

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

#### 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





1

# 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 •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features



**Quantum Mechanics** 

"Cheminformatics"

Friedrich-Alexander-Universität Erlangen-Nürnberg



## Molecular Features

Molecular Feature	Defined Uniquely?
Bonds, bond orders etc.	NO
Atomic Monopoles, polarizabilities etc.	NO
Hydrogen-bond donors, acceptors	NO
Aromatic Rings	NO
Isodensity surface	YES
Molecular electrostatic potential (MEP)	YES
Local ionization potential (IE <sub>L</sub> )	YES
Local Electron Affinity (EA <sub>L</sub> )	(NO)
Local polarizability ( $\alpha_L$ )	NO





#### **Local Properties: MEP**

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

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

See, for instance, P. Politzer and J. S. Murray. *Molecular electrostatic potentials and chemical reactivity*. In K. Lipkowitz and R. B. Boyd, editors, *Rev. Comput. Chem.*, volume 2, 273. VCH, New York, **1998**.







# Local Properties: $\alpha_{L}$

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features



B. Ehresmann, B. Martin, A. H. C. Horn and T. Clark, *J. Mol. Model.* 2003, *9*, 342.
B. Martin, P. Gedeck, T. Clark, *Int. J. Quant. Chem.*, 2000, 77, 473.

0.23 Friedrich-Alexander-Universität Erlangen-Nürnberg









•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features



11

## Intermolecular Interactions

Intermolecular Interaction	Local Property
Coulomb	Molecular electrostatic potential
Lewis acid – Lewis base (electron donor – acceptor)	Local ionization energy, local electron affinity
Dispersion	Local polarizability
Pauli repulsion	Electron density
Which element?	Local electronegativity, local hardness





## **Descriptors From Local Properties**

- Use the statistics of the distribution of the values of the properties over the tesselation points on a molecular surface (J. S. Murray and P. Politzer, *J. Mol. Struct. (THEOCHEM)*, **1998**, *425*, 107-114).
- 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<sup>TM</sup>





#### **Descriptors Based on Local Properties**

- Describe additional properties (electron donor/acceptor characteristics) not captured by "classical" descriptor sets
  - B. Ehresmann, M. J. de Groot, A. Alex and T. Clark, *J. Chem. Inf. Comp. Sci.* 2004, 44, 658-668.
- 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
  - *Modelling the Chemistry: time to break the mould?*, T. Clark in EuroQSAR 2002: *Designing drugs and crop protectants*, M. Ford, D. Livingstone, J. Dearden and H. V. d. Waterbeemd (Eds) Blackwell Publishing, Oxford, **2003**, 111-121.
  - C. Kramer, B. Beck and T. Clark, *Insolubility classification with accurate prediction probabilities using a MetaClassifier*, to be submitted.





#### Surface-Integral Models

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

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

$$P = \int_{O} f(l_1, l_2, \dots, l_n) dO \approx \sum_{i=1}^{ntri} f(l_1^i, l_2^i, \dots, l_n^i) A^i + C \quad ????$$

- Richards, N. G. J.; Williams, P. B.; Tute, M., *IJQC: Quantum Biology Symposium*, **1991**; pp 299-316.
- Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Moeller, A.; Brickmann, J., *J Chem Inf Comp Sci* **1994**, 34, 1309-1319.
- Used with polynomials of the local properties:
  - Ehresmann, B.; de Groot, M. J.; Clark, T. Surface-Integral QSPR Models: Local Energy Properties. *J Chem Inf Model* 2005, 45, 1053-1060.





# Surface-Integral Model For LogPov

- 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  $\rightarrow$  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 LogPov

me of	Model	Training Set (N=9,464)			Validation Set (N=1,350)		
ules		MUE	RMSE	R <sup>2</sup>	MUE	RMSE	$\mathbb{R}^2$
riptor- models ce- al models core rical nics old hops ng es	Descriptor model	0.67	0.88	0.75	0.68	0.89	0.74
	Polynomial SIM	0.68	0.89	0.75	0.69	0.90	0.74
	Binned SIM <sup>a</sup>	0.51	0.68	0.85	0.52	0.70	0.84
	Binned SIM without zwitterions	0.48	0.64	0.86	0.48	0.64	0.86

<sup>a</sup> 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 logP<sub>ow</sub> and a Local Hydrophobicity, to be submitted).



•Extree views molect •Local proper •Describased •Surfa •CypS •Sphe harmo •Scaffo •Bindir feature



# Surface-Integral Model For LogPow

Model	LOGK Se	OW Valio t (N=1,35	dation 60)	In-house Validation Set (N=767)			
	MUE	RMSE	$\mathbb{R}^2$	MUE	RMSE	$\mathbb{R}^2$	
Binned SIM	0.48	0.64	0.86	0.85	1.10	0.53	
ACDLabs logP	0.26	0.45	0.94	1.03	1.36	0.47	
ClogP (BioByte)	0.31	0.52	0.92	0.86	1.36	0.47	
SlogP (MOE)	0.53	0.68	0.85	0.92	1.19	0.51	
logP_o/w (MOE)	0.53	0.77	0.82	1.00	1.28	0.49	
AlogP (TSAR)	0.62	0.86	0.79	0.97	1.24	0.50	







•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features



m-dinitro-benzene logP(exp) = 1.49 logP(AM1) = 1.55

Friedrich-Alexander-Universität Erlangen-Nürnberg **mesitylene** logP(exp) = 3.42 logP(AM1) = 3.11



19



•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features



#### 2,3,4,5-tetrachloroanisole

logP(exp) = 4.50-4.57logP(AM1) = 4.22

Friedrich-Alexander-Universuat Erlangen-Nürnberg



# CypScore

- ... predicts the likely sites of Cytochrome P450 metabolism in small molecules
- ... is a model for a hypothetical CYP P450 superenzyme
  - "Simulates" oxidation reactions by P450 enzymes
  - Currently almost no active-site accessibility restrictions
- ... is based AM1 and ParaSurf<sup>TM</sup>
  - 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**

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

#### 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

Friedrich-Alexander-Universität Erlangen-Nürnberg



# Models For Individual Metabolic Reactions

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

#### **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









#### **CypScore Models**

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

#### Reaction centers:

- Csp<sup>3</sup>
- Csp<sup>2</sup> (Aryl)
- $\operatorname{Csp}^2(C=X)$
- N-oxidation
- NR<sub>2</sub> and NR<sub>3</sub>

#### All centered and scaled

- CypScores range from -10 (very weak) to +10 (stable)
- A CypScore of zero marks equal probability of positive and negative
- M. Hennemann, A. Friedl, M. Lobell, J. Keldenich, A. Hillisch, T. Clark and A. H. Göller, *CypScore: Quantitative Prediction of Reactivity toward Cytochromes P450 Based on Semiempirical Molecular Orbital Theory, ChemMedChem*, **2009**, *4*, 657-669.









% false negative (blue) and % false positive (purple) atoms vs. cutoff score for CypScore

Friedrich-Alexander-Universität Erlangen-Nürnberg

•Extreme views of molecules •Local

Properties
Descriptorbased models
Surface-

Spherical

harmonics

integral models •CypScore

•Scaffold hops •Binding features



# **Influence of the Protonation State**







# **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

Friedrich-Alexander-Universität

Erlangen-Nürnberg

• 73% of metabolic positions found





#### tResults: in-house validation set

#### 39 compounds from 12 in-house projects

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

 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**

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

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

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

 ParaSurf: 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.

ParaFit: Toward high throughput virtual screening using spherical harmonic surface representations, L. Mavridis, B. D. Hudson and D. W. Ritchie, J. Chem. Inf. Model., 2007, 47, 1787-1796.





# **Shape Description**















#### ParaFit Fast Alignment

# DHF/Methotrexate

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features





local electron affinity

RMSD = 0.58 Å





#### ParaFit Fast Alignment

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features Non-nucleotide HIV-1 reverse transcriptase inhibitors (NNRTIs, nerivapine, TIBO, aAPA) D. T. Manallack, *J. Mol. Model.* **2008**, *14*, 797-805.

local ionization energy RMSD = 0.47 Å

Electron density (shape) RMSD = 0.80 Å







X-ray

# Scaffold Hops

## Search Protocol in PubChem



#### **Scaffold Hops**

 Extreme views of molecules I ocal properties Descriptorbased models Surfaceintegral 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







# ParaFit Similarities

	Local Property	Similarity	Exit- vector RMSD (Å)	Overlay
ls S	Shape	0.972	0.7	
	EAL	0.920	1.5	







•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features









Friedrich-Alexander-Universität Erlangen-Nürnberg



#### **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 •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features



40



•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features







41

#### 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<sup>™</sup>

Matthias Hennemann, David Whitley, Ray Crichton, Brian Hudson, Harald Mauser (Hofmann-la-Roche, Basel), Arjen Jakobi





42

# Acknowledgments: Support

•Extreme views of molecules •Local properties •Descriptorbased models •Surfaceintegral models •CypScore •Spherical harmonics •Scaffold hops •Binding features

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

# CEPOS INSILICO Science for Scientists

www.ceposinsilico.com



