3D QSAR Methods: Phase and Catalyst Compared

David Evans
Lilly Research Laboratories
UK-QSAR 18th October 2006
Introduction

- **HypoGen** from Catalyst (Accelrys) and **Phase** (Schrodinger).

- Both aim to quantitatively predict activity using a training set of compounds with known activity.

- Both search for common 3D pharmacophores amongst active compounds in the training set — known as *hypotheses*.
Estimating Activity?

• HypoGen uses the distance between the pharmacophore features (site points) in the hypotheses and matching feature in the compounds.

Active compound – features centred on site points

Less active compound – features cannot be centred
Estimating Activity?

• **Phase** scores compounds using a grid-based 3D QSAR model built from the compounds aligned on the hypotheses.

**Active compound – more blue regions**  
**Less Active compound – less blue**
HypoGen Methodology

1. **Select Training Set**
   - Minimum of 16 compounds, spread of 4 orders of magnitude in activity, includes representatives of different structural classes.

2. **Generate Conformers**
   - CHARMM-like force field, ‘poling’ method to give maximum sampling of conformational space.

3. **Find possible Hypotheses from Actives**
   - ‘Constructive’: Active compounds are those within a threshold of the maximum activity (9-fold by default).
   - Finds all hypotheses which match a minimum (default 4) number of site points in all these actives.

4. **Reject some Hypotheses using Inactives**
   - ‘Subtractive’: Inactives are those with activity 3.5 orders of magnitude below the most active compound.
   - Any hypothesis which matches more than half of these inactives is rejected.

5. **Refine and Score Hypotheses**
   - ‘Optimization’: All remaining hypotheses are scored.
   - Cost = Error + Feature Weight + Complexity
   - Hypotheses are refined by adjusting the site point positions to minimize the cost, using simulated annealing.
Phase Methodology

Select Training Set

No guidelines given, expect HypoGen rules will be useful.

Generate Conformers

MacroModel used, OPLS force field and rapid torsional search used by default.

Find possible Hypotheses from Actives

User selects active compounds (manually or with numerical cutoffs).
User chooses required number of site points and the number of active compounds which every hypothesis must match.

Score Hypotheses

Hypotheses are scored by the geometric alignment of site points in the actives to site points in the hypotheses.
Hypotheses not refined, each suggested hypothesis comes from one conformer of one of the active compounds.

Build QSAR model

For each selected hypothesis, all compounds are aligned and a 3D QSAR model is constructed based on the position of either atom (default) or pharmacophore features. User selects number of PLS coefficients to fit.
User can define training and test set at this stage, with option of random selection, and cross-validation statistics are calculated.
Data Sets

• Eight public data sets. $K_i$ or IC50 values.

• Chosen without reference to previous 3D QSAR work (with exception of ETA set).

- CDK2
  - 52 compounds
  - MPS 0.62

- Mean Pairwise Similarity (MPS):
  - The average Tanimoto Coefficient of MACCS fingerprints across all pairs of compounds in the set (MOE).

www.bindingDb.org
www.cheminformatics.org

• Aim to have similar sized test set and training set
• Assess based on predictions on test set.
Data Sets

- Glutaminyl cyclase
  - 63 compounds
  - MPS 0.47
- CDK1
  - 79 compounds
  - MPS 0.64
- VEGFR2
  - 63 compounds
  - MPS 0.77
- ETA
  - 61 compounds
  - MPS 0.77
- COX2
  - 81 compounds
  - MPS 0.62
- DHFR
  - 73 compounds
  - MPS 0.69
- BZR
  - 73 compounds
  - MPS 0.47
Preparing Training Sets

Catalyst ‘rules’ require:

- 4 orders of magnitude spread of activity
- No similar compounds which are also similar in activity.

Lilly ‘CatScan’ program makes training sets and runs automated Catalyst jobs.

5. Cluster with fingerprint similarity.

6. Include more than one compound per cluster only if they differ significantly in activity.
Generating Models

• Each program: 5 jobs for each data set. Command-line scripts.

• Phase:
  1. Default
  2. Pharmacophore QSAR grid
  3. Cutoff for defining active (1-3 pK units)
  4. 3 + Hypothesis scoring to include activity of reference ligand
  5. 3 + Lower tolerance of matching molecules to site points

• HypoGen:
  1. Default
  2. Minimum spacing of site points.
  3. Weight and tolerances of scoring different features vary
  4. 2 + 3
  5. Add excluded volume as extra site point.
## Results on Test Sets

- 5 x 10 top scoring hypotheses → Predict on Test Set
- Results are presented for the highest $R^2_{\text{test}}$.  

<table>
<thead>
<tr>
<th>System</th>
<th>Phase</th>
<th>HypoGen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CDK2</td>
<td>-0.34</td>
<td>-0.27</td>
</tr>
<tr>
<td>GlutCyc</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>CDK1</td>
<td>0.37</td>
<td>0.46</td>
</tr>
<tr>
<td>VEGFR2</td>
<td>0.18</td>
<td>-0.69</td>
</tr>
<tr>
<td>ETA</td>
<td>0.13</td>
<td>-0.43</td>
</tr>
<tr>
<td>COX2</td>
<td>0.18</td>
<td>0.13</td>
</tr>
<tr>
<td>DHFR</td>
<td>0.34</td>
<td>0.26</td>
</tr>
<tr>
<td>BZR</td>
<td>0.26</td>
<td>0.11</td>
</tr>
<tr>
<td>Mean</td>
<td>0.15</td>
<td>-0.12</td>
</tr>
</tbody>
</table>
Example Correlations

Glutaminyl Cyclase Phase ‘1’ hypothesis

$R^2 = 0.02$

Pearson $r = 0.78$

3D QSAR model projected on most active compound
Results on Test Sets

- Pearson $r$-values
- Measure correlation, not exact match. *Amber* 99.9% chance $r > 0$.

<table>
<thead>
<tr>
<th>System</th>
<th>Phase</th>
<th>HypoGen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CDK2</td>
<td>0.32</td>
<td>0.62</td>
</tr>
<tr>
<td>GlutCyc</td>
<td>0.78</td>
<td>0.52</td>
</tr>
<tr>
<td>CDK1</td>
<td>0.65</td>
<td>0.78</td>
</tr>
<tr>
<td>VEGFR2</td>
<td>0.67</td>
<td>0.41</td>
</tr>
<tr>
<td>ETA</td>
<td>0.51</td>
<td>0.36</td>
</tr>
<tr>
<td>COX2</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td>DHFR</td>
<td>0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>BZR</td>
<td>0.52</td>
<td>0.39</td>
</tr>
<tr>
<td>Mean</td>
<td>0.56</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Comments

• Important to test all hypotheses to find good model, not just top scoring.

• Experimentation with more extensive conformational sampling in both programs did not yield significantly better results.

• Phase gives statistics excluding compounds which do not match the hypothesis
  • Feature as initial step to identifying different classes/binding modes
  • Need to write own code for fair comparison

• Catalyst prediction much slower than Phase (with Best fit), harder to test as many hypotheses.
Transplant Conformers

Assess whether poor conformer generation or poor scoring was responsible for the poor models.

Phase BZR training set compounds aligned with best hypothesis

Input to Catalyst as training set conformer library

Good Catalyst Model?

NO

Pearson $r = 0.43$

• Indicates that scoring is the problem
• Programs unable to model activity even if compounds are input in ‘correct’ overlays?
CDK2: Structure-Based Comparison

- CDK2 data set: PDB structure (2C5P) of the most active compound in complex with receptor.

- Does the conformer generation in the two programs find the crystallised conformation?

  Phase best alignment RMS = 2.4 Å  
  Catalyst best alignment RMS = 3.1 Å
CDK2: Alignments to X-ray structure

- Generate alignments of all compounds to the crystal conformation with ROCS (OpenEye). Shape + feature matching.
- Input these to Phase and Catalyst.

ROCS alignment

Overlay still not convincing – different series of compounds not overlaid correctly?

Input to Phase

Best $r_{test} = 0.51$

Input to HypoGen

Best $r_{test} = 0.56$
CDK2: Alignments to X-ray structure

• Restrict to one series of compounds

Input to Phase

Best $r_{\text{test}} = 0.46$

Input to Catalyst

Best $r_{\text{test}} = 0.56$

Input to CoMFA

$r_{\text{test}} = 0.42$

$r_{\text{LOO}}$ on entire data set = 0.35
Conclusions

• Phase $R^2$ and Pearson $r$ values higher than Catalyst on average.

• Evaluated recommended strategies for improving models.

• Hypothesis: Is the problem with scoring & not conformations?
  • If so why?
    • Data sets just too small to build model?
    • Multiple binding modes for ligands and proteins.
    • $\Delta G$ depends on free ligand state as well as bound state.
Acknowledgements

Lilly, Erl Wood
• Mike Bodkin
• Dave Thorner
• Sandra Milutinovic
• Linda Hirons

Lilly, Indianapolis
• Thompson Doman

Lilly, Alcobendas
• Howard Broughton

Accelrys
• Katalin Nadassy

Schrodinger
• Jas Gata-Aura
• Stuart Murdock
• Steve Dixon

www.bindingDb.org
www.cheminformatics.org
Appendix: 50/50 Train/Test

- 5 x 10 top scoring hypotheses → Predict on Test Set
- Results are presented for the highest $R^2_{\text{test}}$.

<table>
<thead>
<tr>
<th>System</th>
<th>Phase Original</th>
<th>Phase 50/50</th>
<th>Best</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>CDK2</td>
<td>-0.34</td>
<td>-0.27</td>
<td>0.06</td>
<td>-0.20</td>
</tr>
<tr>
<td>GlutCyc</td>
<td>0.11</td>
<td>0.04</td>
<td>0.13</td>
<td>0.20</td>
</tr>
<tr>
<td>CDK1</td>
<td>0.37</td>
<td></td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>VEGFR2</td>
<td>0.18</td>
<td>-0.69</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>ETA</td>
<td>0.13</td>
<td>-0.43</td>
<td>0.16</td>
<td>0.28</td>
</tr>
<tr>
<td>COX2</td>
<td>0.18</td>
<td>0.13</td>
<td>0.18</td>
<td>0.23</td>
</tr>
<tr>
<td>DHFR</td>
<td>0.34</td>
<td>0.26</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>BZR</td>
<td>0.26</td>
<td>0.11</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td>Mean</td>
<td>0.15</td>
<td>-0.12</td>
<td>0.24</td>
<td>0.30</td>
</tr>
</tbody>
</table>
# Appendix: 50/50 Train/Test

- Pearson \( r \)-values

- Measure correlation, not exact match. **Amber** 99.9\% chance \( r > 0 \).

<table>
<thead>
<tr>
<th>System</th>
<th>Phase Original</th>
<th>Phase 50/50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CDK2</td>
<td>0.32</td>
<td>0.62</td>
</tr>
<tr>
<td>GlutCyc</td>
<td>0.78</td>
<td>0.52</td>
</tr>
<tr>
<td>CDK1</td>
<td>0.65</td>
<td>0.78</td>
</tr>
<tr>
<td>VEGFR2</td>
<td>0.67</td>
<td>0.41</td>
</tr>
<tr>
<td>ETA</td>
<td>0.51</td>
<td>0.36</td>
</tr>
<tr>
<td>COX2</td>
<td>0.45</td>
<td>0.41</td>
</tr>
<tr>
<td>DHFR</td>
<td>0.58</td>
<td>0.52</td>
</tr>
<tr>
<td>BZR</td>
<td>0.52</td>
<td>0.39</td>
</tr>
<tr>
<td>Mean</td>
<td>0.56</td>
<td>0.40</td>
</tr>
</tbody>
</table>
VEGFR2: Alignments to X-ray structure

- VEGFR2 data set has analogous ligand in the PDB (1YWN).
**VEGFR2: Alignments to X-ray structure**

- Compare alignment with best Phase hypothesis \((r_{\text{test}} = 0.67)\) to crystal

- Alignment good apart from orientation of urea.

- No reason to expect program to get this correct in absence of receptor structure.

- Urea interacts with Glu \(\text{CO}_2^-\) in protein

Orange = crystal

Urea oriented incorrectly
VEGFR2: Alignments to X-ray structure

Original Results:

<table>
<thead>
<tr>
<th></th>
<th>Phase</th>
<th>Catalyst</th>
</tr>
</thead>
<tbody>
<tr>
<td>VEGFR2</td>
<td>0.67</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Input to Phase

Best $r_{test} = 0.55$

Input to HypoGen

Best $r_{test} = 0.62$

Input to CoMFA

$r_{test} = 0.41$

$r_{LOO}$ on entire data set = 0.62

ROCS alignment