

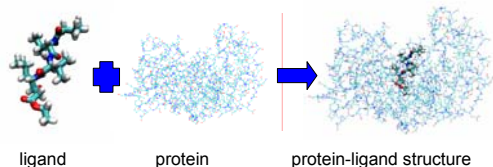
Study of a Highly Accurate and Fast Protein-Ligand Docking Method based on Molecular Dynamics for Desktop Grids

M. Taufer, M. Crowley, D. Price, A. A. Chien, C. L. Brooks III

The Scripps Research Institute (TSRI) – The Center for Theoretical Biological Physics (CTBP) - UCSD Jacobs School of Engineering (CSE/UCSD)

Protein-Ligand Docking

Protein-ligand docking → computational methods for the prediction of ligand-protein structural information



- Most well-known docking methods use **fast, simplified scoring functions** to direct conformational search and select best docked structures
- Significant **inaccuracy** associated with these methods
- Methods based on **molecular dynamics (MD)** and **atomically detailed force fields** (e.g. CDOCKER) are more accurate but even more **time- and computing-expensive**

Maintain **accuracy** and improve **performance** by taking advantage of new computer technologies and new distributed architectures → **desktop grids**

Desktop Grid Computing

- Desktop grids → desktop PCs connected to the Internet and Intranets
- Their number is large and still growing
- 80%-90% of their CPU time is idle time
- Advantage of commodity-priced hardware and open-source software offers good price/performance ratio



Desktop grids are highly **heterogeneous, volatile** computational environments → well-suited desktop grid docking methods require **fine-grained parallelism**

MD Docking versus other Docking Methods

Other Methods

- AutoDock
- DOCK
- FlexX
- ICM
- GOLD

Comparison Metrics

- Docking accuracy (DA)
- RMSD of lowest energy ligands
- Computing time

Definition of docking attempt

- **Our docking method (T10)**: 10 independent trials per attempt - for each ligand orientation in a trial, we run 7000 MD steps (2000 heating steps/5000 cooling steps)
- **Other methods**: controlled by the default or recommended parameter settings of each algorithm

Searching for a MD-based method with high parallelism granularity

Our objective:

Benefit from molecular mechanics force fields and guarantee performance using cost-effective platforms

Our approach:

Docking method based on molecular dynamics simulations with a highly flexible computational granularity

Our final goal:

Screening large set of ligands (10,000 molecules) using our method and open-source desktop grids

Grid-based Representation

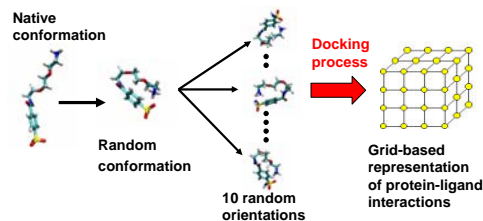
- Rigid protein structures and flexible ligands
- Grid-based representation of protein interactions with ligands
- Preliminary computation of grid maps → 3D lattice of regular spaced points
- Each point stores the potential energy between a "probe" atom of the ligand due to its interaction with the macromolecule

Reduce the floating-point computation by **several orders of magnitude**

Definition of Attempt and Trials

Docking attempt → sequence of independent trials

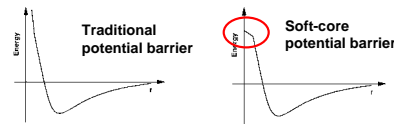
Example of trial for the 1cnx complex:



Predicted ligand → docked ligand with lowest energy over the several attempts

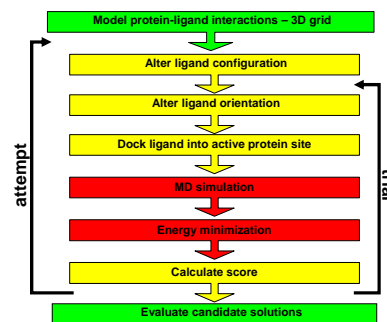
Soft-Core Repulsion

- Van der Waals and electrostatic potentials → soft-core repulsions
- Soft-core repulsion → potential barrier at vanishing interatomic distances as a finite limit



Facilitate the penetration of ligands into proteins

MD-based Docking Algorithm



CHARMM -Chemistry at HARvard Macromolecular Mechanics

- We use CHARMM to investigate the protein-ligand docking process.
- CHARMM uses classical mechanical methods for **molecular dynamics** simulations (MD):

$$\delta^2 s_i / \delta t^2 = F_i / m_i$$

displacement s_i of atom i at time t with force F_i and mass m_i

- The force on the atoms is the negative gradient of the **CHARMM potential energy function**

Protein-Ligand Complexes

- **Trypsin** → 3ptb(3), 1tng(2), 1tnj(3), 1tnk(4), 1tni(5), 1tpp(7), 1pph(11)
- **Cytochrome P-450cam** → 1phf(1), 1phg(5), 2cpp(3)
- **Neuraminidase** → 1nsc(12), 1nsd(11), 1nnb(11)
- **Carboxypeptidase** → 1cbx(5), 3cpa(8), 6cpa(16)
- **L-Arabinose binding protein** → 1abe(4), 1abf(5), 5abp(6)
- **e-Thrombin** → 1etr(15), 1ets(13), 1ett(11)
- **Thermolysin** → 3tmn(10), 5tn(14), 6tmn(20)
- **Penicillopepsin** → 1apt(30), 1aup(29)
- **Intestinal FABP** → 2ifb(15)
- **Carbonic Anhydrase II** → 1cil(6), 1okl(5), 1cnx(13)

Metrics

Docking Accuracy (DA):

$$DA = f_{RMSD \leq 2} + 0.5(f_{RMSD \leq 3} - f_{RMSD \leq 2})$$

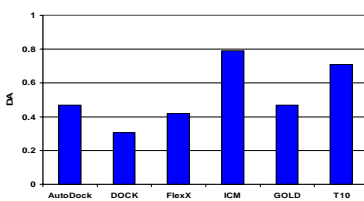
- $f_{RMSD \leq a}$ fraction of predicted ligands docked into a given protein with RMSD lesser or equal to a Å

Simulation time:

- CPU time needed to complete a single trial and docking attempt

Docking Accuracy

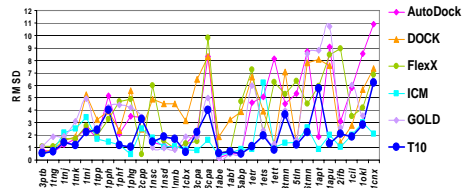
Avg. DA for the 31 protein-ligand complexes:



Our method T10 provides better DA, except for ICM

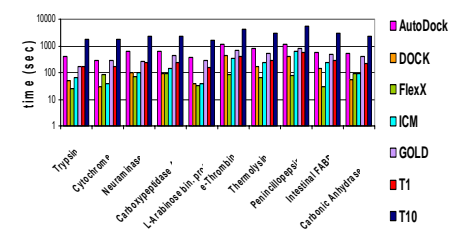
Best RMSD

RMSD of the **predicted ligand** from the corresponding ligand in its published crystal structures



Our method T10 provides an RMSD competitive with ICM

Computing Time



If **enough processors** are available, **time-to-solution** of a docking attempt is **equal to time-to-solution** of a trial (T1)