Marcus Gastreich

Fragment Based *De Novo* Design
Credits

- Val Gillett & John Holliday
- Matthias Rarey & Jörg Degen
- Hans Briem (Bayer-Schering)
- Markus Boehm (Pfizer US)

BioSolveIT colleagues:
- Ingo Dramburg
- Markus Lilienthal
- Holger Claußen
The Tempting Idea

Generating “interesting” compounds...
... is a search problem in entire chemical space.

=> We need to search efficiently.

Strategy: Generate de novo compounds from **fragments**

Key Ingredients:

1. Interesting Fragments
2. Fast and targeted assembly
Fragments
Fragments Can Form Drugs

Gleevec
Not All Fragments Can Be Connected

=> we need LINK compatibility rules
Linker Compatibility Matrix

@link_types 12
L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12

@link_terminals
# link  group   bond  blen  tangle
L1  [O-]  1     1.280  *
L2  [H]   1     1.009  *
...  
L6  [CH3] 1     1.510  *
L7  [CH2] 2     1.340  180
...  
L12 [CH3] 1     1.780  *

@link_connects
# link1 link2 bond   blen  tangle
L1  L2    am   1.355  180
L1  L3    1     1.362  *
L2  L6    am   1.355  180
L2  L12   1     1.656  *
L3  L4    1     1.362  *
L4  L5    1     1.469  *
L4  L8    1     1.469  *
L4  L9    1     1.469  *
L4  L10   1     1.469  *
L7  L7    2     1.316  180
L11 L11   1     1.473  180
Input for Compatibility Rules

- Chemistry Knowledge
- Combinatorial Chemistry
  - Corporate
  - Literature

Constraint for both: synthesisisability
No, we cannot guarantee synthesisability nowadays.

Increase the chance of being synthetically accessible by:

**Clever Shredding**
- Retrosynthetic approaches (e.g., RECAP, Lewell et al., JCICS, 38, 511-522 (1998))
- experience, corporate knowledge
- duplicate removal, filtering

**Massive Filtering (before assembly & on-the-fly)**
- Toxic groups & other knowledge
- Lipinski-like approaches
- property violation thresholds
CoLibri: Fragment Generation & More

mol2
SD
sln
combi mol2

shredder (SMIRKS)
&
filter (SMARTS)

bookkeeping &
redundancy removal

ligand based
FTrees

structure based
FlexX / FlexNovo

mol2
SD
sln
combi mol2
Assembly 1

ligand based
FTrees
Fast, Fuzzy Searching: FTrees

- Molecule = Tree
  - ✓ no bit string
  - ✓ no 3D conformers
  - ✓ Nodes = chemistry
    + physics
  - ✓ Topology preserved

=> “Fuzzy” Searching, “Scaffold Hops”

FTrees Fragment Spaces: Generation

**RECAP RULES**

Fragment Space
FTrees Fragment Spaces: Assembly

- Convert fragments
- Make linkage with special linker nodes
- Combined tree loses linkers
- Back to molecule
FTrees Fragment Spaces: Searching

query

fragment space

dynamic programming *


*deterministic:
- same result for every run
- always identify best mol
  fast: 1 msec/comparison

**Experimental Setup:**

- approx. 35,000 member subset of WDI
- RECAP-like shredding [1]
- maximum number of fragments connected: 5

**Consequences:**

- 16,780 fragments
- a search in $10^{18}$ compounds!

**Results Delivery for this search:**

= a couple of minutes

## WDI & Histamine H1 Antagonists

- **Target similarity level = 0.9**

<table>
<thead>
<tr>
<th>Query</th>
<th>Generated structure</th>
<th>Known Active</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimethindene</td>
<td><img src="image1" alt="Chemical Structure" /></td>
<td><img src="image2" alt="Chemical Structure" /></td>
</tr>
<tr>
<td>Mianserin</td>
<td><img src="image3" alt="Chemical Structure" /></td>
<td><img src="image4" alt="Chemical Structure" /></td>
</tr>
</tbody>
</table>
Pfizer Inhouse Application

- Markus Boehm & colleagues, ACS 2006, S.F.:

- Ca. 1.600 FTrees FS Searches w/ inhouse Comobilbs
  - 1.4 Mio frags, 89.221 unique, FS Generation: 12hrs
  - in 96% of the cases, 343 buried actives were retrieved;
    in 85% of the cases these were on ranks 1-10.
  - Typical search times: between 1 and 15 mins.
  - M.B. gave more than a dozen examples of structures

- Paper in preparation!
FTrees Pipeline

- Web interface for Feature Trees available (free)

  http://www.biosolveit.de/FTreesWeb

- A Graphical User Interface
  - currently in beta stage
  - screening
  - clustering
  - visualisation of the actual Feature Trees alignments

- FTrees in PipelinePilot®
  - available for testing
Assembly 2

structure based
FlexX / FlexNovo
FlexX
FlexX Docking Basics

- Incremental construction
  1. Cut Ligand @ rotatable bonds
  2. Select so-called “base fragment”
  3. Place base fragment
  4. Build-up full complex

- Hundreds of successful virtual screening applications [1]

- Extensions
  “induced fit”: FlexE
  screening: FlexX-Screen
  pharmacophores: FlexX-Pharm
  combinatorics: FlexX\textsuperscript{c}


FlexX has been cited more than 600 times (Source: Web of Science, 1996 JMB article only.)
FlexX - Incremental Buildup
FlexX^C - Combinatorial Docking

**FlexX^C** re-cycles placements

→ tremendous speed up!

© 2007 BioSolve IT GmbH    Marcus Gastreich @ 4th Sheffield Chemoinformatics Conference 20 June 2007
Applications

- inspired by Leach\(^1\): **Place-and-combine on-the-fly**
- Joint project with Schering \(^2\)
- Proof of Concept:
  Dock 22.4 member combilib under pharmacophore constraints

\(^1\) Leach et. al, JMGM 18, 358-367, 2000
\(^2\) Gastreich et al., JCAMD 20 (12) 2006, 717-734.
Generic “Gleevec” Library

Size of combinatorial library: $4 \times 20 \times 20 \times 7 \times 20 \times 20 \times 5 = 22,400,000$ cpds
Filtering the Gleevec Library

- Lipinski “Rule-of-5”-like filter
  - molecular weight < 500
  - number of NH or OH ≤ 5
    (mimicking donors)
  - number of N or O ≤ 10
    (mimicking acceptors)

⇒ 2 constraints must be fulfilled

- No aza-compounds
Results

# cpds. docked: ~70,000 (~0.3% of all)
Rank of closest Gleevec analog: 141 (within Top-0.0005% of all)
Total CPU time: 60h (3h on 20 nodes)
Average run time per cmpd.: 0.001s (vs. 22.4 Mio)
3s (vs. all actually docked)

closest Gleevec analog
Xray structure
Best-ranked:

Gleevec in thinner representation
FlexNovo
FlexX Novo Project Background

- Novel development by Jörg Degen at Rarey‘s Group

- Continuous development in framework of NovoBench project

  Consortium:
  - Altana (now Nycomed)
  - Eli Lilly
  - BioSolveIT
  - ZBH Hamburg (Rarey)
  - Molecular Networks / CCC Erlangen (Gasteiger)
  - 4SC Munich

- Available for beta-testing through BioSolveIT
  (it‘s on the CDs !)
FlexNovo - The Program

Task:
- Find high-scoring drug-like molecules for the target protein

Input:
- Protein structure, [pharm. constraints]
- Fragment space, [property filter criteria]

FlexNovo:
- Based on FlexX-kernel
- Deals with large fragment spaces
- Contains on-the-fly evaluated property, placement and diversity filters
1. Placement of all fragment space fragments (FlexX-Pharm)

2. Extension-Loop ($k$-greedy):
   - Select highest-scoring fragments
   - Add all compatible fragments
   - Filter by score, property, diversity
   - Placement by incremental construction
   - Apply placement filter

3. Results
   - Select highest-scoring molecules
   - Filter by property, diversity
Application Scenarios

Targets
- Dihydrofolate reductase (4dfr)
- Cyclooxygenase-2 (6cox)
- Estrogen receptor (1err)
- Cyclin-dependant kinase 2 (1di8)

Parameters (for each target)
- 2 extension cycles
- 100 intermediate solutions per cycle
- 50 placements per intermediate solution
- 50 molecules in the final solution list
- Target-specific property filters
- No ‘further’ parameter tuning
- Two different fragment spaces
  (Fragments from complexed ligand contained)
Dihydrofolate Reductase

DHFR complexed with Methotrexate (4dfr)

- 2 representative solutions
- hand-picked from 50 molecules
- resulting from a single calculation
- standard parameters used
Cyclooxygenase-2

COX-2 complexed with SC558 (6cox)

- 2 representative solutions
- hand-picked from 50 molecules
- resulting from a single calculation
- standard parameters used
Estrogen Receptor

ER complexed with Raloxifene (1err)

- 2 representative solutions
- hand-picked from 50 molecules
- resulting from a single calculation
- standard parameters used
Cyclin-Dependent Kinase 2

CDK2 complexed with Q4SP1298 (1di8)

- 2 representative solutions
- hand-picked from 50 molecules
- resulting from a single calculation
- standard parameters used
Serendipity?

CDK2 complexed with Staurosporine (1qpd)

CEP-1347
Lundbeck, Cephalon
PKC/JNK/MLK
formerly Phase III

Result for 1di8
No filter criteria used
Easier to synthesize

Technical Issues

Fragment spaces

WDI space
17000 fragments

Filtering

WDIsub space
4500 fragments

Some statistics

<table>
<thead>
<tr>
<th></th>
<th>WDI</th>
<th>WDIsub</th>
</tr>
</thead>
<tbody>
<tr>
<td># of compounds evaluated</td>
<td>1.2 M - 1.5 M</td>
<td>350 k - 478 k</td>
</tr>
<tr>
<td># of compounds placed</td>
<td>28 k - 62 k</td>
<td>24 k - 62 k</td>
</tr>
<tr>
<td>Total process time</td>
<td>7.0 h - 15.5 h</td>
<td>5.0 h - 10.0 h</td>
</tr>
<tr>
<td>Total memory usage</td>
<td>750 MB</td>
<td>200 MB</td>
</tr>
</tbody>
</table>

→ per second, 1-2 compounds can be placed

Timings with standard parameters on a single CPU workstation computer
Wrap-Up

- Find new compounds from fragment space searches.

- Fragment Spaces now more easily accessible (CoLibri).

- Redundancies of $n$ libraries are taken care of.

- Ligand based (FTrees) and structure based approaches (FlexX/FlexNovo) perform very well in validation studies.

Thank you!