Molecular Similarity and QSAR based on Molecular Surface Property Graphs

Vishwesh Venkatraman, David Whitley, Brian Hudson and Martyn Ford

Centre For Molecular Design, Institute of Biomedical and Biomolecular Sciences, University of Portsmouth, UK

Introduction

The identification of iso-functional but structurally dissimilar chemotypes (scaffold hopping) is an important aspect of drug design. Drug-receptor binding is influenced strongly by the behaviour of various properties at the molecular surface. Here we describe how molecular surface properties can be represented as a graph on the molecular surface that is independent of the underlying chemical structure and encapsulates both molecular shape and the property's major features. For implicitly defined surfaces, we show how these surface graphs may be used to obtain property-based alignments and derive QSARs.

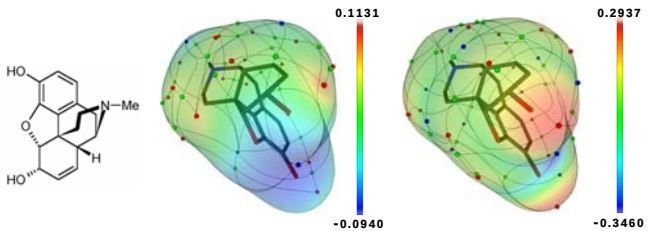

Figure 1. MEP (Gasteiger-Huckel charges) and Gaussian curvature MSPGs for Morphine

MOLECULE	PROPERTY	MAXIMA	MINIMA	SADDLES	χ
MORPHINE	Molecular Electrostatic Potential (MEP)	18	5	21	2
	Gaussian Curvature (GC)	15	12	25	2

Table 1. Critical points for the MEP and Gaussian curvature and the Euler Characteristic (χ)

Molecular Surface Property Graphs

A molecular surface **S** may be defined implicitly as the level set **S=F^{-1}(0)**, where **F** is a sum of atom centred Gaussians. For a property **P** defined on **S**, the gradient vector field **x'=grad P(x)** defines a graph on **S** called a molecular surface property graph or MSPG.

- The vertices of the MSPG are the critical points of **P**; *i.e.* the fixed points of the gradient field where **grad P(x) = 0**.
- These are classified as local **maxima**, **minima** and **saddles**. They are also the attractors (sinks), repellors (sources) and saddle points of the gradient field.
- The edges of the MSPG are the trajectories of the gradient field (lines of steepest ascent) that pass through the saddles.
- The vertices are labelled by the critical values of **P** and the edges by their arc length on **S**.

As an example, **Figure 1** shows the MSPGs for the *Molecular Electrostatic Potential (MEP)*, defined in terms of atomic partial charges, and a geometric property, the *Gaussian Curvature*, on the surface of morphine. To calculate an MSPG, the implicit representation of **S** is used to find local coordinate charts on **S**. Each chart is searched for critical points using Newton's method and the edges of the graph are found by numerically integrating the gradient field. To ensure a consistent set of critical points, a local test using a winding number around each chart and a global test using the **Euler Characteristic** are applied. **Table 1** illustrates the latter: the alternating sum of maxima - saddles + minima in each case is 2, the Euler Characteristic of a sphere.

3D Alignment and Similarity

A property-based alignment of two molecules is obtained by calculating their MSPGs, finding the *maximum common edge subgraph* (MCES) and performing a least squares fit of the corresponding vertices. In calculating the MCES only critical points of the same type are paired and constraints on the difference between critical values at paired vertices and the lengths of matched edges may be included. A numerical measure of similarity for a given property can be obtained as

$$sim(A,B) = \frac{(V(C)+E(C))^2}{(V(A)+E(A))(V(B)+E(B))}$$

where A and B are the MSPGs for the two molecules, C is their MCES, and V(A) and E(A) are the number of vertices and edges in A, *etc.*

MEP and mean curvature (MC) based alignments are shown for two thermolysin inhibitors 1THL and 1TLP in **Figures 2-4**. These results are in close agreement with those reported by Hofbauer et al[2] who used a more elaborate procedure.

1THL 1TLP

Figure 2. 2D structures for thermolysin inhibitors 1THL and 1TLP.

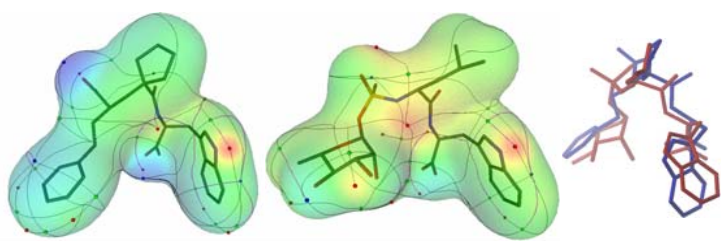

Figure 3. MEP alignment for 1THL and 1TLP. MCES shown in magenta. 10 edges matched with RMSD 1.3Å. MEP similarity = 0.09.

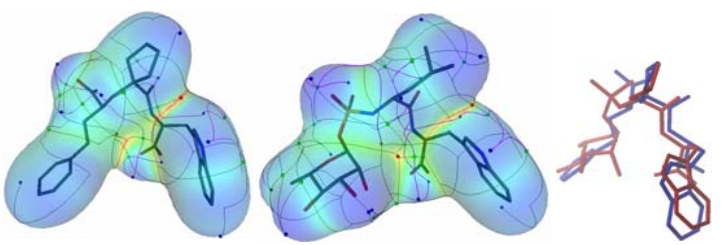

Figure 4. Mean curvature alignment for 1THL and 1TLP. MCES shown in magenta. 16 edges matched with RMSD 1.1Å. Curvature similarity = 0.16

QSAR

Many of the topological indices commonly calculated for molecular graphs can be adapted easily for use with MSPGs, by replacing atoms and bonds in the usual formulae with the critical points and connecting edges of MSPGs, and employing critical property values and edge lengths where appropriate. These indices can then be used as descriptors for developing QSARs.

To illustrate this, graphs based on the MEP, a Molecular Lipophilicity Potential (MLP) and the Gaussian and mean curvatures were calculated for a set of 25 D4-antagonists[3,4]. For each MSPG 49 topological indices were calculated. A variable selection procedure combined with continuum regression[5] (CR) was applied to the resulting 196 descriptors to derive the following QSAR for experimental pK_i values.

$$pK_i = 8.58 - 0.05\ HTI_{MEP} - 5.17\ QTOTMI_{MEP} - 0.0045\ ZM1_{MLP} + 0.1035\ IIMM_{MC}$$
$$(\pm 1.50) \quad (\pm 2.49) \quad (\pm 2.18E{-}03) \quad (\pm 6.7E{-}02)$$

	Description	Activity	HTI_{MEP}	$QTOTMI_{MEP}$	$ZM1_{MLP}$
HTI_{MEP}	Harmonic Topological Index	-0.4224			
$QTOTMI_{MEP}$	Sum of critical values of all minima	-0.5132	0.3596		
$ZM1_{MLP}$	Zagreb Index	-0.5656	0.2643	-0.0171	
$IIMM_{MC}$	Information content based on max-min composition	0.4028	-0.2063	-0.0529	0.0022

Table 2. Selected variables and correlation matrix

Method	R^2	Q^2	Model
COMFA-PLS[3]	0.99	0.75	7 variable, 9 component
Spherical harmonics – Stepwise regression[4]	0.89	0.75	4-variable
MSPG-Continuum regression	0.71	0.73	4-variable, 2 component

Table 3. Comparison of MSPG-CR model with literature models

The selected variables are described in **Table 2**, with their correlation matrix. **Table 3** compares the MSPG-CR result with two other QSARs for the same dataset. While the MSPG-CR model has a poorer fit than the other models it has similar predictive power. It does not require the alignment needed for the COMFA model and it uses atom-centred partial charges and lipophilicities so might have been expected to perform far worse than the semi-empirically based spherical harmonic model.

Summary and Future Directions

MSPGs provide a natural and concise representation of the behaviour of properties defined on molecular surfaces. This work illustrates how existing methods developed for treating chemical structure graphs can be applied to MSPGs with only minor modifications to address problems of molecular similarity, 3D alignment and QSARs.

Here, implicitly defined surfaces and atom-centred properties have been used. The same approach, however, can be applied to many other types of surfaces and properties. Current efforts are focussed on calculating MSPGs for spherical harmonic approximations to electron iso-density surfaces and quantum mechanics based properties.

REFERENCES

1. Whitley, D.C. (2002). Molecular Surface Property Graphs, *Euro QSAR*. Bournemouth, UK, 428-430.
2. Hofbauer, C., Lohninger, H. and Aszodi, A. SURFCOMP: A novel graph-based approach to molecular surface comparison, *J Chem. Inf. Model.* 2004;44:837-847.
3. Lanig, H., Utz, W. and Gmeiner, P. Comparative molecular field analysis of dopamine D4 receptor antagonists including 3-[4-(4-chlorophenyl)piperazin-1-ylmethyl]pyrazolo[1,5-a]pyridine 3-[4-(4-chlorophenyl)piperazin-1-ylmethyl]-1h-pyrrolo[2,3-b]pyridine (l-745,870), and clozapine. *J. Med. Chem.* 2001; 44, 1151–1157
4. Lin J-H, Clark T. An analytical, variable resolution, complete description of static molecules and their intermolecular binding properties. *J Chem. Inf. Model.* 2005;45:1010-1016
5. Whitley, D.C. , Ford, M.G. and Livingstone, D.J. Unsupervised Forward Selection: A Method for Eliminating Redundant Variables, *J Chem. Inf. Model.* 2000; 40; 1160-1168.