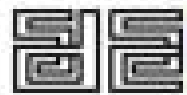


Study of a Highly Accurate and Fast Protein Ligand Docking Algorithm Based Upon Molecular Dynamics

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



SDSC



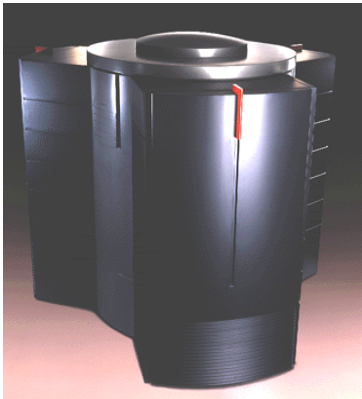
Search for More Cost-Effective Computing Resources

Late 80s

Middle 90s

Early 2000

time →



Cray C90



Cluster of PCs

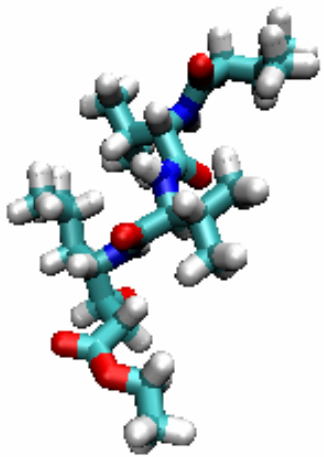


Desktop grid

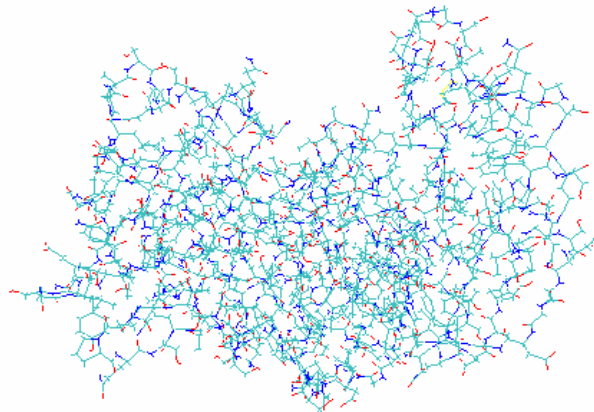
Desktop grids → By scavenging for available and idle cycles, they provide computing power at a significant cost savings

Protein-Ligand Docking

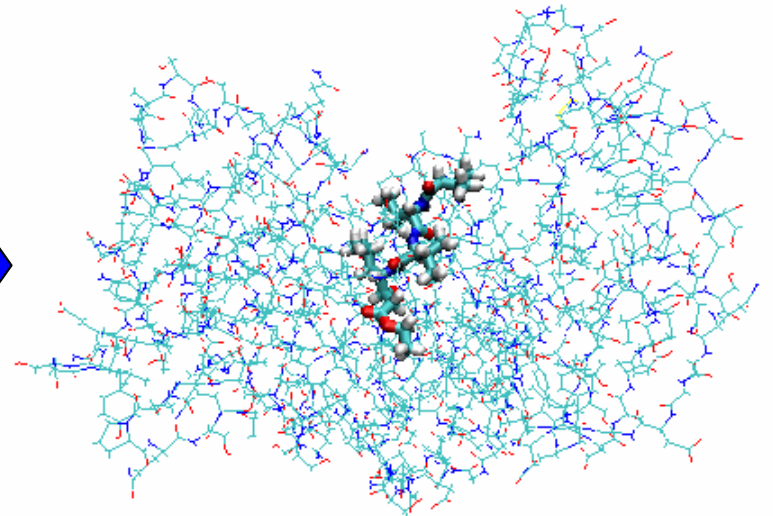
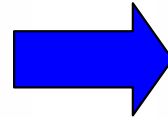
Drug development → use of small molecules (ligands) to turn on or off a protein function



ligand



protein



protein-ligand structure

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

Existing Docking Methods

- Most docking methods are **fast and use simplified scoring functions** to direct conformational search and select best docked structures
- Significant **inaccuracy** associated with these methods [Ferrara *et al.*, J.Med.Chem., 2004]
- Methods based on **molecular dynamics (MD)** and atomically detailed force fields (e.g., CDOCKER) are more accurate but **time- and resource-expensive**

On desktop grids, **too long turn-around time MDs are likely to fail** and their results are not guaranteed

Our MD-based Docking Method

Our goals

- Assure accuracy → benefit from the **molecular mechanics force fields**
- Guarantee performance → return **docking results** in a **short turnaround time** using **cost-effective platforms**

Our approach

- Docking method based on **CHARMM molecular dynamics** simulations and with a **highly flexible computational granularity**

Outline

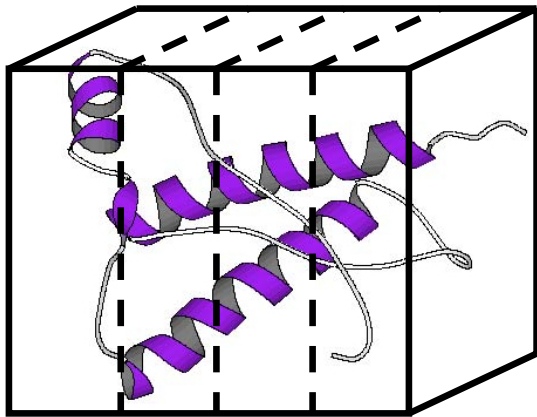
- The scientific computing code CHARMM
- MD-based docking algorithm:
 - Dealing with large molecular systems
 - Decoupling tightly couple computation
- Characterization our docking method:
 - Accuracy and length of MDs
 - Number and length of docking attempts
- Our docking method vs. other docking methods
- Conclusion and future research opportunities

CHARMM – Chemistry at Harvard Macromolecular Mechanics

- CHARMM is a computational chemistry code for studying the **structure** and the **dynamics** of relevant **macromolecules** such as proteins, DNA, RNA
- CHARMM uses classical mechanical methods for **molecular dynamics simulations** (MD)
- CHARMM MD simulations are used to investigate **thermodynamics, kinetics** of macromolecules

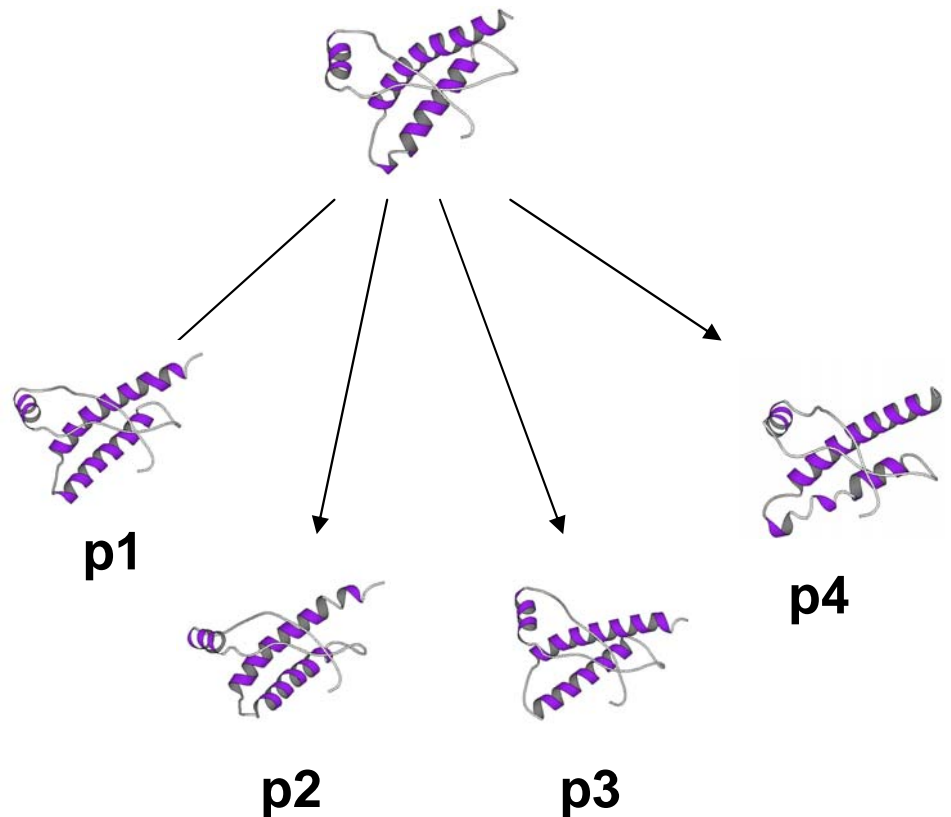
Parallelism in CHARMM

data parallelism



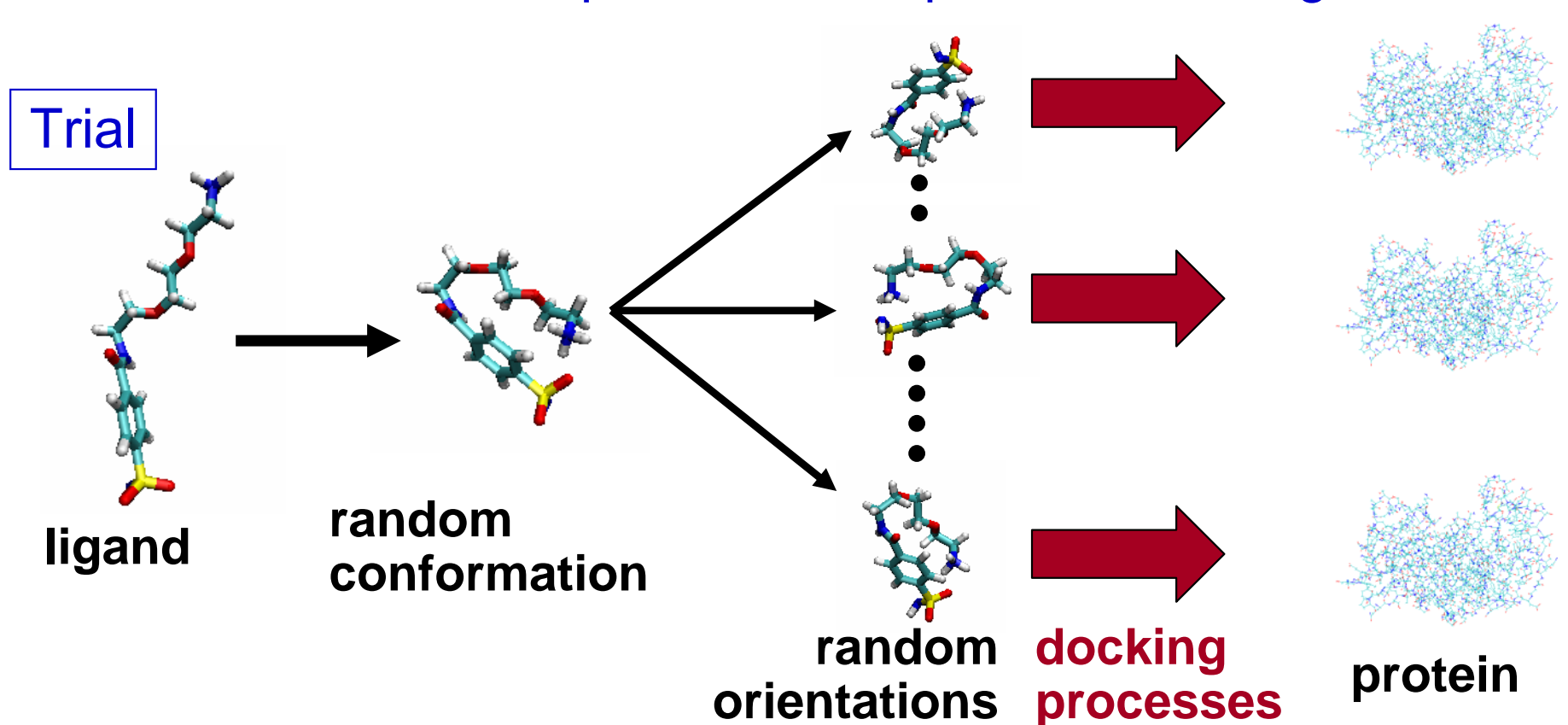
Nodes: p1 p2 p3 p4

task parallelism



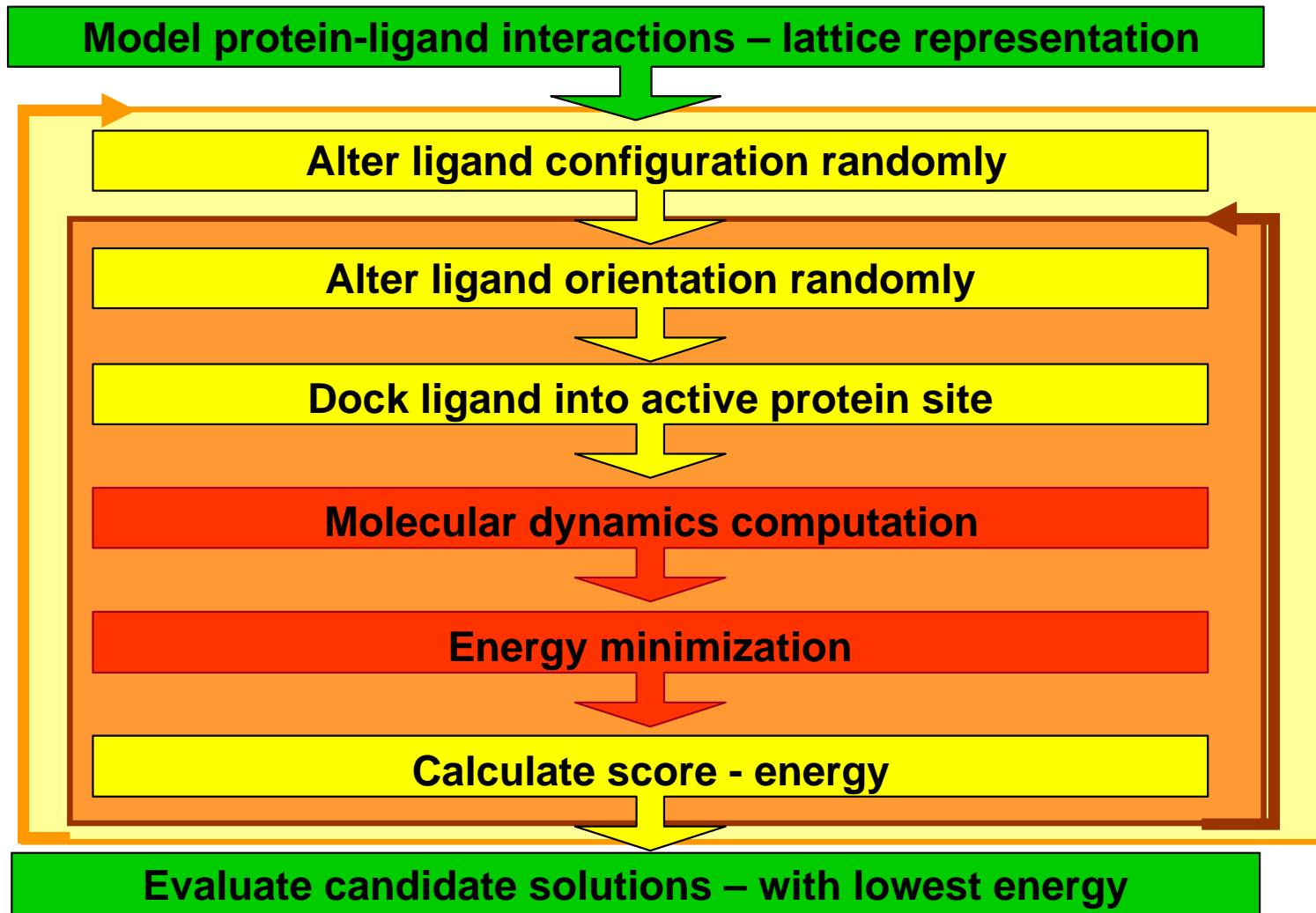
Decoupling the Docking Process

Docking attempt → for each protein-ligand complex, we run a sequence of independent docking trials



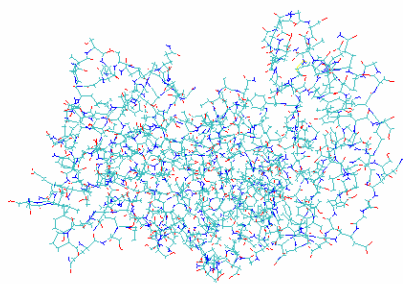
Predicted ligand → docked ligand in protein with lowest energy

MD-based Docking Algorithm

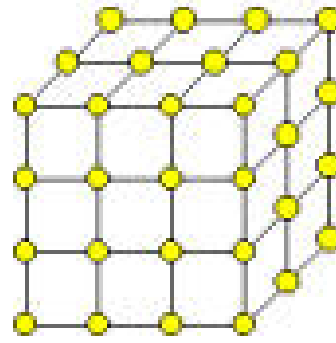


Modeling Protein-Ligand Interactions (I)

- Rigid protein structures and flexible ligands
- Representation of protein interacting with ligands:
3D lattice of regular spaced points



protein



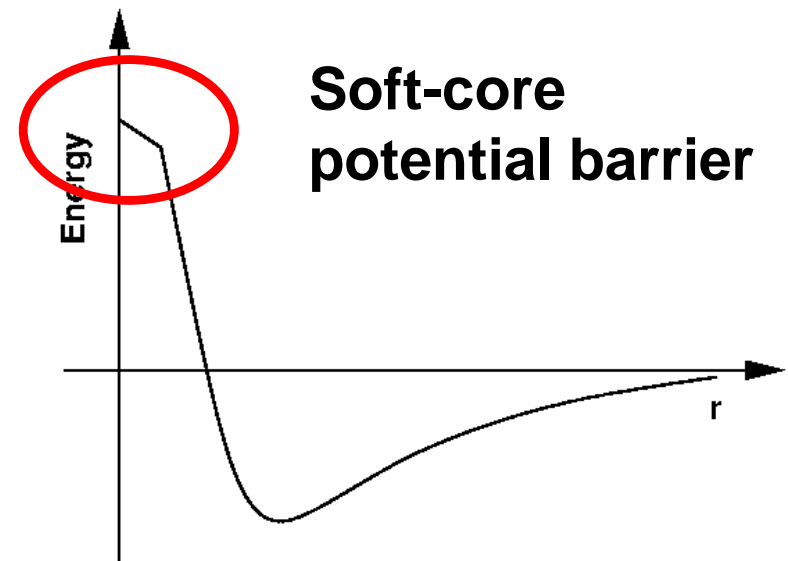
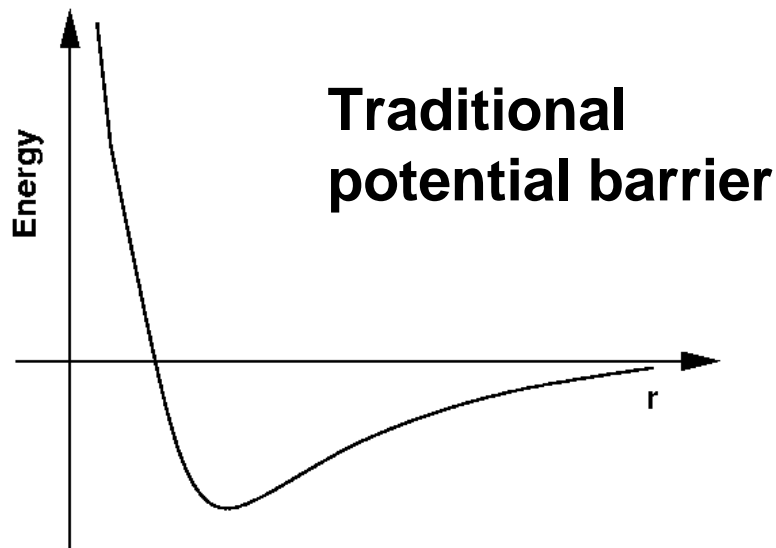
3D lattice

Each point → 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**

Modeling Protein-Ligand Interactions (II)

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



Facilitate the penetration of ligands into proteins

Algorithm Evaluation

Issues under investigation

- Characterization of our docking method:
 - Does the MD length affect the docking accuracy?
 - Does the number of trials affect the docking accuracy?
- Comparing our docking method with other well-known docking methods

Experimental testbed

- Data set: 31 protein-ligand complexes
 - 10 proteins
 - 31 ligands with different levels of complexity (number of rotatable bonds)

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 $a \text{ \AA}$
[Bursulaya, *et al.*, J.Comp.-Aided M.Design, 2003]

Simulation time:

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

Four Different Molecular Dynamics Settings

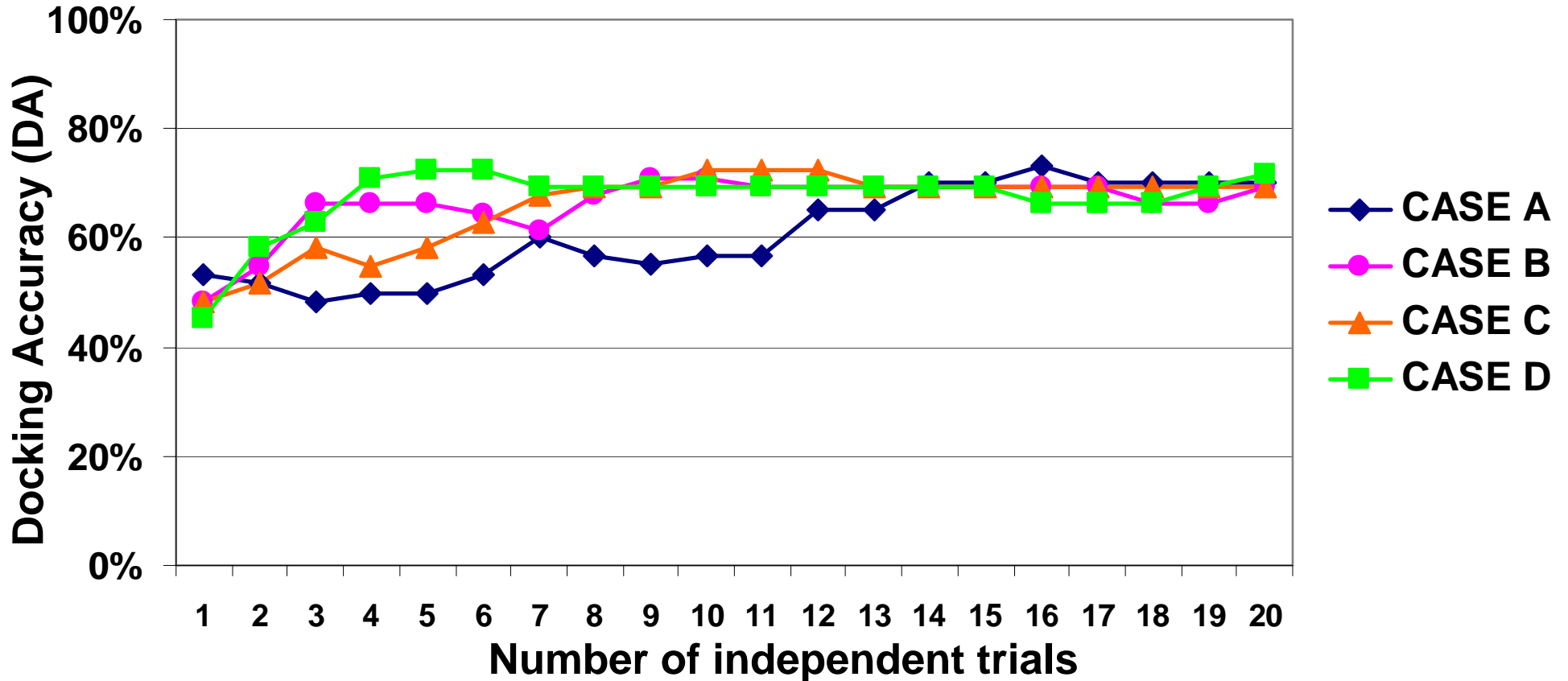
	Heating Phase 300K → 700K # of MD steps	Cooling Phase 700K → 300K # of MD steps
Case A	1000	2500
Case B	2000	5000
Case C	4000	10000
Case D	8000	20000

time



Docking Accuracy

Average DA for the 31 protein-ligand complexes:



- Longer MD simulations **do not improve** the DA
- Ten trials per attempt ensure enough accuracy (T10)

MD Docking versus other Docking Methods

Other Methods:

- AutoDock
- DOCK
- FlexX
- ICM
- GOLD

Comparison Metrics:

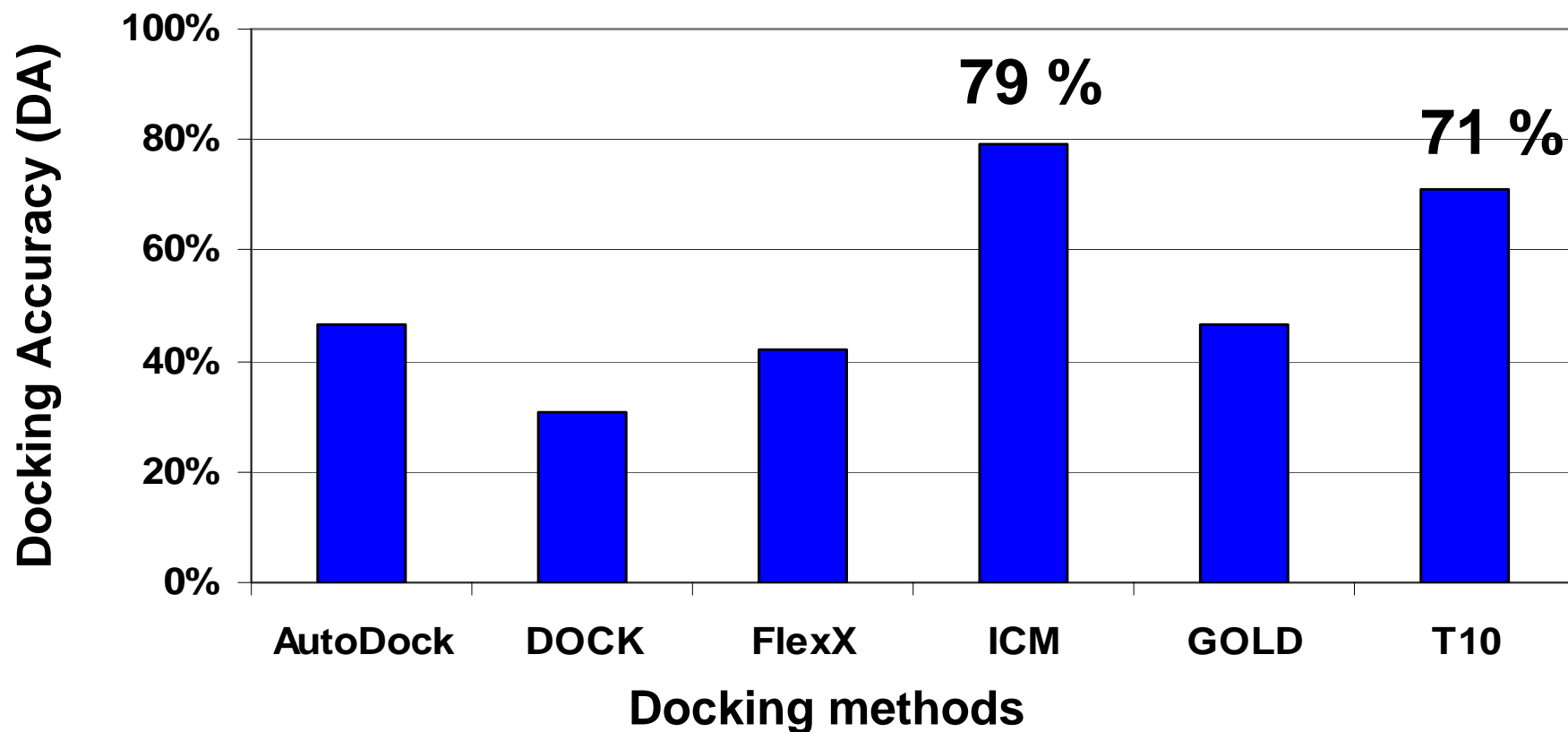
- Docking Accuracy (DA)
- RMSD of predicted ligands
- CPU time per attempt

Definition of attempt:

- Our docking method: for each protein-ligand complex we consider **CASE B** and **10 trials per attempt (T10)**
- Other methods: docking attempt controlled by **recommended parameter settings** [Bursulaya, *et al.*, J.Comp.-Aided M.Design, 2003]

Comparison of Docking Accuracies

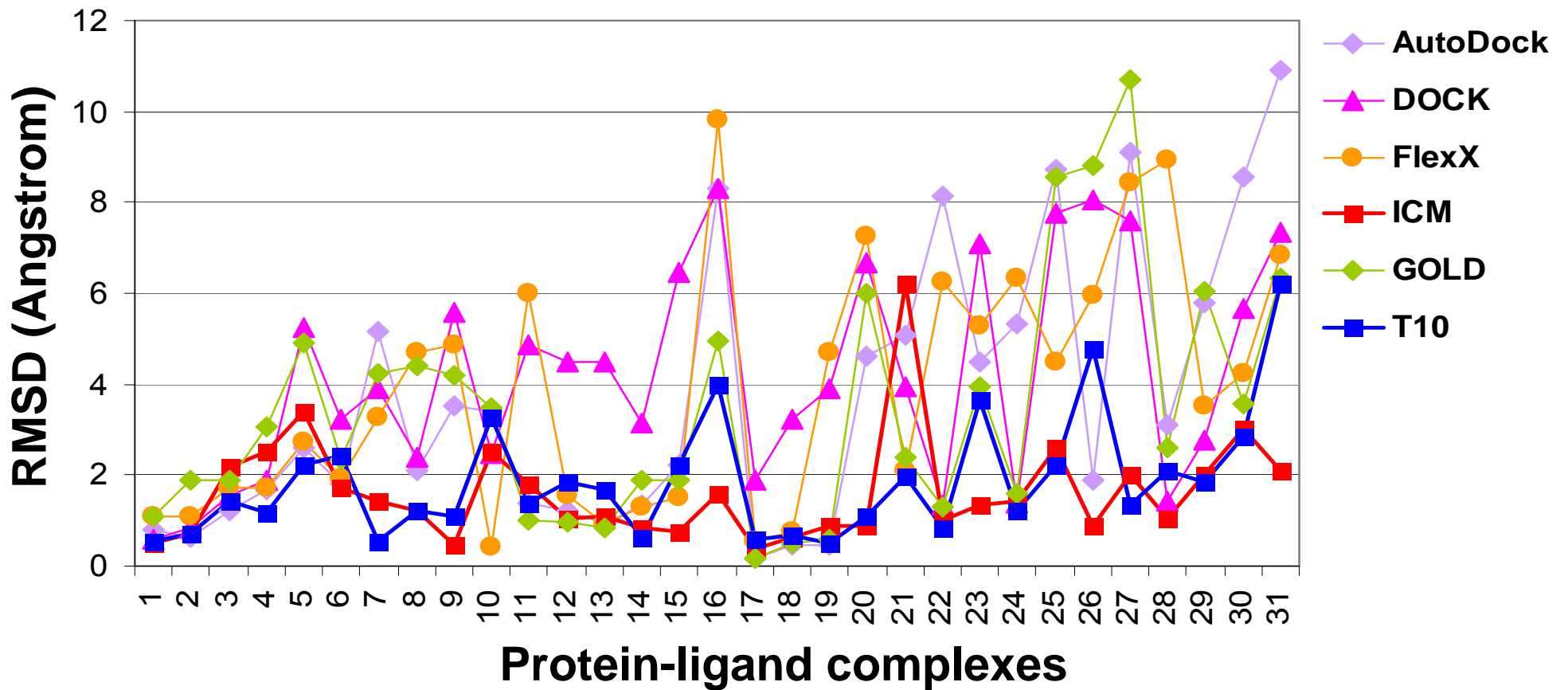
Average DA for the 31 protein-ligand complexes:



Our method T10 provides better DA, except for ICM

Best RMSD of Predicted Ligands

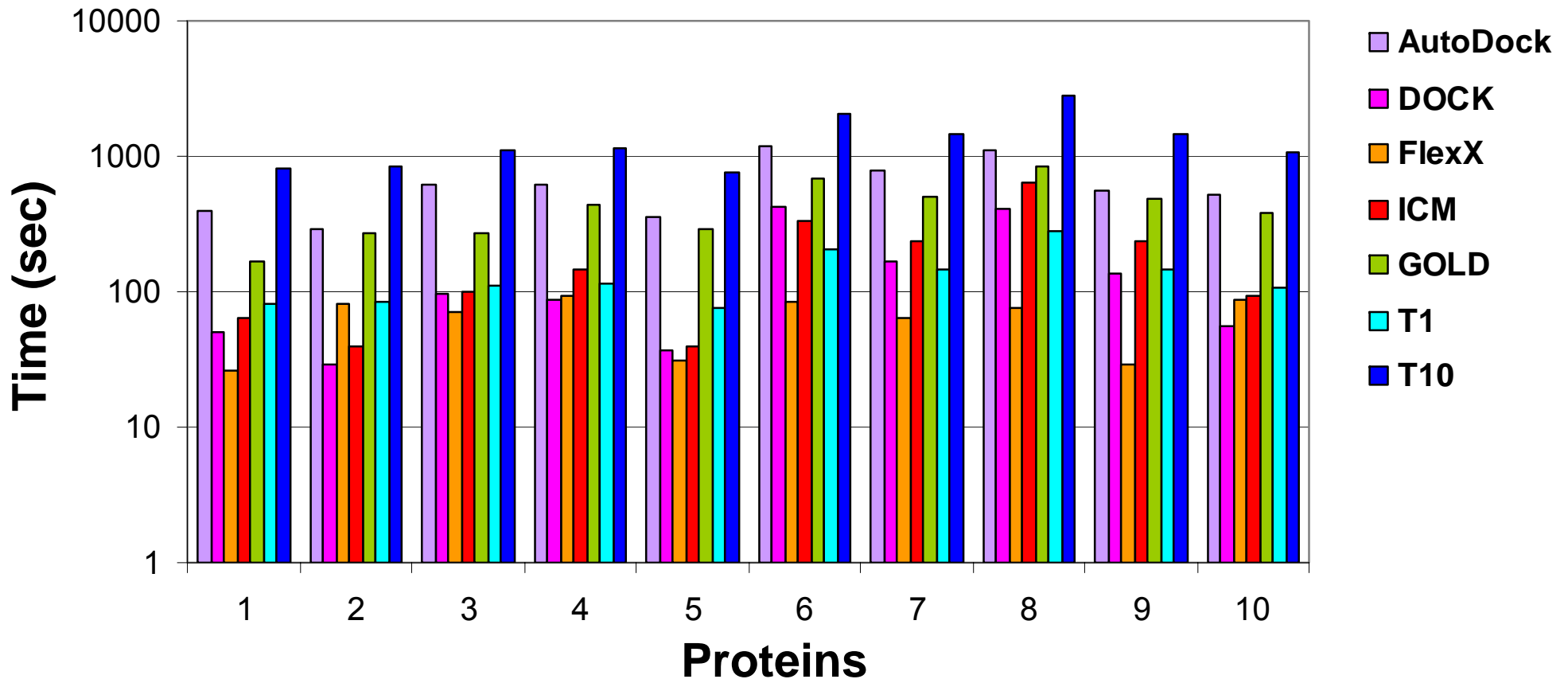
RMSD of predicted ligands from their published crystal structures



Our method provides an RMSD competitive with ICM

Time Comparison

Average docking time in seconds per protein:



If **enough processors** are available, the time for completing a protein-ligand docking **is competitive** with the other methods

Summary

- Accurate, tightly coupled algorithms for docking can be decoupled in flexible, independent and variable-in-length computation
- We have presented a **MD-based docking method**:
 - Reaches an average **accuracy of 71%**
 - Can easily be adapted to the desktop grid characteristics
- Still a lot of **exciting research** has to be addressed both at the application and system levels → to get self-adapting docking on self-adapting desktop grids.

Future Work in Decoupling Protein-Ligand Docking

System level

- Dynamic adaptation at run-time of:
 - Number of trials per protein-ligand complex based on its complexity
 - Number of ligand orientations per trial based on resources and node reliability
 - Desktop workers based on resources availability

Application level

- Screening of large sets of protein-ligand complexes (10,000 molecules) using our adaptive algorithm and desktop grids