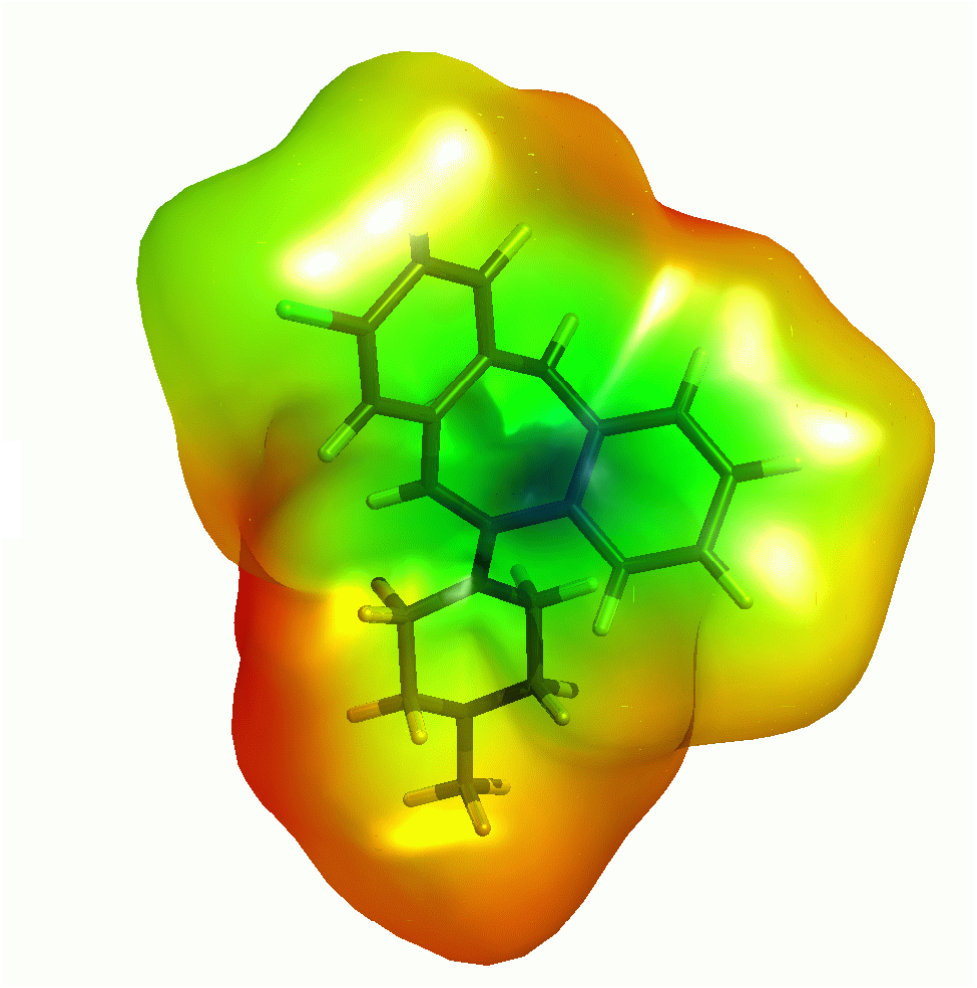
MEP

L=4



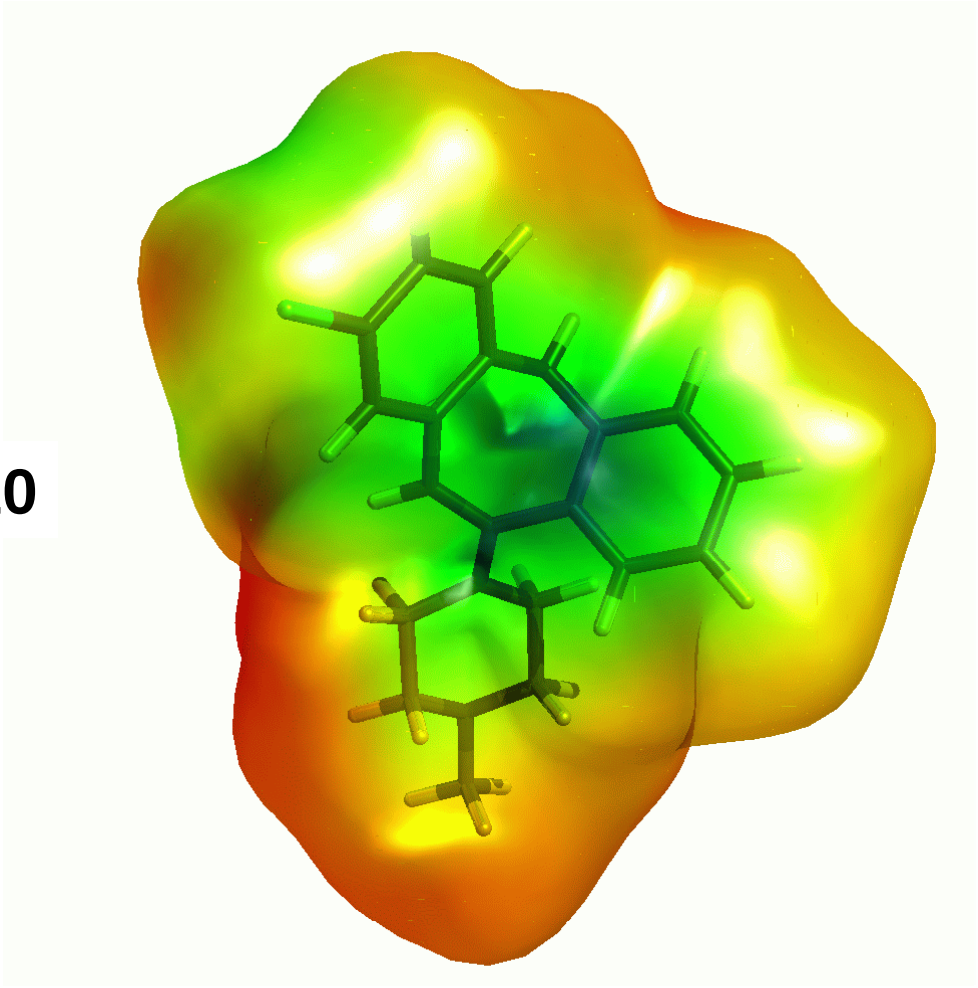
MEP

L=5



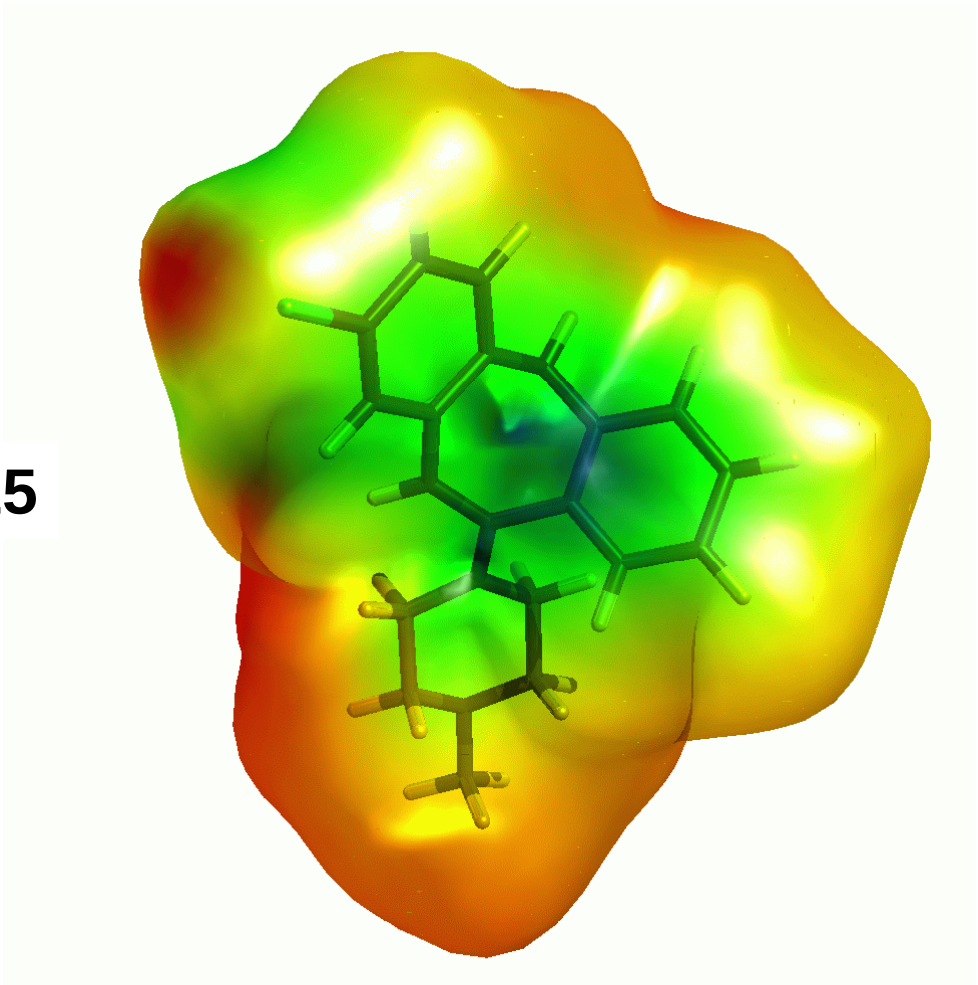
MEP

L=10



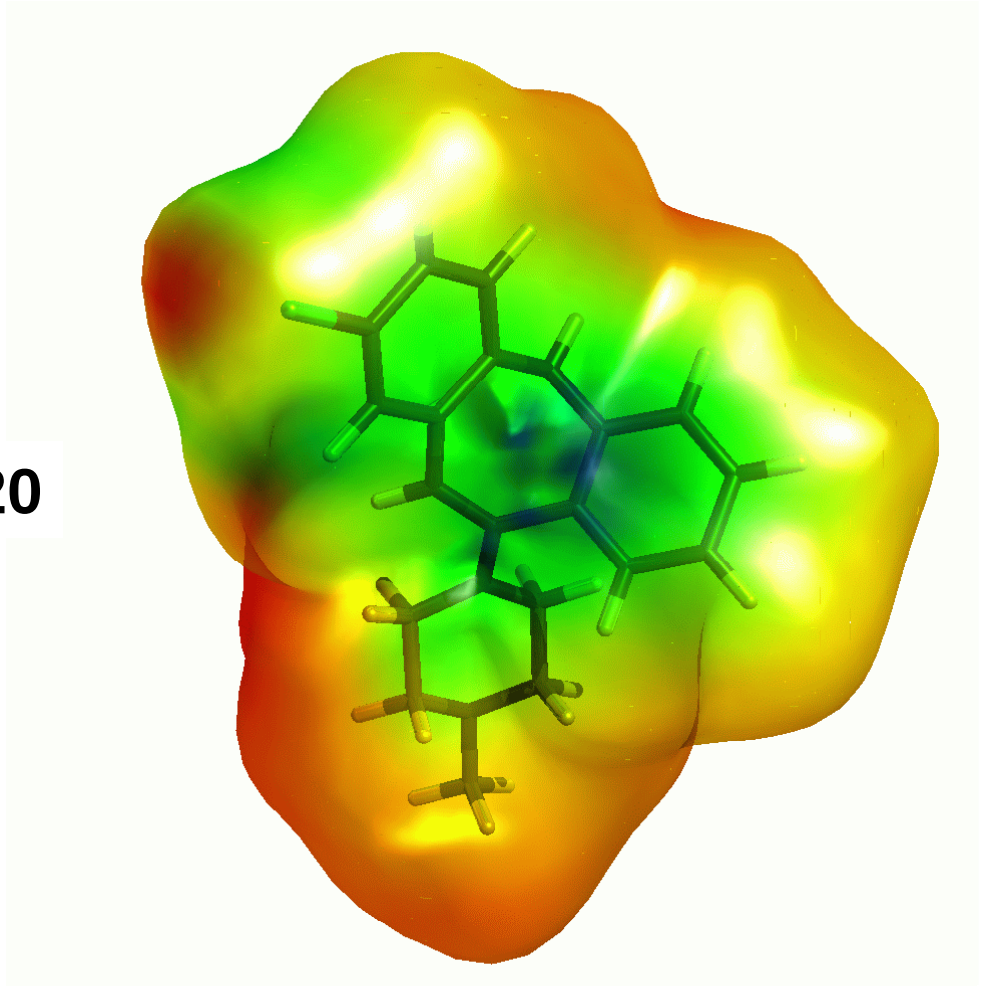
MEP

L=15



MEP

L=20



New Developments

- "ParaLeap"
  - High-throughput visualization of molecular similarity using Lambert azimuthal equal-area projections
  - Joint development with Boehringer-Ingelheim, Biberach (Bernd Beck, Thomas Fox)
  - Uses ParaSurf™ and ParaFit™ technology

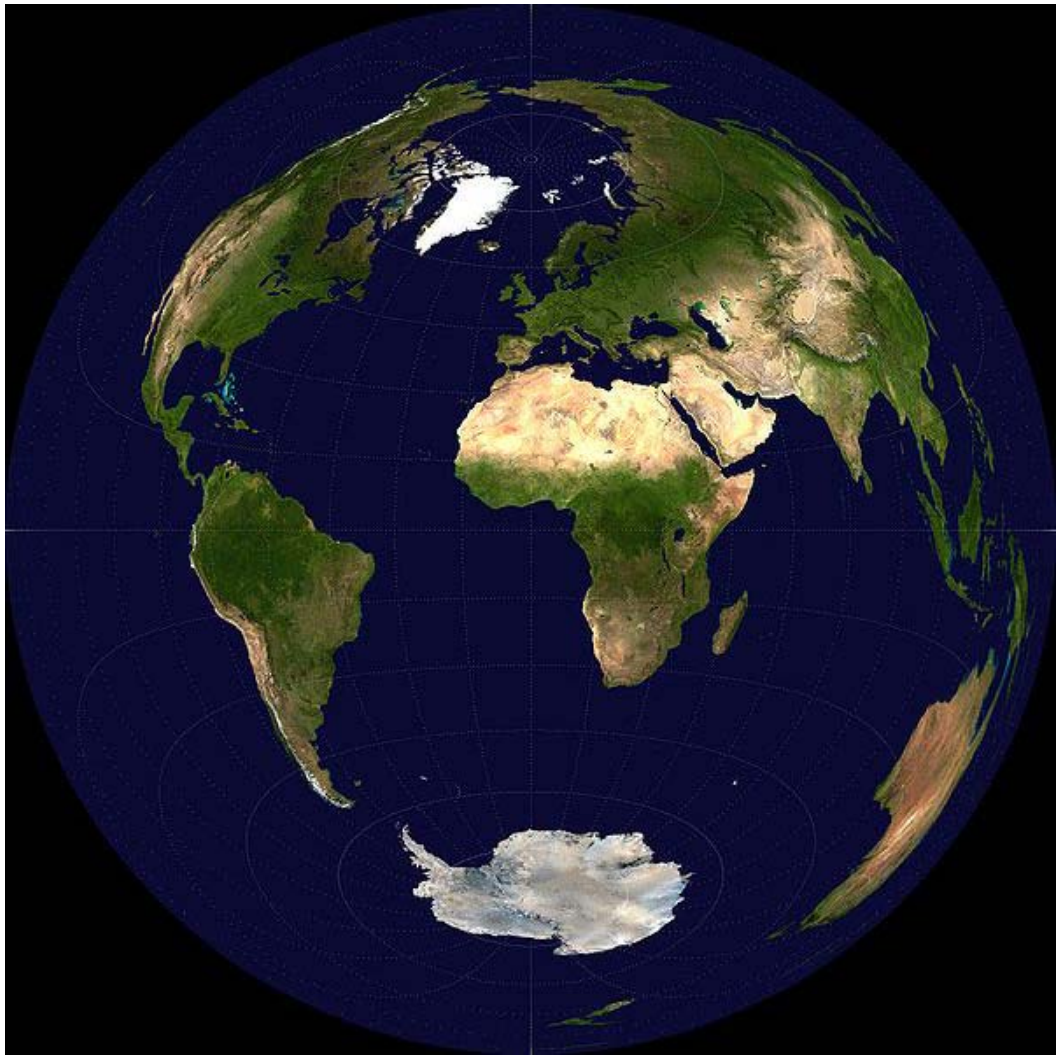
Lambert Azimuthal Equal Area Projection

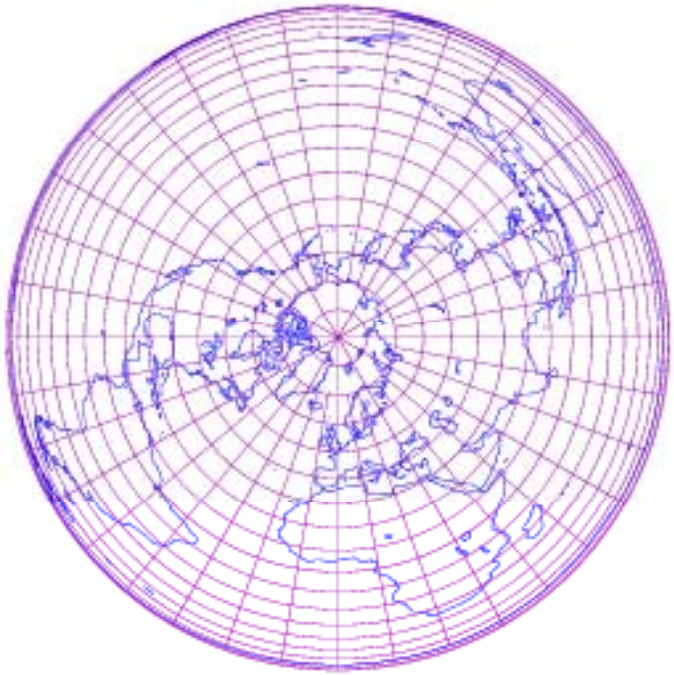
- **Johann Heinrich Lambert**
- 1728 – 1777
- Swiss mathematician, physicist and astronomer

- In 1764 Lambert was invited by Euler to join the Prussian Academy of Sciences in Berlin under the sponsorship of Frederick II of Prussia. He worked there until his death in 1777.

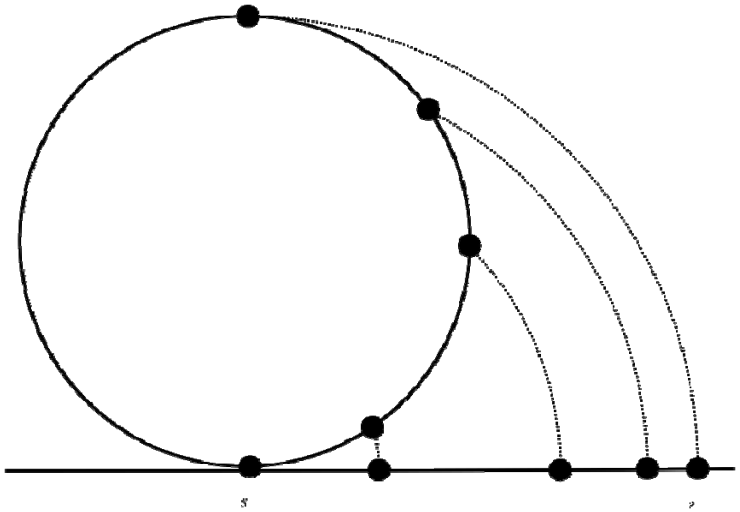
- Also known for the Lambert-Beer law

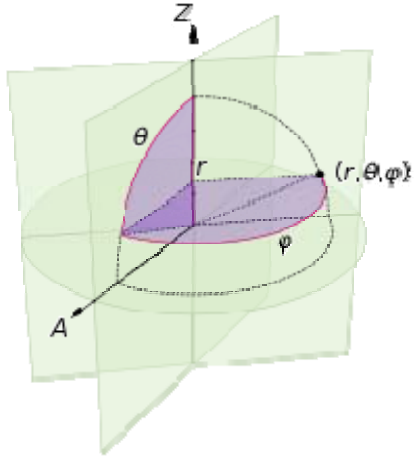
Lambert Projection of the Earth

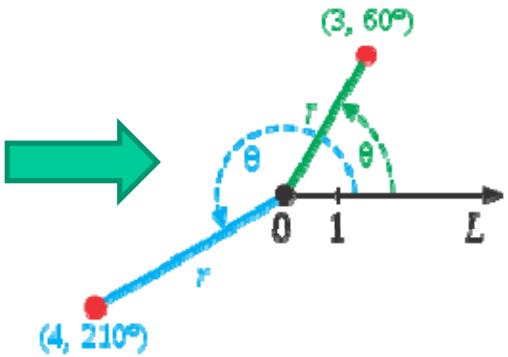




Lambert Azimuthal Equal Area Projection







Spherical coordinates, θ, ϕ

Polar coordinates, R, Θ

$$(R,\Theta)=\left(2\cos\left(\frac{\phi}{2}\right),\theta\right)$$

Lambert Azimuthal Equal Area Projection for Molecules

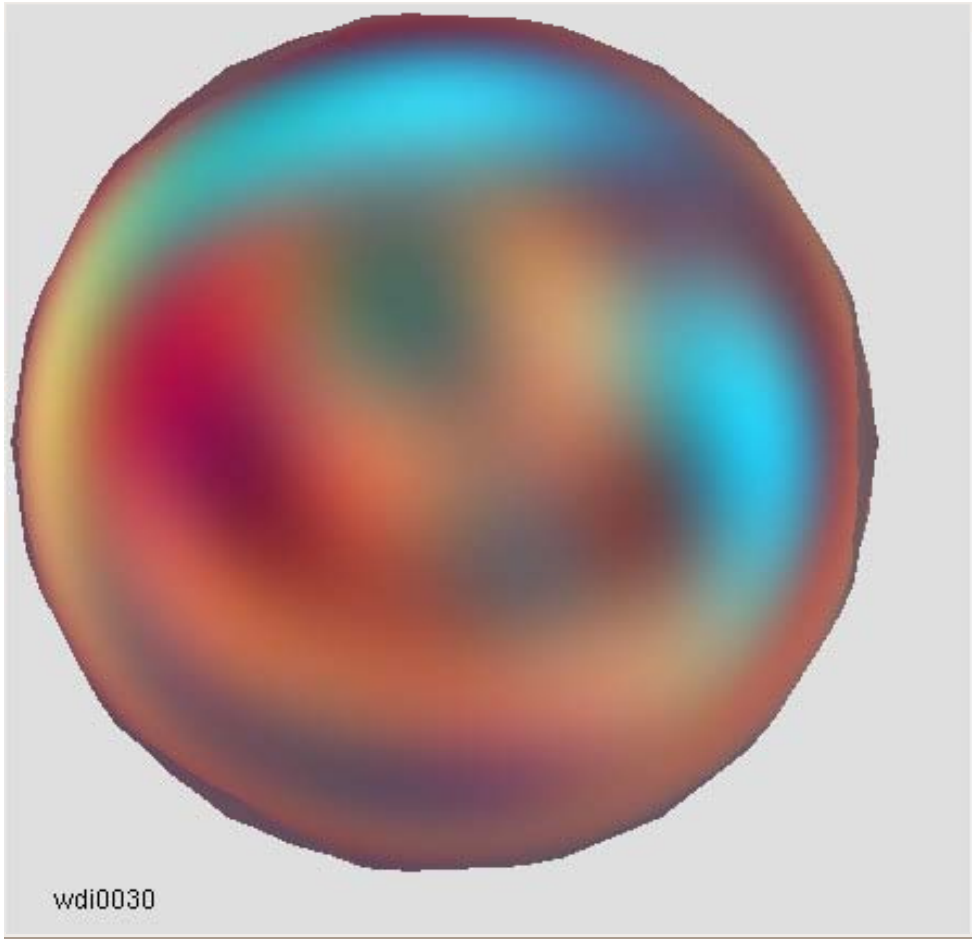
Lambert projection:

$$(R,\Theta) = \left(2\cos\left(\frac{\phi}{2}\right), \theta\right)$$

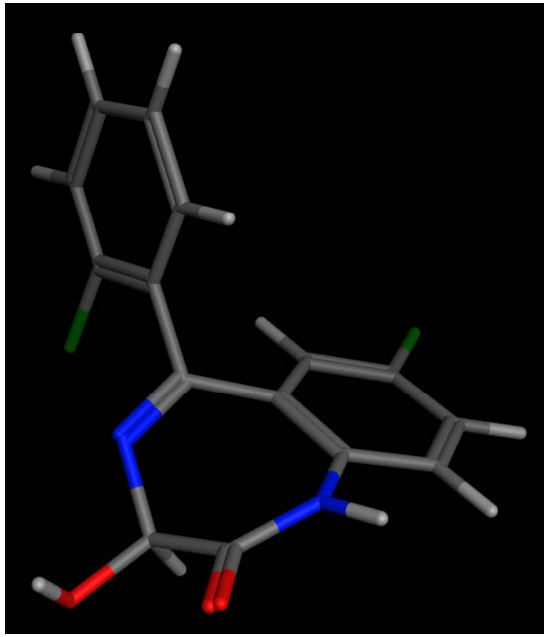
Spherical-harmonic representation:

$$r(\theta,\phi) = \sum_{l=0}^{N}\sum_{m=-l}^{l} a_{lm} y_{lm}(\theta,\phi)$$

RGB-Encoding

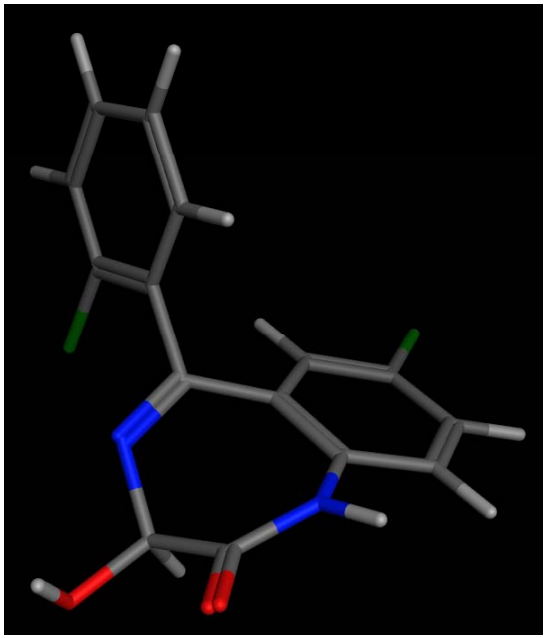


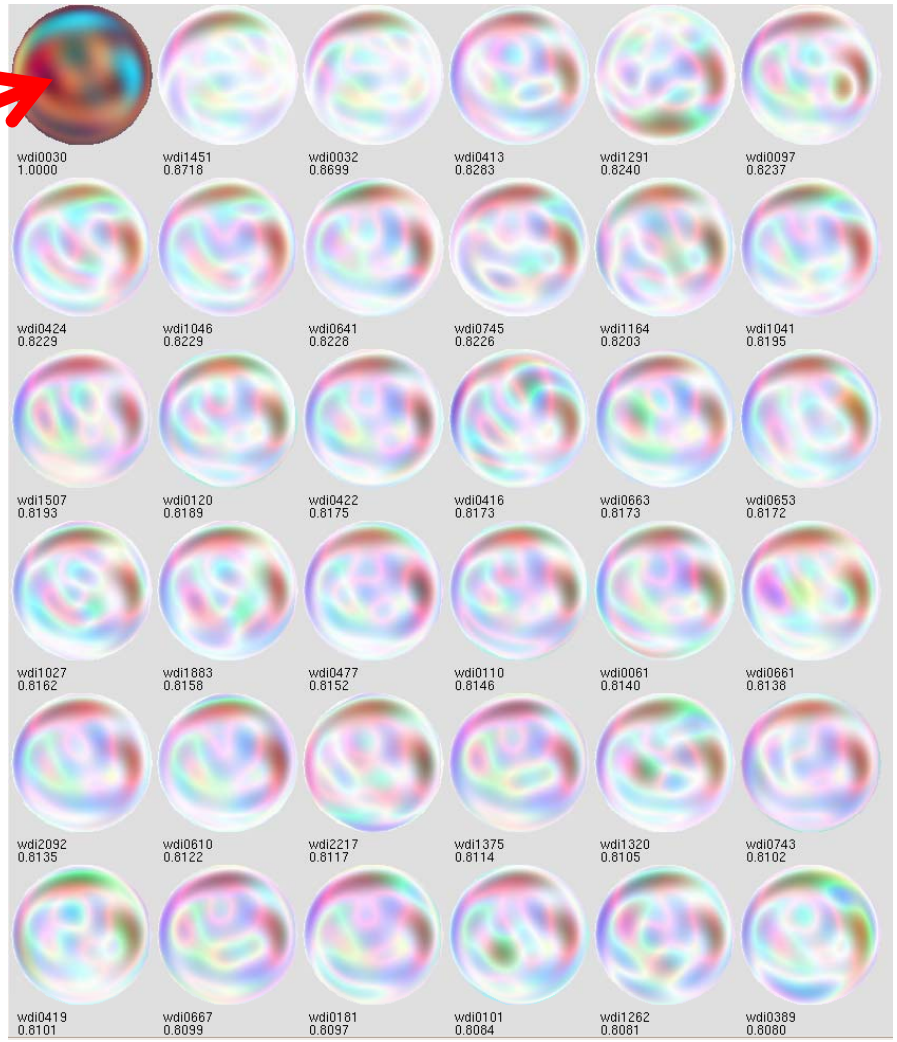
- **Lambert projection**
- Three properties encoded on the surface:
  - R=local ionization energy
  - G=local electron affinity
  - B=local polarizability



Similarities

Lead Compound





Similarities

CEPOS

Thank You