Electronic properties and intermolecular binding in drug-like molecules

Tim Clark

Computer-Chemie-Centrum
Friedrich-Alexander-Universität
Erlangen-Nürnberg
Germany
Tim.Clark@chemie.uni-erlangen.de

Centre for Molecular Design
University of Portsmouth
United Kingdom
Tim.Clark@port.ac.uk
Electronic properties and intermolecular binding in drug-like molecules

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Scaffold hops
- Binding features
Extreme Views of Molecules

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

“Cheminformatics”

Quantum Mechanics
# Molecular Features

<table>
<thead>
<tr>
<th>Molecular Feature</th>
<th>Defined Uniquely?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonds, bond orders etc.</td>
<td>NO</td>
</tr>
<tr>
<td>Atomic Monopoles, polarizabilities etc.</td>
<td>NO</td>
</tr>
<tr>
<td>Hydrogen-bond donors, acceptors</td>
<td>NO</td>
</tr>
<tr>
<td>Aromatic Rings</td>
<td>NO</td>
</tr>
<tr>
<td>Isodensity surface</td>
<td>YES</td>
</tr>
<tr>
<td>Molecular electrostatic potential (MEP)</td>
<td>YES</td>
</tr>
<tr>
<td>Local ionization potential (IE$_L$)</td>
<td>YES</td>
</tr>
<tr>
<td>Local Electron Affinity (EA$_L$)</td>
<td>(NO)</td>
</tr>
<tr>
<td>Local polarizability ((\alpha_L))</td>
<td>NO</td>
</tr>
</tbody>
</table>
Local Properties: MEP

\[ V(r) = \sum_{A} \frac{Z_{A}}{R_{A} - r} - \int \frac{\rho(r')}{|r' - r|} dr' \]

Local Properties: $\text{IE}_L$

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

Local Properties: $E_{A_L}$

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

Local Properties: $\alpha_L$ 

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


**Extreme views of molecules**
**Local properties**
**Descriptor-based models**
**Surface-integral models**
**CypScore**
**Spherical harmonics**
**Scaffold hops**
**Binding features**
Local Properties: Local Electronegativity, $\chi_L$

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

Local Properties: Local Hardness, $\eta_L$

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

Local Properties

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

- MEP
- $\alpha_L$
- $\chi_L$
- $\eta_L$
- $IE_L$
- $EA_L$
### Intermolecular Interactions

<table>
<thead>
<tr>
<th>Intermolecular Interaction</th>
<th>Local Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coulomb</td>
<td>Molecular electrostatic potential</td>
</tr>
<tr>
<td>Lewis acid – Lewis base (electron donor – acceptor)</td>
<td>Local ionization energy, local electron affinity</td>
</tr>
<tr>
<td>Dispersion</td>
<td>Local polarizability</td>
</tr>
<tr>
<td>Pauli repulsion</td>
<td>Electron density</td>
</tr>
<tr>
<td>Which element?</td>
<td>Local electronegativity, local hardness</td>
</tr>
</tbody>
</table>

• For each property, we can calculate the maximum, minimum, range, mean, variance, skew, kurtosis.

• ... and additionally for properties with positive and negative values the variance within the positive and negative areas, the proportion of each and the balance parameter.

• This gives a total of 68 descriptors in ParaSurf’09™.
Descriptors Based on Local Properties

- Describe additional properties (electron donor/acceptor characteristics) not captured by “classical” descriptor sets

- Do not require element-specific descriptors

- However, perform very similarly to “classical” descriptors for most properties

- This is possible because the performance of the models is limited by the accuracy of the experimental data
  - C. Kramer, B. Beck and T. Clark, *Insolubility classification with accurate prediction probabilities using a MetaClassifier*, to be submitted.
Surface-Integral Models

- Property obtained by integrating a function of the local properties over a molecular surface:

\[ P = \int_{\mathcal{O}} f(l_1, l_2, \ldots, l_n) d\mathcal{O} \approx \sum_{i=1}^{n_{tri}} f(l_1^i, l_2^i, \ldots, l_n^i) A_i^i + c \]


- Used with polynomials of the local properties:

• Total dataset of 10,814 compounds taken from the LOGKOW database
  – Sangster, J. LOGKOW - A databank of evaluated octanol-water partition coefficients (Log P), Sangster Research Laboratories: Montreal, Quebec, 2009.
  – Incorrect SMILES strings corrected, molecular structures checked
  – Limited to compounds with the elements H, C, N, O, F, P, S, Cl, Br and I.
  – Zwitterions treated separately.
  – Calculation protocol
  – SMILES → 3D-SDF file (CORINA. Molecular Networks)
  – Geometry optimized with AM1 or AM1* (VAMP10.0. Accelrys)
  – Isodensity marching-cube surface and local properties generated with ParaSurf’09 (Cepos InSilico)

• 1,350 random compounds selected as validation dataset
• Additional 767 compounds (Boehringer-Ingelheim in-house data) used as an additional validation dataset.
### Surface-Integral Model For LogPO\text{W}

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Set (N=9,464)</th>
<th>Validation Set (N=1,350)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MUE</td>
<td>RMSE</td>
</tr>
<tr>
<td>Descriptor model</td>
<td>0.67</td>
<td>0.88</td>
</tr>
<tr>
<td>Polynomial SIM</td>
<td>0.68</td>
<td>0.89</td>
</tr>
<tr>
<td>Binned SIM^a</td>
<td>0.51</td>
<td>0.68</td>
</tr>
<tr>
<td>Binned SIM without zwitterions</td>
<td>0.48</td>
<td>0.64</td>
</tr>
</tbody>
</table>

^a Uses only binned integrals of the local properties themselves and their pairwise products

(C. Kramer, B. Beck and T. Clark, *A Surface-Integral Model for logPO\text{W} and a Local Hydrophobicity*, to be submitted).
### Surface-Integral Model For LogPO\textsubscript{w}

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

<table>
<thead>
<tr>
<th>Model</th>
<th>LOGKOW Validation Set (N=1,350)</th>
<th>In-house Validation Set (N=767)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MUE</td>
<td>RMSE</td>
</tr>
<tr>
<td>Binned SIM</td>
<td>0.48</td>
<td>0.64</td>
</tr>
<tr>
<td>ACDLabs logP</td>
<td>0.26</td>
<td>0.45</td>
</tr>
<tr>
<td>ClogP (BioByte)</td>
<td>0.31</td>
<td>0.52</td>
</tr>
<tr>
<td>SlogP (MOE)</td>
<td>0.53</td>
<td>0.68</td>
</tr>
<tr>
<td>logP_o/w (MOE)</td>
<td>0.53</td>
<td>0.77</td>
</tr>
<tr>
<td>AlogP (TSAR)</td>
<td>0.62</td>
<td>0.86</td>
</tr>
</tbody>
</table>

- MUE: Mean unsigned error
- RMSE: Root mean square error
- R\textsuperscript{2}: Coefficient of determination
Local Hydrophobicity

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

\[
\begin{align*}
\text{m-dinitro-benzene} & \\
\log P(\text{exp}) &= 1.49 \\
\log P(\text{AM1}) &= 1.55 \\
\text{mesitylene} & \\
\log P(\text{exp}) &= 3.42 \\
\log P(\text{AM1}) &= 3.11
\end{align*}
\]
Local Hydrophobicity

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

**2,3,4,5-tetrachloroanisole**

\[
\text{logP(exp)} = 4.50-4.57 \\
\text{logP(AM1)} = 4.22
\]
... 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 almost 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 CypScore range
  - Restricted to Phase I reactions
Atomic Surface Descriptors

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. Works for any surface (isodensity, SES, marching tetrahedron or spherical harmonic)
5. Atoms are described by the statistical characteristics (max, min, mean, variance etc.) of the distribution of the local properties on “their” surface
6. Provides an excellent description of local reactivity
• **Dataset:**
  – MajorMetaboliteDB from Bayer Healthcare 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

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

Compound 1

Experimental negative
Experimental positive

Susceptibility to metabolism (-CypScore)

Compound 2

“Experimental false negatives”
• **Reaction centers:**
  - Csp³
  - Csp² (Aryl)
  - Csp² (C=X)
  - N-oxidation
  - NR₂ and NR₃

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

% false negative (blue) and % false positive (purple) atoms vs. cutoff score for CypScore
Influence of the Protonation State

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features
Results: “Public” validation Set

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

- 187 weak positions in 1,797 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
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 Healthcare**
Spherical Harmonics and ParaFit

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

\((N+1)^2\) Coefficients \(a_{lm}\) for a complete \(N^{th}\) order description


Shape Description

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

N = 15
• Extreme views of molecules
• Local properties
• Descriptor-based models
• Surface-integral models
• CypScore
• Spherical harmonics
• Scaffold hops
• Binding features

N=20
ParaFit Fast Alignment

DHF/Methotrexate

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

X-ray  local electron affinity

\[ \text{RMSD} = 0.58 \text{ Å} \]
Non-nucleotide HIV-1 reverse transcriptase inhibitors (NNRTIs, nerivapine, TIBO, aAPA)

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

**X-ray**  
local ionization energy  
RMSD = 0.47 Å

**Electron density (shape)**  
RMSD = 0.80 Å
Scaffold Hops

Search Protocol in PubChem

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features
• Search revealed nine high-quality new scaffold hops and more at a lower confidence level
• ParaFit is independent of the 2D-structure and was therefore used to calculate scaffold similarities
• Example: (PKA inhibitors)

19.9 μM  24.6 μM
### ParaFit Similarities

<table>
<thead>
<tr>
<th>Local Property</th>
<th>Similarity</th>
<th>Exit-vector RMSD (Å)</th>
<th>Overlay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>0.972</td>
<td>0.7</td>
<td>![Overlay Image]</td>
</tr>
<tr>
<td>$E_A$</td>
<td>0.920</td>
<td>1.5</td>
<td>![Overlay Image]</td>
</tr>
</tbody>
</table>
Comparison

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features
Identifying Binding Features

• Calculate H-Bond energies at B3LYP/6-311+G(d,p) for fixed positions of the H-Bond acceptor (grid)
  – Donor water; acceptor formaldehyde, water, hydroxide
  – Donor formamide, acceptor water

• Use a decision tree to distinguish between bonding and non-bonding positions around the donor

• Apply to detect H-bond donor sites (or their equivalents) in drug-like molecules.
H-Bond Donors

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features

- Water-water (-2 kcal mol\(^{-1}\))
- Water-ammonia (-2 kcal mol\(^{-1}\))
- Formamide-water (-2 kcal mol\(^{-1}\))
- Water-hydroxide (-10 kcal mol\(^{-1}\), wireframe -8 kcal mol\(^{-1}\))
H-Bond Acceptor Positions

- Extreme views of molecules
- Local properties
- Descriptor-based models
- Surface-integral models
- CypScore
- Spherical harmonics
- Scaffold hops
- Binding features
Acknowledgments: People

• **Local properties, descriptors**
  Marcel de Groot, Alexander Alex (Pfizer, Sandwich), Stefan Güßregen (Sanofi-Aventis, Frankfurt), Bernd Ehresmann, Kendall Byler

• **Surface-integral models**
  Bernd Beck (Boehringer-Ingelheim, Biberach), Bernd Ehresmann, Christian Kramer, Marcel Fouka Youmbi

• **CypScore**
  Andreas Göller, Alexander Hillisch (Bayer Healthcare, Wuppertal), Matthias Hennemann

• **Spherical Harmonics**
  David Ritchie (LORIA, Nancy), David Manallack (Victoria School of Pharmacy, Melbourne), Brian Hudson, Jr-Hung Lin

• **Scaffold Hops**
  Andreas Teckentrup (Boehringer-Ingelheim, Biberach), Kerstin Höhfeld

• **Binding Features**
  Matthias Hennemann, Christian Wick

• **ParaSurf™**
  Matthias Hennemann, David Whitley, Ray Crichton, Brian Hudson, Harald Mauser (Hofmann-la-Roche, Basel), Arjen Jakobi
Acknowledgments: Support

- Boehringer-Ingelheim, Biberach
- Bayer Healthcare, Wuppertal
- Hofmann-la-Roche, Basel

www.ceposinsilico.com