The importance of understanding ligand-binding sites in proteins

- **Rapid site identification and ranking:** Locate binding sites in the entire protein whose size, functionality, and extent of solvent exposure to assess their propensity for ligand binding.

- **Site visualization tools:**
  - Highlight regions within the binding site suitable for occupancy by hydrophobic groups or by ligand hydrogen-bond donors, acceptors, or metal-binding functionality.
  - Distinguishing the different binding site sub-regions allows for ready assessment of a ligand's complementarity.

- **Tools for exploiting targets of opportunity:** Affinity maps in these pockets show where modifications to a ligand structure would be expected to promote binding.

- **Integration with docking programs:** Identified sites can easily be used to set up virtual screening experiments for structure-based drug design work.
Validated Targets + Combinatorial Chemistry Libraries

THE "DOCKING" PROBLEM

Ligand-Receptor Complexes

CASTp:
A Server for Identification of Protein Pockets & Cavities
- Identifies all pockets and cavities.
- Measures the volume and area analytically.

http://sts.bioengr.uic.edu/castp/

A software tool for analysis and visualization of tunnels and channels in protein structures

http://www.caver.cz/

"THE DOCKING PROBLEM"

- SITE/LIGAND REPRESENTATION (treatment of H atoms?)
- JUXTAPOSITION OF THE LIGAND AND SITE FRAMES OF REFERENCE (docking engine)
- EVALUATION OF COMPLEMENTARITY (scoring functions)

AIM: To obtain the lowest free energy structure(s) for the receptor-ligand complex
H++ is an automated system that computes pK values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment. Given a (PDB) structure file on input, H++ outputs the completed structure in several common formats (PDB, PQR, AMBER inpcrd/prmtop) and provides a set of tools useful for analysis of electrostatic-related molecular properties.

Drug Design by the Method of Receptor Fit

Peter J. Goodford
Laboratory of Molecular Biophysics, Oxford, England. Received October 17, 1983

What, then, does the method of receptor fit offer for a future in which doctors may have access to every one of every patient?

- One should affinity
- System enable
- Sequence specific
- All receptors are different until proved identical.

These tentative forecasts point toward:

- a new generation of more potent, specific, effective therapeutic agents with less toxicity, reduced side effects, and fewer aberrant responses, which is what people and society at large are seeking.
- more costly research, which is the price that must be paid.

One last conclusion seems very probable. Mountaineers climb because the mountains are there and offer them a worthwhile challenge, and scientists will try to design drugs to fit receptors for similar reasons.
Trimethoprim (TMP), a widely used antibacterial drug, is a potent inhibitor of bacterial DHFRs but a much weaker inhibitor of the vertebrate enzymes (e.g., IC₅₀ values against Escherichia coli and human enzyme are, respectively, 5 x 10⁻⁹ M and 3 x 10⁻⁴ M).

To provide information on the action of this drug at the molecular level, we have determined the structure of the binary complex of E. coli (strain RT500) form I DHFR with TMP and compared it with that of the complex of DHFR with methotrexate (MTX), a drug which binds tightly to both bacterial and vertebrate DHFR. The structure of our TMP-enzyme complex differs from that in [Science 1977; 197, 452-455] of an MTX-enzyme complex from a different strain (MB1428) of E. coli. The amino acid sequences of the two enzymes are currently thought to differ at 3 positions.
"GRID: A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules"

Peter Goodford, Oxford University


*ibid.* 32, 1083-1094 (1989); 36, 140-147 (1993); 36, 148-156 (1993)

http://www.moldiscovery.com/

---

**Aromatic carbon probe**

- Grid point value range: -5.45 to 5.0 kcal/mol
- Contour level: -2.5 kcal/mol

**Hydrophobic probe**

- Grid point value range: -2.86 to 0.0 kcal/mol
- Contour level: -1.0 kcal/mol

**Carbonyl oxygen probe**

- Grid point value range: -8.03 to 5.0 kcal/mol
- Contour level: -5.0 kcal/mol

**Hydroxyl oxygen probe**

- Grid point value range: -12.30 to 5.0 kcal/mol
- Contour level: -7.0 kcal/mol

Didemnin B bound to human elongation factor eEF-1A

Marco, E.; Martín-Santamaría, S.; Cuevas, C.; Gago, F.


http://www.moldiscovery.com/
Prokaryotic elongation factor EF-Tu

PDB code 2C78

PDB code 2C77

http://farmamol.uah.es/

Open3DGRID
An open-source software aimed at high-throughput generation of molecular interaction fields (MIFs)

Paolo Tosco, a, b Thomas Balle b

 a Department of Drug Science and Technology, University of Turin, via Pietro Giuria 9, 10125 Turino, Italy
 b Faculty of Pharmacy, University of Sydney, Pharmacy Building (A15), Camperdown Campus, Sydney NSW

http://open3dgrid.sourceforge.net/

Molecular Interaction Fields

The Basic Principles of GRID

1.1 Introduction
1.2 Philosophy and Objectives
1.3 Priorities
1.4 The GRID Method
1.4.1 GRID Probes Are Anisometric
1.4.2 The Target "Responds" to the Probe
1.4.3 The Target is Immersed in Water
1.5 The GRID Force Field
1.5.1 The Lennard-Jones Terms
1.5.2 The Electrostatic Terms
1.5.3 The Hydrogen Bond Terms
1.5.4 The Other Terms
1.6 Nomenclature
1.6.1 "ATOM" Records
1.6.2 "HEATM" Records
1.7 Calibrating the GRID Force Field
1.7.1 Checking the Calibration
1.7.2 Checking Database GRUB
1.8 The Output from GRID
1.8.1 GRID Maps from Macromolecules
1.8.2 GRID Maps from a Small Molecule
1.9 Conclusions
Docking algorithms

- Require 3D atomic structure for protein, and 3D structure for compound (“ligand”)
- May require initial rough positioning for the ligand
- Will use an optimization method to try and find the best rotation and translation of the ligand in the protein, for optimal binding affinity

Molecular Docking

- **SYSTEMATIC SEARCH** (brute force algorithm):
  - All binding orientations of all conformers of the ligand and the receptor (impractical for most situations).

- **AUTOMATED SEARCH**:
  - **GEOMETRIC METHODS**: Matching of ligand and receptor site descriptors (descriptors, grids, fragments...).
  - **FORCE FIELD METHODS**: Minimizing the ligand-receptor interaction energy - Molecular dynamics and Monte Carlo simulations.

What is Docking?

- “Best way(s) to put two molecules together”
- Three steps:
  1. Definition of the **structure** of the target molecule.
  2. Location of the **binding site**.
  3. Determination of the **binding mode**: are there any conformational changes in the ligand and/or the receptor upon binding?

What is Docking?

- “Best way(s) to put two molecules together”
  - Need to **quantify** solutions for **ranking**;
  - **Scoring function**: force field, knowledge-based, empirical.
- “Best ways to put two molecules together.”
  - (plural) Experimental structure may be amongst one of **several predicted solutions**.
- “Best way(s) to put two molecules together.”
  - Need a **search method**
**Scoring functions**

**Force field-based:** calculation of van der Waals and electrostatic interaction energies between the receptor and the ligand atoms

**Knowledge-based:** statistical analysis of 3D complex structures to derive a sum of potentials of mean force between receptor and ligand atoms

**Empirical:** the binding free energy is broken down into a number of different weighted contributions (supposed to be additive: number of hydrogen bonds, ionic interactions, apolar contacts, entropy penalties...)

### Empirical Scoring Functions

The binding free energy is broken down into a number of different weighted contributions (supposed to be additive: number of hydrogen bonds, ionic interactions, apolar contacts, entropy penalties...)

<table>
<thead>
<tr>
<th>Functional Form</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>PMF</th>
<th>Functional Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta G_{\text{PMF}} = \Delta G_{\text{B,bind}} + \sum_{i=1}^{n} f(\Delta R_i, \Delta x_i) + \Delta G_{\text{olv}} + \sum_{i=1}^{n} f(\Delta R_i, \Delta x_i) + \Delta G_{\text{olv}}$</td>
<td>$PMF = \sum \left[ \sum_{i=1}^{n} \left( \frac{b_i}{k_i} \right) \right] - kT \ln \left[ \sum_{i=1}^{n} \left( \frac{b_i}{k_i} \right) \right] \left( \frac{P_{\text{mol}}}{P_{\text{mol}}^0} \right)$</td>
</tr>
</tbody>
</table>

where $a_i$ is the Boltzmann constant; $b_i$ are the volume correction factors; $c_i$ and $d_i$ are the radial distribution functions for a protein pair $(i,j)$ (ligand-ligand) or a protein pair $(i,j)$ (protein-ligand).

<table>
<thead>
<tr>
<th>Doxygen v3.45</th>
<th>Functional Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta G_{\text{D}} = \sum_{i=1}^{n} f(\Delta R_i, \Delta x_i) + \Delta G_{\text{olv}} + \sum_{i=1}^{n} f(\Delta R_i, \Delta x_i) + \Delta G_{\text{olv}}$</td>
<td>$\Delta G = \sum \left( -kT \ln \left( \frac{P_i}{P_j} \right) \right) \left( 1 - \frac{\Delta x_i}{\Delta x_j} \right)$</td>
</tr>
</tbody>
</table>

where $a_i$ and $b_i$ are the distance-dependent pairwise potentials, and $c_i$ are the adjustable weight factors, typically set to 0.5.
SuperStar

- Calculate binding positions for specific probe atoms in protein active sites
- Identify functional groups in binding-site
- Look up relevant IsoStar scatterplots and overlay on functional groups
- Contour - combining by taking products

Examples of algorithms to dock a ligand into a receptor site

**Rigid ligand:**
- Fast shape matching (DOCK)

**Flexible ligand:**
- Shape matching (DOCK 4.0)
- Incremental construction (FlexX)
- Simulated annealing (AutoDock 2.4)
- Monte Carlo simulation (MCDOCKER)
- Genetic algorithm (AutoDock 3.0, GOLD, GAMBLER)

Some popular docking programs

- **DOCK**
  - Developed in Tak Kuntz’s group at UCSF
  - Shape algorithm – [http://www.cmpharm.ucsf.edu/kuntz/dock.html](http://www.cmpharm.ucsf.edu/kuntz/dock.html)
  - Recent versions allow for ligand flexibility

- **GOLD**
  - Developed at Sheffield University, distributed by CCDC
  - Uses genetic algorithm
  - Flexible ligand - [http://www.ccdc.cam.ac.uk/](http://www.ccdc.cam.ac.uk/)

- **FLEXX**
  - Flexible ligand – [http://www.biosolveit.de/FlexX/](http://www.biosolveit.de/FlexX/
  - Binding mode prediction and virtual high-throughput screening (vHTS)

- **FRED**
  - Rigid, but able to use multiple, well chosen conformers
  - Very fast

- **AUTODOCK**
  - Uses Genetic Algorithm

- **LIGANDFIT**

SuperStar features

- Cavity detection
- Surface or pharmacophore point display
- Metal coordination
- Hyperlinking to IsoStar scatterplots
- Choice of CSD- or PDB-based maps
- Gaussian fits

map for aromatic CH carbon probe generated at the binding site of the protein-ligand complex 1CPS.
"A Geometric Approach to Macromolecule-Ligand Interactions"

"Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding Site of Known Three-Dimensional Structure"

"Automated Docking with Grid-Based Energy Evaluation"
E. C. Meng, B. K. Soichet, I. D. Kuntz
J. Comp. Chem. 13, 505-524 (1991)
AutoDock: Why Use Grid Maps?

- **AutoGrid** computes grid maps
  - Representation of macromolecule
    - Regular orthogonal lattice of points
  - Ligand ‘probe’ samples force field
  - One map for each ligand atom type

- AutoDock uses **trilinear interpolation**
  - to compute interaction energy between ligand and target

- Non-bonded energy is pre-calculated

- **Saves time**: ~100x faster than traditional non-bonded pair list method

AutoGrid Grid Box

- Grid box depends on:
  - Orientation with respect to protein.
  - Where should I center the grid box?
    - Center on ligand;
    - Center on macromolecule;
    - Pick atom;
    - Type in x-, y- and z-coordinates.
  - Spacing (0.2 Å - 1.0 Å; default 0.375 Å).
  - Specify an **Even** Number of x-, y-, z-points (2×2×2 - 126×126×126).
- `% makebox mol.gpf > mol.gpf.box.pdb`

Ligand Flexibility

- **Set Root of Torsion Tree**:
  - By interactively picking, or
  - Automatically.
    - Smallest ‘largest sub-tree’.

- **Interactively Pick Rotatable Bonds**:
  - No ‘leaves’;
  - No bonds in rings;
  - Can freeze:
    - Peptide/amide/selected/all;
  - Can set the number of active torsions that move either the most or the fewest atoms

Choose the Docking Algorithm

- **SA.dpf** → **Monte Carlo Simulated Annealing**
- **GA.dpf** → **Genetic Algorithm**
- **LS.dpf** → **Local Search**
  - Solis-Wets (SW)
  - Pseudo Solis-Wets (pSW)
- **GALS.dpf** → **Genetic Algorithm with Local Search, i.e. Lamarckian GA**
AutoDock 3 Scoring Function

\[ \Delta G_{\text{binding}} = \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} + \Delta G_{\text{hbond}} + \Delta G_{\text{desolv}} + \Delta G_{\text{tors}} \]

- **\( \Delta G_{\text{vdW}} \)**
  12-6 Lennard-Jones potential
- **\( \Delta G_{\text{elec}} \)**
  Coulombic with Solmajer-dielectric
- **\( \Delta G_{\text{hbond}} \)**
  12-10 Potential with Goodford Directionality
- **\( \Delta G_{\text{desolv}} \)**
  Stouten Pairwise Atomic Solvation Parameters
- **\( \Delta G_{\text{tors}} \)**
  Number of rotatable bonds

Viewing Conformational Clusters by RMSD

- List of available RMSD tolerances
  - Separated by spaces
- Histogram of conformational clusters
  - Number in cluster versus energy
- Pick a cluster
  - Makes a list of the conformations in that cluster;
  - Makes this the current sequence for states player.

Sample AutoDock output for GSK3β

Docked energy

Number in cluster

Cluster Rank
Kenpaullone

Docked energy

Number in cluster

Sample AutoDock output for GSK3β

AutoDock/Vina plugin:

to set up docking runs and view the docking results.

http://www.pymolwiki.org/index.php/AutoDock_plugin

AutoDock/Vina plugin:

http://www.biosolveit.de/FlexX/

PROGRAM FlexX

Main applications:

(1) Binding mode prediction

For a protein with known three-dimensional structure and a small ligand molecule, FlexX predicts the geometry of the protein-ligand complex and estimates the binding affinity in less than 15 seconds.

(2) Virtual high-throughput screening (vHTS)

With FlexX a database consisting of ~100,000 compounds can be screened in about 8 hours on a 30-node cluster – fully automated

Algorithmic details:

- Incremental construction.
- The conformational flexibility of the ligand is taken into account
- The MIMUMBA database is used for determination of low-energy torsion angles, while an interaction geometry database is used to exactly describe intermolecular interaction patterns. For scoring, FlexX uses an adapted Böhm’s function.
**FLEXX: Incremental construction**

- Select rigid portion as the base fragment
- Dock the base fragment into the receptor site, optimizing steric and electrostatic interactions.
- Sequentially add the remaining ligand fragments.

**FLEXX: Evaluation of the interaction energy**

\[
\Delta G = \Delta G_0 + (\Delta G_{\text{rot}} \times N_{\text{rot}}) + \Delta G_{\text{hb}} \sum f(\Delta R, \Delta \theta) + \Delta G_{\text{ion}} \sum f(\Delta R, \Delta \theta) + \Delta G_{\text{aro}} \sum f(\Delta R, \Delta \theta) + \Delta G_{\text{lipo}} \sum f(\Delta R, \Delta \theta)
\]

- Loss of entropy during ligand binding
- Hydrogen bonds between neutral atoms
- Ion bridges and ionic hydrogen bonds
- Interactions between aromatics
- Lipophilic contacts (mainly van der Waals)

\[0 \leq f(\Delta R, \Delta \theta) \leq 1\] Geometry penalty function

---

**Program GOLD**


- Product of a collaboration between the University of Sheffield, GlaxoSmithKline plc and CCDC
- Uses a genetic algorithm for optimization
- Can output multiple solutions (i.e. output multiple final population members)
- Full ligand and partial protein flexibility
- Fitness function combination of four elements:
  - protein-ligand hydrogen bond energy (*external H-bond*)
  - protein-ligand van der Waals (vdw) energy (*external vdw*)
  - ligand internal vdw energy (*internal vdw*)
  - ligand torsional strain energy (*internal torsion*)

**Genetic Algorithms**

- Create a “population” of possible solutions, encoded as “chromosomes”
- Use “fitness function” to score solutions
- Good solutions are combined together (“crossover”) and altered (“mutation”) to provide new solutions
- The process repeats until the population “converges” on a solution
How GAs Work

A way is found of encoding possible solutions into a bitstring (chromosome), and of specifying the 'goodness' of a chromosome (fitness function)

1. Initialize a population of chromosomes
2. Evaluate the fitness of each chromosome
3. Create new chromosomes from the current population
4. Delete population members to make room for new ones
5. Evaluate the new chromosomes and put them in population
6. If we want to keep going, go back to step 3

Genetic Algorithms

- For our purpose, we can encode rotation and translation of a molecule, and bond torsion angles in a chromosome, e.g.:

\[ T_X \ T_Y \ T_Z \ R_X \ R_Y \ R_Z \ \tau_1 \ \tau_2 \]

where we have 3 translation values (T), 3 rotation values (R) and as many torsion angles (\( \tau \)) as the molecule has rotatable bonds

Genetic Algorithms

- Initially, our population will be initialized with random values, e.g.

<table>
<thead>
<tr>
<th></th>
<th>( T_X )</th>
<th>( T_Y )</th>
<th>( T_Z )</th>
<th>( R_x^\circ )</th>
<th>( R_y^\circ )</th>
<th>( R_z^\circ )</th>
<th>( \tau_1^\circ )</th>
<th>( \tau_2^\circ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-3.2</td>
<td>-1.6</td>
<td>4.5</td>
<td>130</td>
<td>126</td>
<td>228</td>
<td>131</td>
<td>114</td>
</tr>
<tr>
<td>C2</td>
<td>2.8</td>
<td>1.3</td>
<td>-4.6</td>
<td>97</td>
<td>231</td>
<td>149</td>
<td>126</td>
<td>144</td>
</tr>
<tr>
<td>C3</td>
<td>-8.7</td>
<td>2.9</td>
<td>3.1</td>
<td>143</td>
<td>261</td>
<td>12</td>
<td>83</td>
<td>29</td>
</tr>
<tr>
<td>C4</td>
<td>-2.2</td>
<td>-2.9</td>
<td>-3.6</td>
<td>27</td>
<td>280</td>
<td>141</td>
<td>312</td>
<td>216</td>
</tr>
<tr>
<td>C5</td>
<td>5.8</td>
<td>4.1</td>
<td>4.9</td>
<td>19</td>
<td>25</td>
<td>26</td>
<td>341</td>
<td>18</td>
</tr>
<tr>
<td>C6</td>
<td>0.3</td>
<td>-2.7</td>
<td>5.6</td>
<td>14</td>
<td>81</td>
<td>27</td>
<td>155</td>
<td>75</td>
</tr>
<tr>
<td>C7</td>
<td>4.4</td>
<td>-0.3</td>
<td>-0.2</td>
<td>12</td>
<td>46</td>
<td>22</td>
<td>26</td>
<td>98</td>
</tr>
</tbody>
</table>

Fitness Function

- Used to score chromosomes to determine “goodness”
- For our purposes, we are concerned with how well the molecule in a particular orientation binds to the protein
- So a fitness function for a docking GA might be a combination of the following elements:
  - Energy (binding, potential)
  - Number and strength of hydrogen bonds formed
  - Hydrophobic effects
  - Electrostatic effects
Fitness Function

- To score a chromosome, the GA will place the molecule inside the protein using the given translation, rotation and torsion parameters, and the fitness function will calculate the score based on an analysis of the joint 3D structure.

Fitness function scoring of population

- Initially, our population will be initialized with random values, e.g.

<table>
<thead>
<tr>
<th></th>
<th>T_x</th>
<th>T_y</th>
<th>T_z</th>
<th>R_x</th>
<th>R_y</th>
<th>R_z</th>
<th>t_x</th>
<th>t_y</th>
<th>t_z</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-3.2</td>
<td>-1.6</td>
<td>4.5</td>
<td>130</td>
<td>126</td>
<td>228</td>
<td>131</td>
<td>114</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>2.8</td>
<td>1.3</td>
<td>-4.6</td>
<td>97</td>
<td>231</td>
<td>149</td>
<td>126</td>
<td>144</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>-8.7</td>
<td>2.9</td>
<td>3.1</td>
<td>143</td>
<td>261</td>
<td>12</td>
<td>83</td>
<td>29</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>-2.2</td>
<td>-2.9</td>
<td>-3.6</td>
<td>27</td>
<td>280</td>
<td>141</td>
<td>312</td>
<td>216</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>5.8</td>
<td>4.1</td>
<td>4.9</td>
<td>19</td>
<td>25</td>
<td>341</td>
<td>18</td>
<td>75</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>0.3</td>
<td>-2.7</td>
<td>5.6</td>
<td>14</td>
<td>81</td>
<td>27</td>
<td>155</td>
<td>75</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>4.4</td>
<td>-0.3</td>
<td>-0.2</td>
<td>12</td>
<td>46</td>
<td>22</td>
<td>26</td>
<td>98</td>
<td>0.61</td>
<td></td>
</tr>
</tbody>
</table>

Create new population members

- Initially, our population will be initialized with random values, e.g.

<table>
<thead>
<tr>
<th></th>
<th>T_x</th>
<th>T_y</th>
<th>T_z</th>
<th>R_x</th>
<th>R_y</th>
<th>R_z</th>
<th>t_x</th>
<th>t_y</th>
<th>t_z</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-3.2</td>
<td>-1.6</td>
<td>4.5</td>
<td>130</td>
<td>126</td>
<td>228</td>
<td>131</td>
<td>114</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>2.8</td>
<td>1.3</td>
<td>-4.6</td>
<td>97</td>
<td>231</td>
<td>149</td>
<td>126</td>
<td>144</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>-8.7</td>
<td>2.9</td>
<td>3.1</td>
<td>143</td>
<td>261</td>
<td>12</td>
<td>83</td>
<td>29</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td>-2.2</td>
<td>-2.9</td>
<td>-3.6</td>
<td>27</td>
<td>280</td>
<td>141</td>
<td>312</td>
<td>216</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>5.8</td>
<td>4.1</td>
<td>4.9</td>
<td>19</td>
<td>25</td>
<td>341</td>
<td>18</td>
<td>75</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>0.3</td>
<td>-2.7</td>
<td>5.6</td>
<td>14</td>
<td>81</td>
<td>27</td>
<td>155</td>
<td>75</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>4.4</td>
<td>-0.3</td>
<td>-0.2</td>
<td>12</td>
<td>46</td>
<td>22</td>
<td>26</td>
<td>98</td>
<td>0.61</td>
<td></td>
</tr>
</tbody>
</table>

Crossover

<table>
<thead>
<tr>
<th></th>
<th>T_x</th>
<th>T_y</th>
<th>T_z</th>
<th>R_x</th>
<th>R_y</th>
<th>R_z</th>
<th>t_x</th>
<th>t_y</th>
<th>t_z</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>2.8</td>
<td>1.3</td>
<td>-4.6</td>
<td>97</td>
<td>231</td>
<td>149</td>
<td>126</td>
<td>144</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>-8.7</td>
<td>2.9</td>
<td>3.1</td>
<td>143</td>
<td>261</td>
<td>12</td>
<td>83</td>
<td>29</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>-8.7</td>
<td>1.3</td>
<td>-4.6</td>
<td>97</td>
<td>261</td>
<td>12</td>
<td>83</td>
<td>144</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mutation

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

Score new chromosomes

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Delete poor chromosomes

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample GOLD output

GMP into RNaseT1
Program **FRED** (OpenEye Scientific Software)

- **Docking is exhaustive**
  Unlike most docking programs, FRED does not use stochastic sampling to dock ligands. Rather, it begins with the set of all possible orientations (to a resolution of one Angstrom, by default) of each conformer near the receptor site and selects the docked position of the ligand from this set.
- **Speed**
  FRED typically docks from 7 to 15 conformers per second on a single PIII-800MHz CPU.
- **Multi-processor**
  FRED fully supports PVM (Parallel Virtual Machine) on Linux and SGI platforms. This allows FRED to take advantage of multiple processors on multiple machines while still returning a single centralized set of output.
- **Multiple scoring functions**
  FRED currently supports Chemscore, PLP, ScreenScore and Gaussian shape scoring. Scoring with ZAP (a Poisson-Boltzmann solver).
- **Alternative docking positions for ligands**
  FRED returns alternative docked poses for each ligand as well as the top scoring ligand.
- **Graphical receptor site preparation (with VIDA)**
  While FRED is fully functional as a command line program, the graphics program VIDA has a FRED wizard which can be used to set up the receptor site for Fred.

**Glide Fragment Library**
Set of 441 unique small fragments (1-7 ionization/tautomer variants; 6-37 atoms; MW range 32-226) derived from molecules in the medicinal chemistry literature. The set includes a total of 667 fragments with accessible low energy ionization and tautomeric states and metal and state penalties for each compound from Epik. These can be used for fragment docking, core hopping, lead optimization, de novo design, etc.

Program **GLIDE** (Grid-based Ligand Docking with Energetics)

- **Funnel:** site point search → diameter test → subset test → greedy score → refinement → grid-based energy optimization → GlideScore.
- **Approximates** a complete systematic search of the conformational, orientational, and positional space of the docked ligand.
- **Hierarchical filters**, including a rough scoring function that recognizes hydrophobic and polar contacts, dramatically narrow the search space.
- **Torsionally flexible energy optimization** on an OPLS-AA nonbonded potential grid for a few hundred surviving candidate poses.
- **The very best candidates** are further refined via a Monte Carlo sampling of pose conformation.
- **A modified ChemScore** (Eldridge et al. 1997) that combines empirical and force-field-based terms.
- **Validation:** 282 complexes, new ligand conformation, the top-ranked pose: 50%<1 Å, 33% >2 Å.
Program **eHITS** *(SimBioSys Inc.)*

- **Accurate**: validation test runs demonstrate that eHITS can reproduce X-ray structures with very high accuracy (low RMSD). Not limited to local energy minima dihedral angle samples.
- **Fast**: million compound libraries can be screened in special VHTS mode in a matter of hours on a Linux cluster. Exhaustive flexible ligand docking can also be performed on a single CPU under three minutes.
- **Easy to use, fully automated**: automatic pocket detection on the protein surface, automatic assignment of partial charges to atoms, consideration of alternative hydrogen protonation states, etc.
- **Customizable scoring function**: parameters and weights of all scoring components can be adjusted in a human readable, well documented configuration file.
- **Parallel execution**: built-in support for SMP (e.g. SGI Origin) and distributed (e.g. Linux clusters) architectures, and also grid computing.
- **Output postprocessing**: hierarchical file structure and clustering utility.

### Some important questions....

- Is there any relationship between **docking** and **ranking** accuracies?
- Will **docking/scoring combinations** provide better results in terms of hit rates? If so, which ones?
- Does "**consensus scoring**" from two or three independent scoring lists outperform single scoring?
- Will it be possible to find a **universal scoring function**?

Combined use of **3 docking algorithms** (Dock, FlexX, GOLD) with **7 scoring functions** (Dock, FlexX, GOLD, Pmf, Chemscore, Fresno, Score) for screening a 1000-compound library against two different protein targets, thymidine kinase (TK) and the ligand-binding domain of the estrogen receptor R subtype (ERR).

A specific database comprising **990 random** and **10 known ligands** was specifically created for each target.

Results of the virtual screening examined in terms of:

(i) docking accuracy (rmsd to known solutions),
(ii) scoring accuracy (prediction of the absolute binding free energy),
(iii) "consensus" versus single scoring,
(iv) discrimination of active from random compounds,
(v) hit rates and enrichment factors among the top scorers.


<table>
<thead>
<tr>
<th>ligand</th>
<th>Dock</th>
<th>FlexX</th>
<th>GOLD</th>
</tr>
</thead>
<tbody>
<tr>
<td>deoxythymidine</td>
<td>0.82</td>
<td>0.78</td>
<td>0.72</td>
</tr>
<tr>
<td>5-iododeoxuridine</td>
<td>9.33</td>
<td>1.03</td>
<td>0.77</td>
</tr>
<tr>
<td>5-iodouracil-anhydrohexitol</td>
<td>1.16</td>
<td>0.88</td>
<td>0.63</td>
</tr>
<tr>
<td>dbht (not publicly available)</td>
<td>2.02</td>
<td>3.65</td>
<td>0.93</td>
</tr>
<tr>
<td>6-(3-hydroxy-propyl-thymine)</td>
<td>1.02</td>
<td>4.18</td>
<td>0.49</td>
</tr>
<tr>
<td>6-(6-hydroxymethyl-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl)-5-methyl-1H-pyrimidin-2,4-dione</td>
<td>9.62</td>
<td>13.30</td>
<td>2.33</td>
</tr>
<tr>
<td>(North)-methanocarbathymidine</td>
<td>7.56</td>
<td>1.11</td>
<td>1.19</td>
</tr>
<tr>
<td>aciclovir</td>
<td>3.08</td>
<td>2.71</td>
<td>2.74</td>
</tr>
<tr>
<td>ganciclovir</td>
<td>3.01</td>
<td>6.07</td>
<td>3.11</td>
</tr>
<tr>
<td>penciclovir</td>
<td>4.10</td>
<td>5.96</td>
<td>3.01</td>
</tr>
</tbody>
</table>

Only one set of protein (TK) coordinates used: PDB code 1KIM
**Misdocked complexes** can be categorized as soft and hard failures.

**Soft failures**: the search algorithm is unable to locate the global energy minimum corresponding to the crystal structure but this conformation, after minimization with the force field chosen, yields a lower energy than that of the lowest energy found in the docking simulations.

**Hard failures**: the global energy minimum corresponds to a misdocked structure, i.e. the method is unable to reproduce the differences in relative energies of alternate binding modes.

Comparison of the three docking methods each with its best performing scoring function (TK ligands)

Ranking versus rms deviations from X-ray pose for TK ligands screened with the three best dock/scoring combinations

C. Bissantz, G. Folkers & D. Rognan

Only partial discrimination of true hits from random ligands

- **Dock**: 10 true hits, 774 random
- **FlexX**: 10 true hits, 488 random ligands
- **Gold**: 10 true hits, 927 random ligands

Docking methods each with its best performing scoring function (TK ligands)

- **Dock**: 10 true hits, 774 random
- **FlexX**: 10 true hits, 488 random ligands
- **Gold**: 10 true hits, 927 random ligands

% = percentages of the total number of ligands for which a docking solution was found

Ranking (position in the scoring list)

- 3 independent docking poses

- **Dock**: 10 true hits, 774 random
- **FlexX**: 10 true hits, 488 random ligands
- **Gold**: 10 true hits, 927 random ligands
raloxifen 4-hydroxy-tamoxifen

reference protein coordinates: PDB code = 3ERT

Dock/Dock (10.47 Å, 17th)
FlexX/FlexX (2.24 Å, 2nd)
Gold/Gold (1.26 Å, 25th)

Dock/Dock (0.68 Å, 18th)
FlexX/FlexX (1.47 Å, 11th)
Gold/Gold (0.76 Å, 14th)

% = percentages of the total number of ligands for which a docking solution was found

Dock: 10 true hits, 907 random
FlexX: 9 true hits, 876 random ligands
Gold: 10 true hits, 926 random ligands

Only partial discrimination of true hits from random ligands

Note that docking/scoring combinations are different from those found optimal for TK inhibitors

Enrichment of inhibitors for seven targets calculated with FlexX and four scoring functions

Martin Stahl & Matthias Rarey

Consensus scoring: Comparison of the FlexX and PLP scoring functions with the FlexX-PLP combination ScreenScore.

For each target, the left column of the triplet shows FlexX results, the middle column PLP results, and the right column results calculated with ScreenScore.

Consensus scoring: Only those compounds are regarded that receive high ranks with two or more scoring functions.

Considerable reduction of false positives ("enriched hit-rate")


17 pairs of complexes of the same protein bound to 2 related ligands

Molecular mechanics (AMBER) and statistical potentials (PMF)

Exhaustive enumeration of all possible docking solutions

Reconstruction of the shape of the energy landscape (coverage-error plots)

Calculation of physico-chemical descriptors

Quantitative evaluation of success

Linear discriminant analysis

Physical origin of failures/successes

Desolvation effects

Dispersive interactions

Directional effects of hydrogen bonds

C. Pérez & A. R. Ortiz


MOLDOCK – an extension of the piecewise linear potential (PLP)

http://www.molegro.com/products.html

Superoxide dismutase
RAS proteins
Cyclooxygenase (COX-2)
Fibroblast Growth Factor Receptor
RAF
Protein-Tyrosine-Phosphatase 1B

Vascular Endothelial Growth Factor
Insulin Tyrosine Kinase
c-ABL Tyrosine Kinase
CDK-2
Farnesyltransferase
VEGF1

Prof. W. Graham Richards

Gi-Rodondo R, Estrada J, Morraale A, Herranz F, Sancho J, Ortiz AR.
VSDMIP: virtual screening data management on an integrated platform.
ALFA (Automatic Ligand Flexibility Assignment)

- Input Molecule (PDB or Mol2 format)
- Select and Classify Rotatable Bonds
- Check Number of Possible Conformers
- Refinement of Selected Conformations

Main Algorithm


PyMOL

QUESTIONS WELCOME
E-mail: federico.gago@uah.es
http://www3.uah.es/farmamol/