Structure-Based 3D Pharmacophores: An Alternative to Docking?

Gerhard Wolber*,
Johannes Kirchmair and Thierry Langer

*wolber@inteligand.com
Abstract & Outline

• **Pharmacophores & the Protein Data Bank**
  - 3D pharmacophore methodology
  - Primary data source: The Protein Data Bank
  - Motivation: Structure-based pharmacophore creation tool

• **LigandScout**
  - Ligand perception
  - 3D pharmacophore generation
  - Shared feature pharmacophores
  - Application example

• **Docking Comparison**
  - Compared active pose prediction
  - 58 relevant protein-ligand complexes
Structure-based pharmacophores

Pharmacophore models

**Pharmacophore** = Ensemble of universal chemical features that represent a specific mode of action in 3D

Chemical Features: Hydrogen bonds, charge interactions, hydrophobic areas
Why use structure-based pharmacophores?

- **Universal**
  Pharmacophores represent chemical functions, valid not only for the currently bound, but also unknown molecules

- **Computationally efficient**
  Due to simplicity (Suitable for virtual screening)

- **Comprehensive & Editable**
  Selectivity-tuning by adding or omitting feature constraints
PDB age!

The graph shows the growth in the number of PDB entries over time, with a significant increase starting around 1996. The y-axis represents the number of entries, measured in thousands, and the x-axis represents the years from 1972 to 2004.
LigandScout: A structure-based pharmacophore creation tool

Structure-based pharmacophore creation from all PDB complexes:

1. Extract, identify and interpret ligands (hybridization states, bonds)
2. Create pharmacophores
3. Visualize, allow user interaction and export for virtual screening
Hybridization state determination

Quantitative Geometry Templates
for all geometry types:

- $sp^3$: tetrahedral
- $sp^2$: trigonal planar
- $sp$: linear

Align along the first two points, numerically turn to match the third point
Geometry templates: Better than bond angles?
Hybridization state: Error determination

\[ d_a = \sum_{i=0}^{n} \sqrt{(I_i - O_i)^2} \]

\[ d_r = \frac{d_a}{n} \]

- \( d_a, d_r \) relative/absolute geometric deviation
- \( I_i \) ideal template positions
- \( O_i \) neighbor atom positions
- \( n \) number of atoms
Hybridization state: Error determination
Ring geometry is different

Planar rings show different bond angles than non-ring sp² atoms: all planar ring atoms are to be sp² hybridized
Using PCA for planarity detection

Distance from PCA plane < 0.4 A
Double bond distribution among sp² atoms

- No exact solution in many cases (e.g., Keto-enol tautomere)

- Use of patterns explicitly covering all known cases from the view of a central atom

- Weighted distribution of the maximum number of double bonds for the rest of the cases (nonbipartite maximum matching)

Patterns by Roger Sayle: Bioinformatics Group, Metaphorics LLC, Santa Fe, see http://www.daylight.com/meetings/mug01/Sayle/m4xbondage.html
Nonbipartite weighed matching

- **Double Bond Distribution along adjacent sp² paths**
  - Create bond classes: Identify longest and shortest bonds with non-linear geometry
  - Shortest bonds: high weights
  - Apply maximum number of double bonds using weighed nonbipartite complete matching
Pharmacophore creation

Chemical Features that are likely to occur in the complex:

- Hydrogen Bond Donors
- Hydrogen Bond Acceptors
- Negative Ionizable Areas
- Positive Ionizable Areas
- Hydrophobic Interactions

**Vectors:** Direction and Distance constraint

**Location Spheres:** Distance constraint only

Chemical features always refer to the ligand side.
Chemical feature constraints

Distance Constraints

Relation between two points, one located on ligand side, one on macromolecular side.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-Bond</td>
<td>2.5-3.8 Å</td>
</tr>
<tr>
<td>Charge Transfer</td>
<td>1.5-5.6 Å</td>
</tr>
<tr>
<td>Hydrophobic</td>
<td>1.0-5.8 Å</td>
</tr>
</tbody>
</table>

Direction Constraints

Relation between two atom groups, one located on ligand side, one on macromolecular side.

Groups form a rigid reference geometry, which are the basis for a directed vector.

Result: one tolerance sphere on ligand side
Chemical feature constraints: Rigid H-bonds
Chemical feature constraints: Flexible H-bonds

\[ \alpha = \delta - a \sin \left( \frac{b \cdot \sin \chi}{c} \right) \]
### Chemical feature universality layers

<table>
<thead>
<tr>
<th>Layer 1</th>
<th>Layer 2</th>
<th>Layer 3</th>
<th>Layer 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subgraph</strong></td>
<td><strong>Without geometry constraint</strong></td>
<td><strong>Including geometry constraint</strong></td>
<td><strong>Without geometry constraint</strong></td>
</tr>
<tr>
<td><strong>Chemical Function</strong></td>
<td><strong>Hydroxylic group, Phenol Group</strong></td>
<td><strong>Hydrogen bond Donor/Acceptor</strong></td>
<td><strong>Lipophilic area, positive ionizable area</strong></td>
</tr>
<tr>
<td><strong>Chemical Function</strong></td>
<td><strong>Including geometry constraint</strong></td>
<td><strong>Without geometry constraint</strong></td>
<td><strong>Including geometry constraint</strong></td>
</tr>
<tr>
<td><strong>Subgraph</strong></td>
<td><strong>Phenol group facing a parallel benzene</strong></td>
<td><strong>Hydrogen bond Donor/Acceptor</strong></td>
<td><strong>Lipophilic area, positive ionizable area</strong></td>
</tr>
</tbody>
</table>

LigandScout creates pharmacophores using the universal Layer 3 and Layer 4 features.
Application example: Gleevec

Gleevec in PDB complex
1IEP, 1OPJ; variant 1FPU
Shared feature pharmacophore

1iep

1fpu

1opj
Pharmacophore overlaying

Pharmacophore model derived from one single bound ligand may not be able to retrieve other related compounds …

Starting set: Several ligand-protein complex pharmacophores

Creation of compatibility graphs

Maximum clique detection

Feature alignment

Calculation of combined features

… new shared feature pharmacophore
Shared feature pharmacophore
Shared feature pharmacophore

Exported to **Catalyst** using hypoedit tool:

- 4 lipophilic aromatic areas
- 2 hydrogen bonding interactions
Virtual screening setup

• Virtual screening using Catalyst
• 3 Databases:
  o **PDB singleConf**: all PDB ligands with one single entry per conformation [67k]
  o **PDB multiConf**: all PDB ligands with one single entry per unique molecule and 50 conformers each (multiConf; 50 FAST) [7k]
  o **Maybridge 2003** (multiConf) [55k]
Virtual screening results

### Gleevec
shared feature pharmacophore

<table>
<thead>
<tr>
<th>Database</th>
<th>Hits</th>
<th>Drug-like hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB singleConf (~67k)</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>PDB multiConf (~7k)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Maybridge (~55k)</td>
<td>19</td>
<td>7</td>
</tr>
</tbody>
</table>
LigandScout summary

LigandScout

- Extracts and interprets ligands and their protein environment from PDB files
- Automatically creates and visualizes 3D pharmacophore models
- Creates overlaid “shared feature” pharmacophores to broaden the scope of a single model
Docking comparison

Is it possible to predict the active pose of a ligand using a 3D pharmacophore?

Is fitting ligands to structure-based 3D pharmacophores as accurate as docking?
## Method comparison: Discussion

<table>
<thead>
<tr>
<th>Pharmacophores</th>
<th>Docking</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Pharmacophore biased to specific binding mode (multi-feature binders less)</td>
<td>• Not biased to bound ligand</td>
</tr>
<tr>
<td>• Editable</td>
<td>• Generic – might detect different binding locations and modes</td>
</tr>
<tr>
<td>• Fully automated</td>
<td>• Black Box</td>
</tr>
<tr>
<td>• <strong>Suitable for virtual screening</strong> (60,000 compounds in minutes)</td>
<td>• Pre-processing necessary</td>
</tr>
<tr>
<td>• Conformer generation might become a limit</td>
<td>• <strong>Suitability for VS questionable</strong> (30 sec to minutes per compound)</td>
</tr>
</tbody>
</table>
Docking and Pharmacophore Fitting

1. Docked bio-active ligands into 58 pharmacologically relevant complexes [1] using FlexX and Gold

3. Generated unbiased conformers and fitted into LigandScout hypotheses using Catalyst (maxConfs=50, FAST) [2]

5. Compared best fitting conformation to best scored docking pose (CScore, GoldScore)


Docking and fitting

Docking into active site with FlexX and GOLD

Fitting to pharmacophore in original coordinate frame
1CKP

RMS = 0.63

< 1.5: „perfect fit“
1DI8
RMS = 2.18
< 3.5: „acceptable fit“
1G49
RMS = 3.69

3.5 < RMS < 6:
„inadequate fit“
1UPJ
RMS = 3.78

LigandScout case „inadequate fit“

But bioisosteric!
1KZN

RMS = 8.76

„inacceptable fit“

> 6
Results

<table>
<thead>
<tr>
<th>Category</th>
<th>FlexX (CScore worst RMS)</th>
<th>FlexX (CScore best RMS)</th>
<th>GOLD (GoldScore)</th>
<th>LigandScout (FAST, maxConfs=50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1.5</td>
<td>35</td>
<td>25</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>1.5 to 3.5</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>3.5 to 6</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>more than 6</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Cumulative percentage

- **GOLD (GoldScore)**
- **LigandScout** (FAST, maxConfs=50)
- **FlexX** (CScore worst RMS)
- **FlexX** (CScore best RMS)

X-axis: < 1.5, 1.5 to 3.5, 3.5 to 6, more than 6
Y-axis: 0%, 20%, 40%, 60%, 80%, 100%
Results summary

• More than 80% of the LigandScout complex fits are below an RMS of 3.5!

• „Binding site bias“ can be seen as an advantage

• Better conformer generation might further improve results
Conclusions

- 3D pharmacophores perform considerably well in predicting poses
- Accuracy is comparable to docking (with fewer complete failures)
- Virtual screening using 3D pharmacophores is much faster (pre-sampled multi-conformer databases)

>> Structure-based 3D pharmacophores are a viable alternative to docking!
• G. Wolber and T. Langer. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.*; 2005; 45; 160-169

Acknowledgements

Alois Dornhofer
Robert Kosara
Martin Biely
Fabian Bendix
Eva Maria Krovat
Theodora Steindl

Thierry Langer
Johannes Kirchmair
Christian Laggner
Daniela Schuster
Oliver Funk