

DAPLDS: a Dynamically Adaptive Protein-Ligand Docking System based on Multi-Scale Modeling

M. Taufer, P.J. Teller, M. Ceberio, D.P. Anderson, R. Armen, C.L. Brooks III

University of Texas at El Paso – University of California at Berkeley – The Scripps Research Institute

Protein-Ligand Docking

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

DAPLDS deploys molecular dynamics (MD) and a force field based scoring energy function (like that in CHARMM) to rank docked ligands with respect to their binding affinity within a multi-scale modeled space of protein-ligand complexes.

Docking Model based on MD

Example of docking trial:

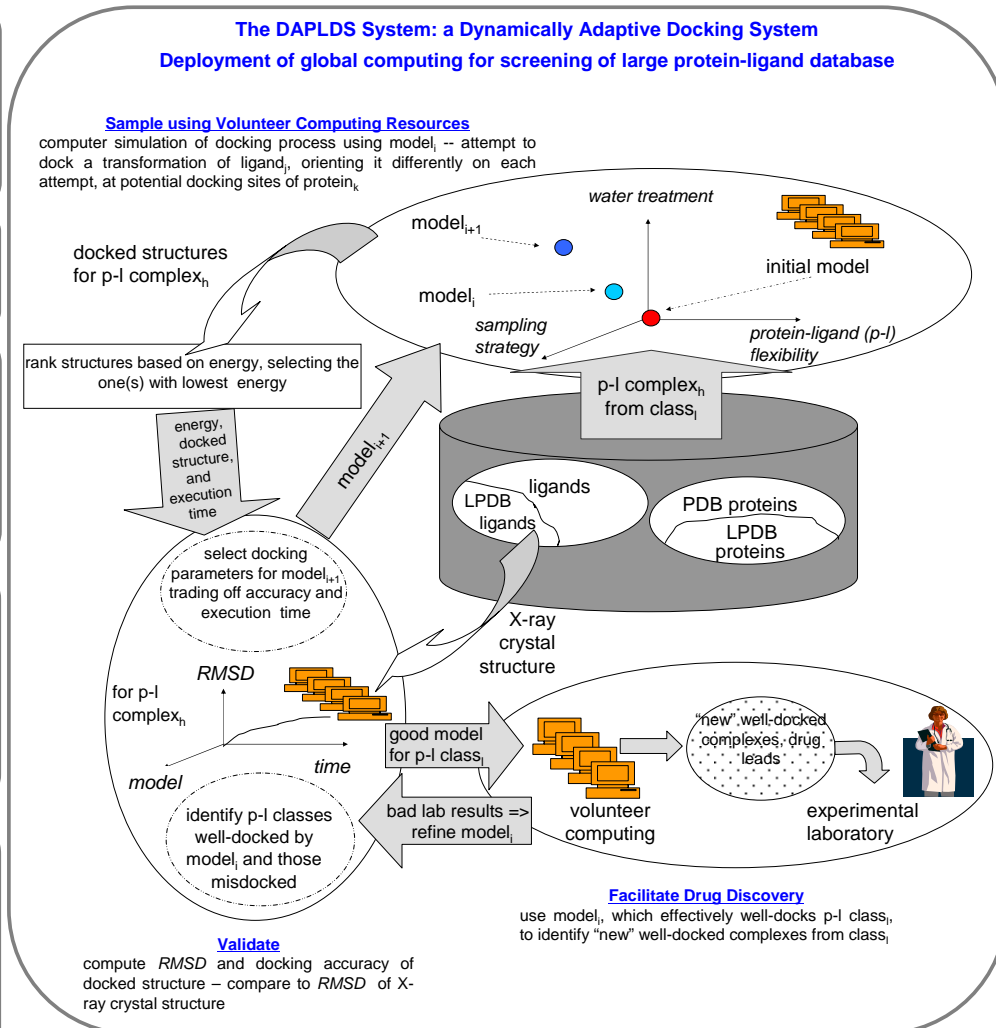
Spanning Scales of Docking

- Protein-Ligand representation:**
 - Spanning scale from rigid to flexible representation of protein-ligand interactions
- Solvent representation:**
 - Spanning scale from less accurate to more accurate modeling of water treatment
- Sampling strategy:**
 - Spanning scale from fixed to adaptive sampling of the protein-ligand docking space

Preliminary Results

p-l complex	10 trials (RMSD)	20 trials (RMSD)	Adaptive trials (RMSD)
1pgh	4.00	0.53	0.43
2cpp	3.27	3.27	0.40
6cpa	4.00	4.00	2.90
6lmm	2.21	2.21	1.99
1apt	5.72	4.79	3.30

- Adaptive docking trials → developed upon inspection of cases in the 10 and 20 trials by improving protein representation and increasing number of trials per complex
- Quality of docking → RMSD (root mean square deviation) lowest energy p-l structure vs. experimental result



The DAPLDS Project: Objectives

- Explore the nature of dynamic protocol model adaptations for protein-ligand docking
- Develop methods and models that efficiently accommodate these computational adaptations in global computing deploying resources owned by the public (volunteers)
- Extend knowledge with respect to protein-ligand complexes and make this knowledge accessible to the scientific community via an easy-to-access data repository
- Provide opportunities for faculty and students at the three collaborating institutions to become involved in this cross-disciplinary research

DAPLDS deploys global computing based on volunteer resources and flexible constraints to efficiently explore the multi-scale modeled space of protein-ligand complexes.

Global Computing

- Deployment of PCs connected to the Internet and owned by general public → volunteer computing
- Emerged in mid-1990s with projects such as GIMPS and SETI@home
- Number of PCs predicted in 2015: 1 billion → many PetaFLOPs of computing power

BOINC

Berkeley Open Infrastructure for Network Computing

- BOINC is the global computing system that provides the distributed framework for DAPLDS

BOINC adaptations in DAPLDS:

- Dynamic scheduling:** work generated based on previous results (docked structures) and resource availability
- Data management:** collection or discharge of docked structures based on their energy
- Data integrity:** redundant instances of tasks distributed among homogenous computing machines (homogenous redundancy)

Flexible Constraints

- Control task scheduling via flexible constraints
 - Use constraint solving techniques
 - But volunteer computing is not stable -- cannot enforce strict constraints
- Find acceptable dockings via flexible processes
 - Selection criteria are available (e.g., minimization of energy)
 - Use constraints to reverse-engineer these expectations/models
 - But model knowledge may be insufficient and/or models may change
- Any-time, speculative algorithms via optimization
 - Ensure consistency/quality of results
 - Deal with delays in result replies