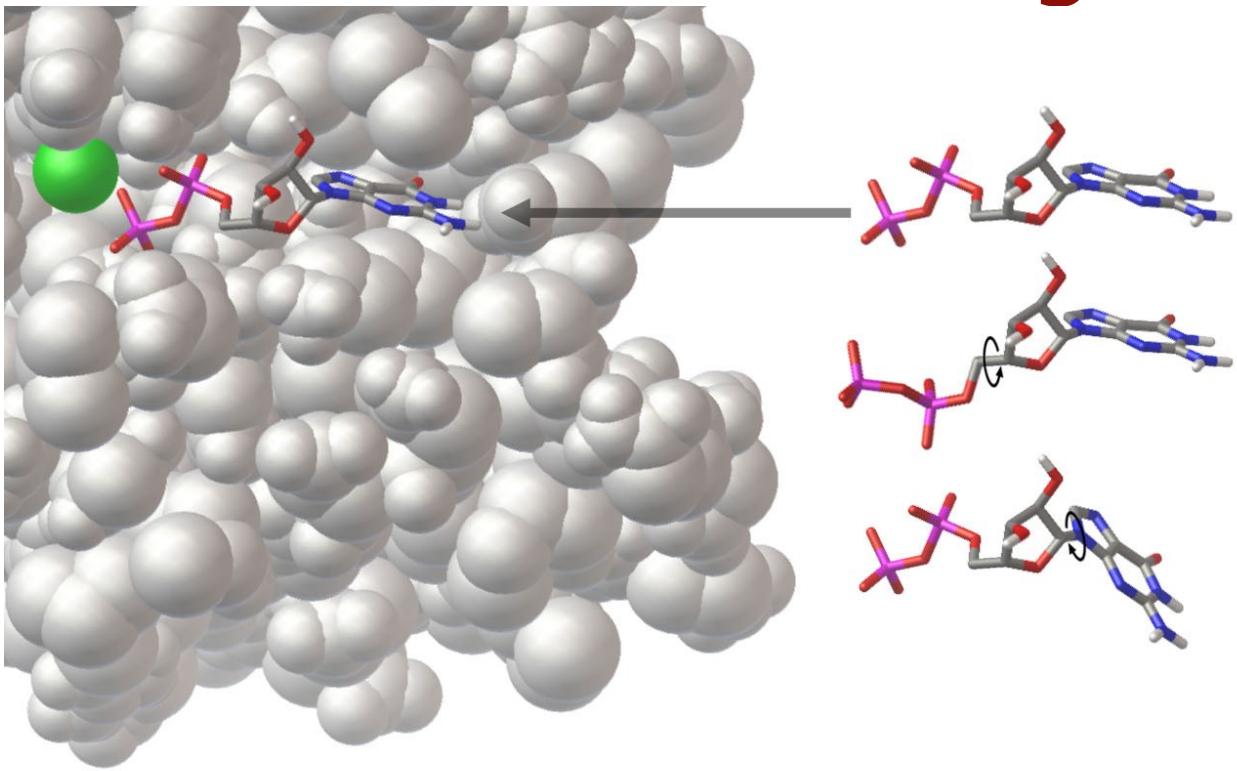
Using AutoDock 4 and AutoDock Vina with AutoDockTools

Ruth Huey
Sargis Dallakyan
Stefano Forli
David S. Goodsell

What is Docking?



Given the 3D structures of two molecules, determine the best binding modes.

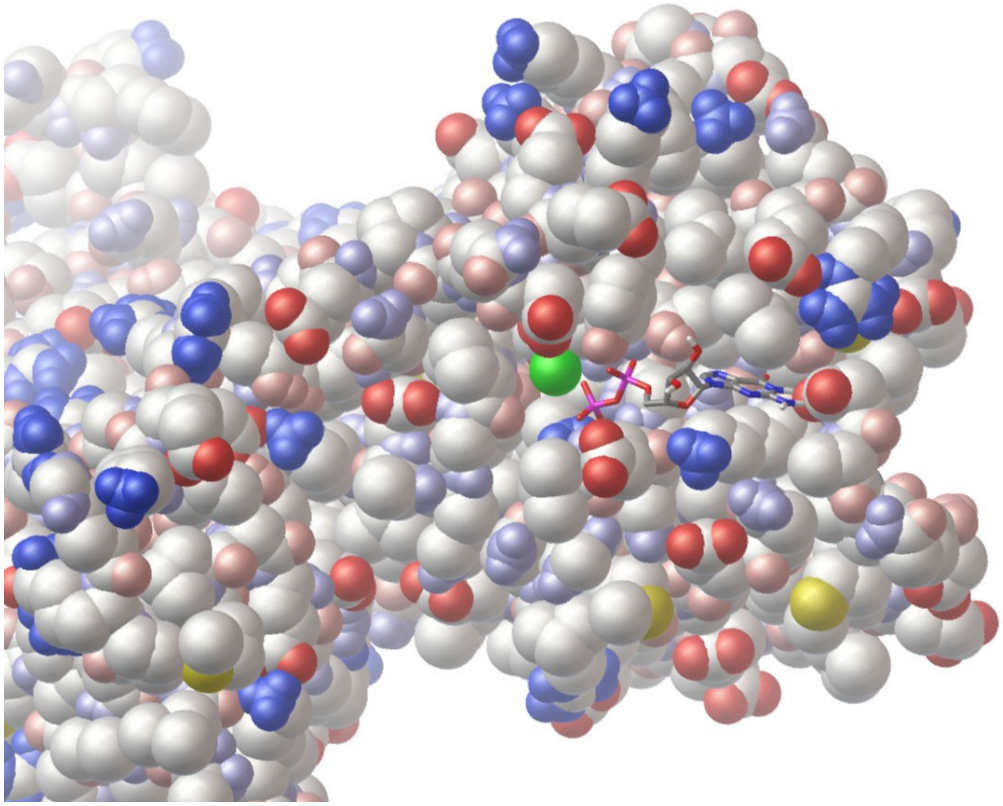
Key aspects of docking...

- Scoring Functions
  - *Predicting the energy of a particular pose*
  - *Often a trade-off between speed and accuracy*
- Search Methods
  - *Finding an optimal pose*
  - *Which search method should I use?*
- Dimensionality
  - *Can we trust the answer?*

AutoDock History

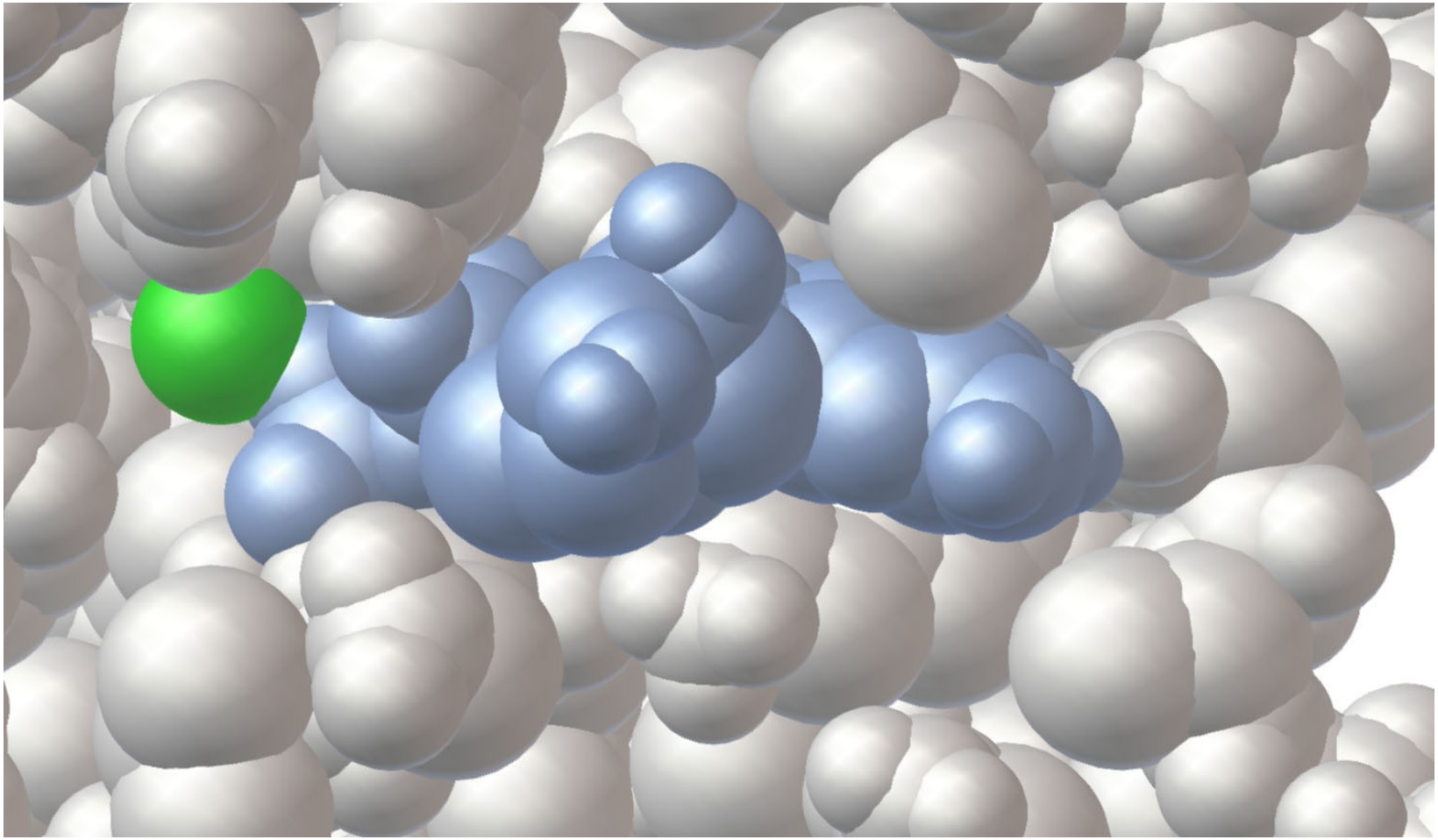
- 1990 - AutoDock 1
  - *First docking method with flexible ligands*
- 1998 - AutoDock 3
  - *Free energy force field and advanced search methods*
  - *AutoDockTools Graphical User Interface*
- 2009 - AutoDock 4
  - *Current version of AutoDock*
  - *Many parameters available to user*
- 2009 - AutoDock Vina
  - *Rewritten by Oleg Trott, new approach to scoring and search*
  - *One step solution to docking*

Scoring Functions



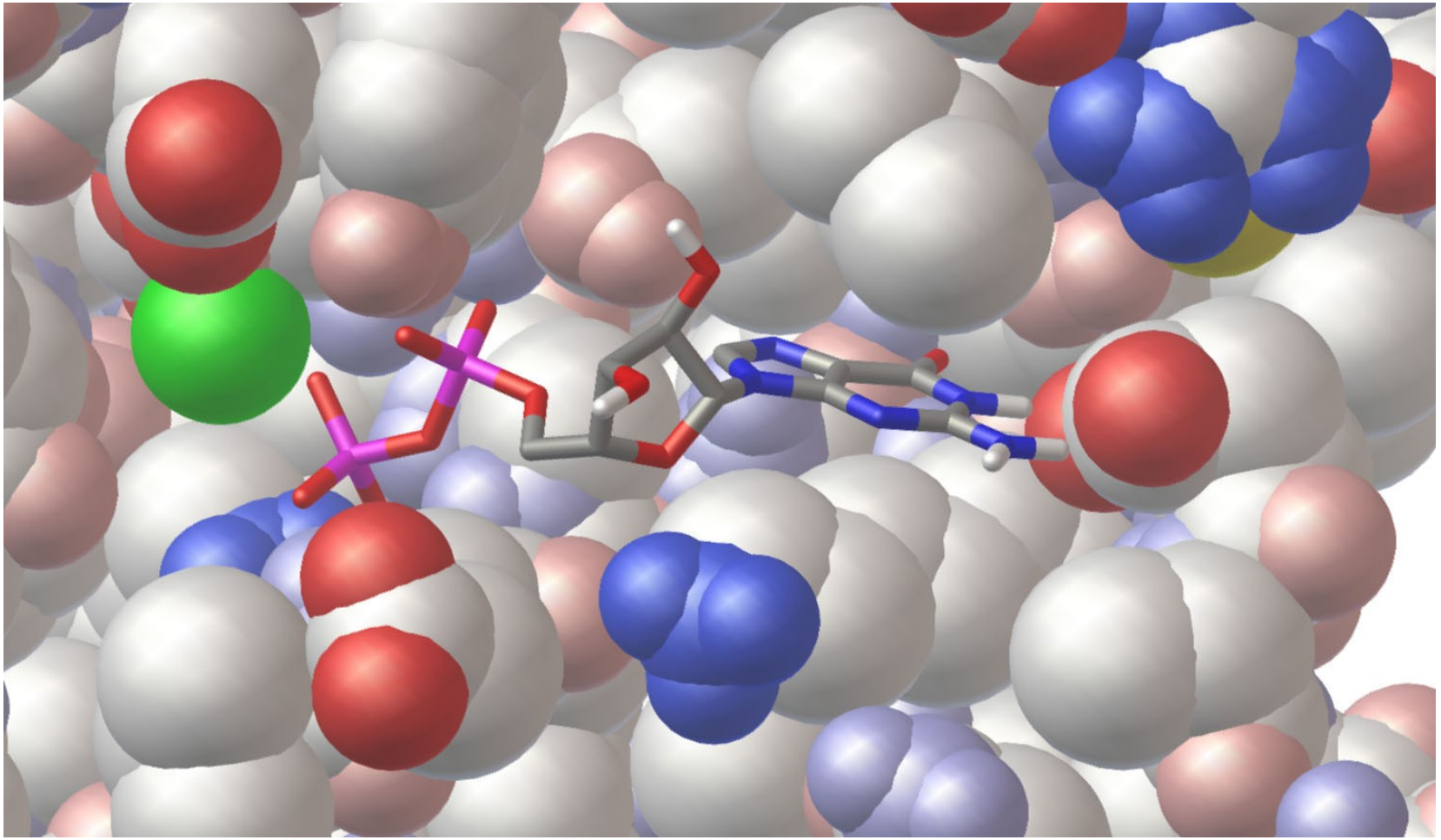
$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

Dispersion/Repulsion



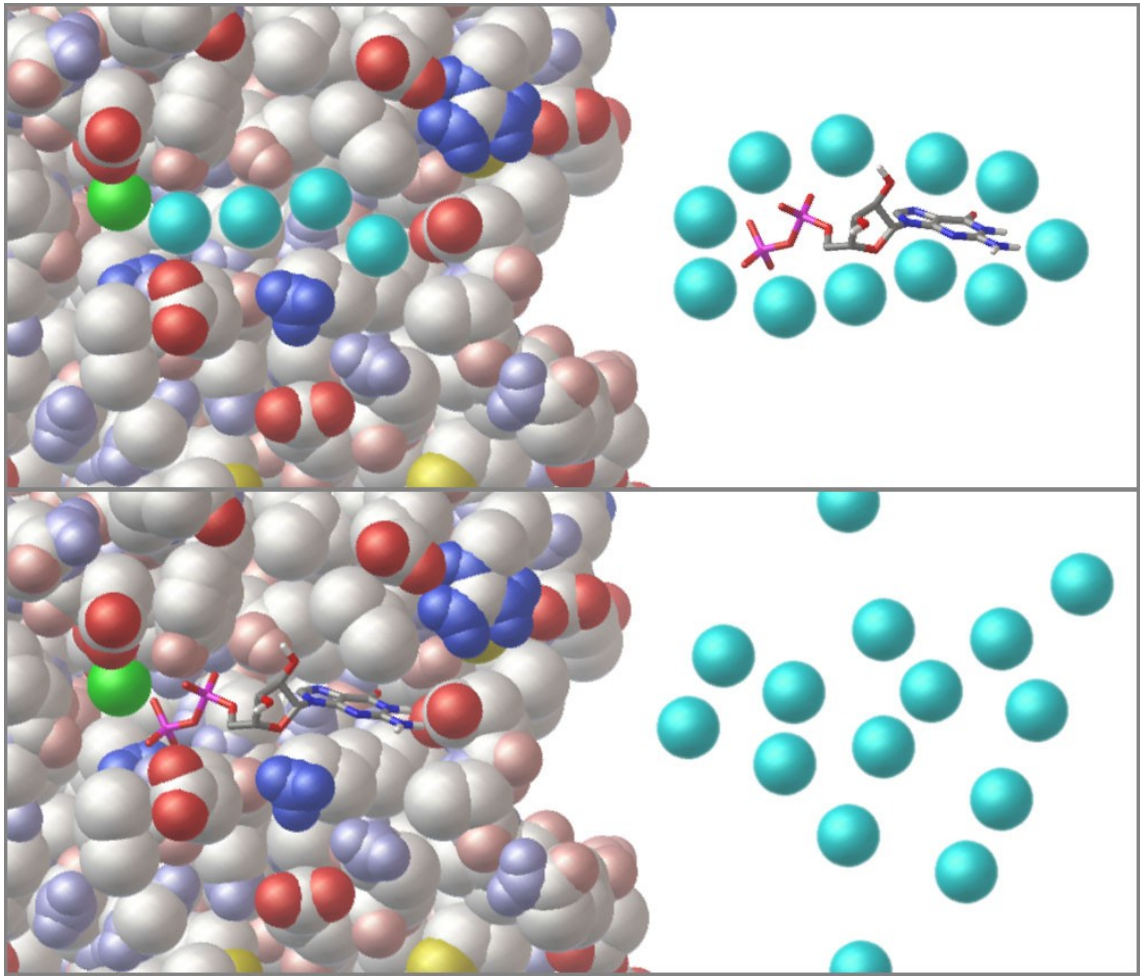
$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

Electrostatics and Hydrogen Bonds



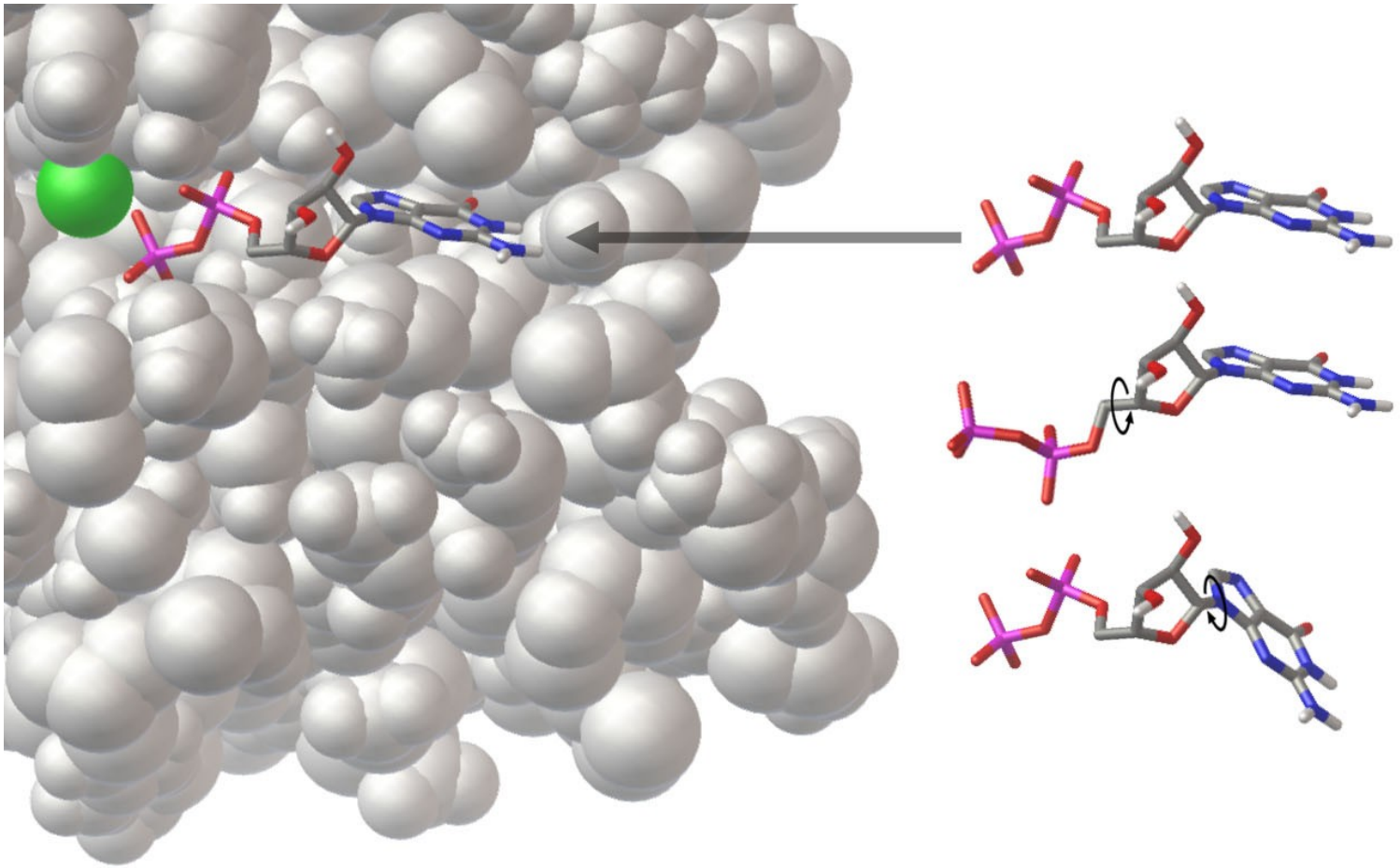
$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

Desolvation



$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

Torsional Entropy



$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

AutoDock Empirical Free Energy Force Field

$$W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right)+$$

$$W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}}-\frac{D_{ij}}{r_{ij}^{10}}\right)+$$

$$W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}}+$$

$$W_{sol}\sum_{i,j}(S_i V_j + S_j V_i)e^{(-r_{ij}^2/2\sigma^2)}+$$

$$W_{tor}N_{tor}$$

- Physics-based approach from molecular mechanics
- Calibrated with 188 complexes from LPDB, K_i's from PDB-Bind
- Standard error = 2.52 kcal/mol

AutoDock Vina Scoring Function

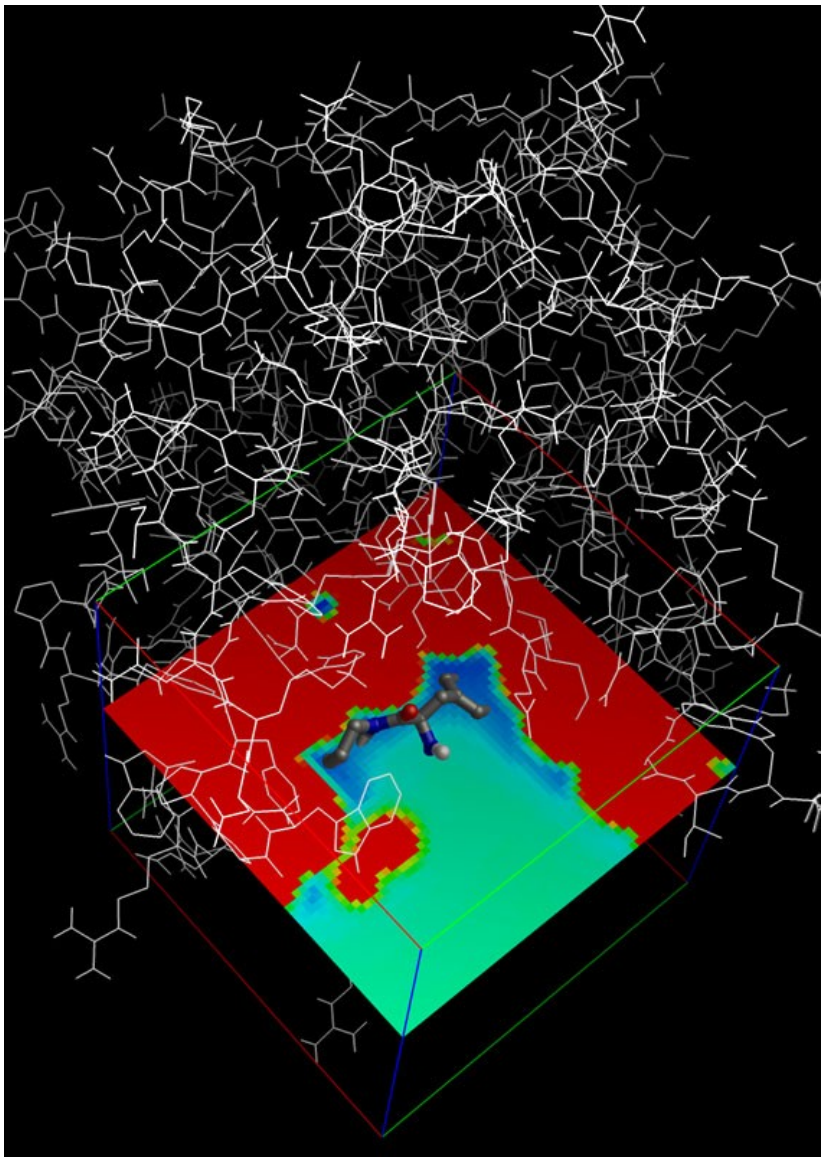
Combination of knowledge-based and empirical approach

$$\Delta G_{binding} = \Delta G_{gauss} + \Delta G_{repulsion} + \Delta G_{hbond} + \Delta G_{hydrophobic} + \Delta G_{tors}$$

- ΔG_{gauss}
  Attractive term for dispersion, two gaussian functions
- $\Delta G_{repulsion}$
  Square of the distance if closer than a threshold value
- ΔG_{hbond}
  Ramp function - also used for interactions with metal ions
- $\Delta G_{hydrophobic}$
  Ramp function
- ΔG_{tors}
  Proportional to the number of rotatable bonds

- Calibrated with 1,300 complexes from PDB-Bind
- Standard error = 2.85 kcal/mol

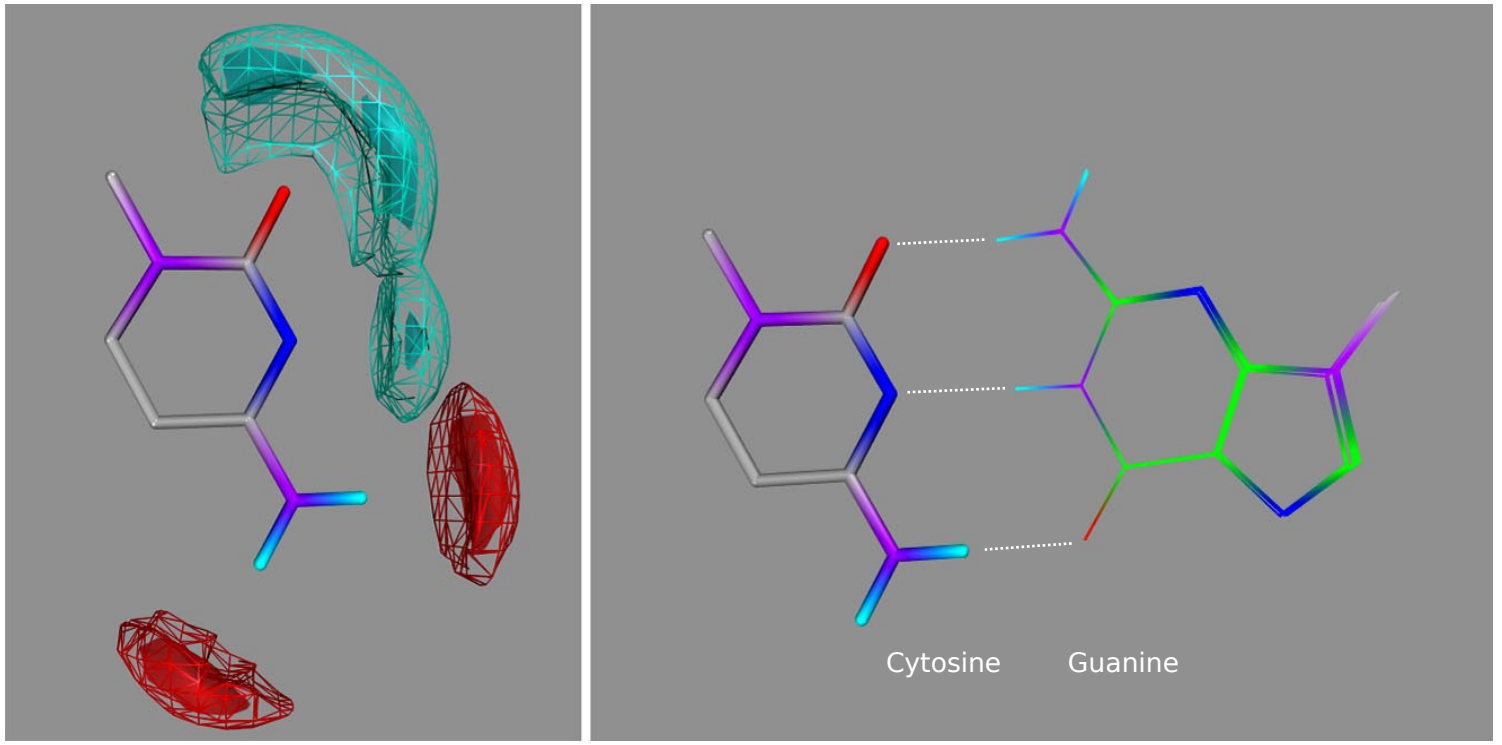
http://vina.scripps.edu



Grid Maps

- Precompute interactions for each type of atom
- 100X faster than pairwise methods
- Drawbacks: receptor is conformationally rigid, limits the search space

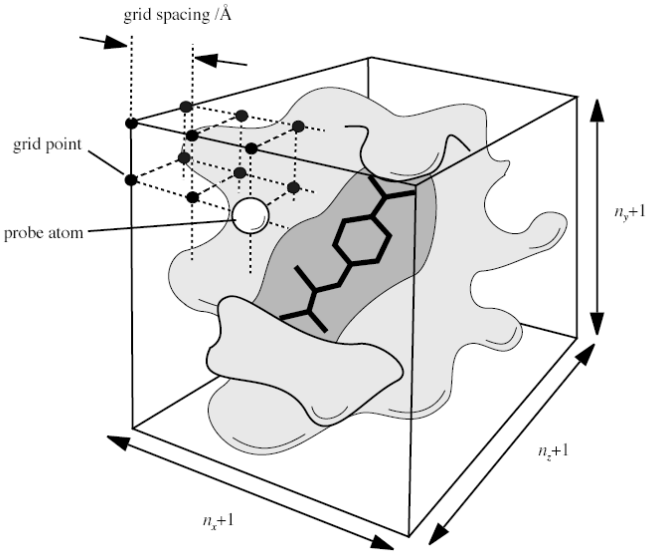
H-bond Grid Map



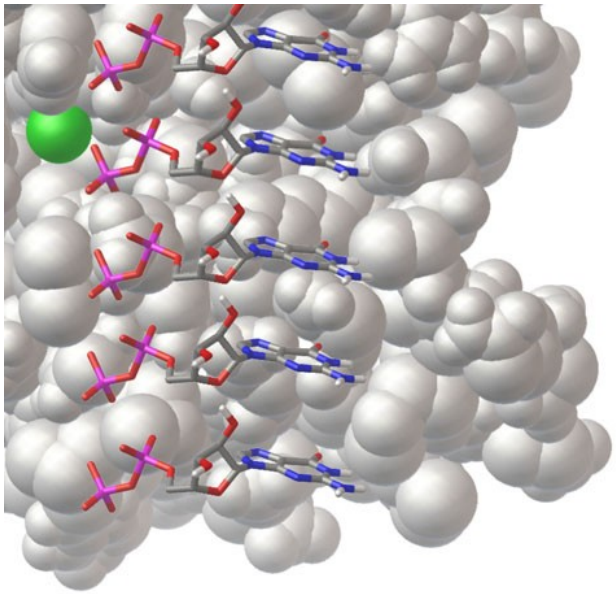
Huey, Goodsell, Morris, and Olson (2004) *Letts. Drug Des. & Disc.*, **1**: 178-183

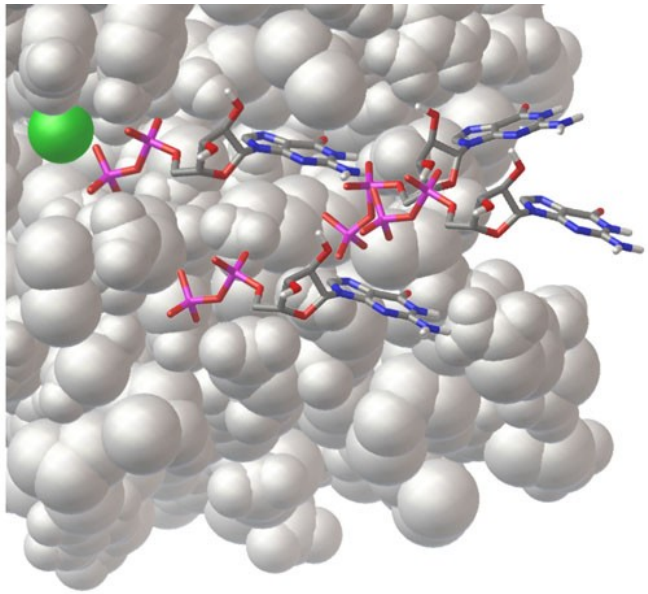
Setting up the AutoGrid Box

- Center:
  - center of ligand;
  - center of macromolecule;
  - a picked atom; **or**
  - typed-in x-, y- and z-coordinate
- Grid point spacing:
  - default is 0.375Å (from 0.2Å to
- Number of grid points in each dime
  - from $2 \times 2 \times 2$ to $126 \times 126 \times$
- Make sure all the flexible parts of the macromolecule are inside the grid
- Make sure that the entire binding site is inside the grid



Two Kinds of Search





Systematic

- Exhaustive, deterministic
- Outcome is dependent on granularity of sampling
- Feasible only for low-dimensional problems

Stochastic

- Random, outcome varies
- Must repeat the search or perform more steps to improve chances of success
- Feasible for larger problems

AutoDock and Vina Search Methods

- Global search algorithms:
  - Simulated Annealing (Goodsell *et al.* 1990)
  - Genetic Algorithm (Morris *et al.* 1998)
- Local search algorithm:
  - Solis & Wets (Morris *et al.* 1998)
- Hybrid global-local search algorithm:
  - Lamarckian GA (Morris *et al.* 1998)
- Iterated Local Search:
  - Genetic Algorithm with Local Gradient Optimization (Trott and Olson 2010)

Dimensionality of Molecular Docking

Degrees of Freedom include:

- **Position** / Translation (3)
  - *x*,*y*,*z*
- **Orientation** / Quaternion (3)
  - *qx*, *qy*, *qz*, *qw (normalized in 4D)*
- **Rotatable Bonds** / Torsions (*n*)
  - τ_1, τ_2, … τ_n

Dimensionality, $D = 3 + 3 + n$

Sampling Hyperspace

- Say we are hunting in D-dimensional hyperspace…
- We want to evaluate each of the D dimensions N times.
- The number of "evals" needed, n, is: $n = N^D$

  $\therefore N = n^{1/D}$
- For example, if $n = 10^6$ and…
  - D=6, $N = (10^6)^{1/6} = 10$ evaluations per dimension
  - D=20, $N = (10^6)^{1/20} = \sim 2$ evaluations per dimension
- Clearly, the more dimensions, the tougher it gets.

Practical Considerations

- What problems are feasible?
  - Depends on the search method:
    - Vina > LGA > GA >> SA >> LS
    - AutoDock SA : can output trajectories, $D < 8$ torsions.
    - AutoDock LGA : $D <$ 8-16 torsions.
    - Vina : good for 20-30 torsions.
- When are AutoDock and Vina not suitable?
  - Modeled structure of poor quality;
  - Too many torsions (32 max);
  - Target protein too flexible.
- Redocking studies are used to validate the method

Using AutoDock: Step-by-Step

Prepare the Input Files

- Ligand PDBQT file
- Rigid Macromolecule PDBQT file
- (Flexible Macromolecule PDBQT file)
- AutoGrid Parameter File (GPF) and AutoDock Parameter File (DPF)
- Or AutoDockVina Parameter File

Run AutoGrid 4

Run AutoDock 4

Run ADT to Analyze DLG

Run AutoDock Vina

Run ADT to Analyze results

PDBQT Format

Coordinates from the Protein Data Bank with:

- Polar hydrogen atoms
- Atomic partial charges
- Types (aromatic/aliphatic carbon, hydrogen bond donors/acceptors)
- Center of rotation and rotatable bonds identified