High-Performance Computing in Pharmaceutical Research:
From Virtual Screening to All-Atom Simulations of Biomolecules

Jan Kriegl, Computational Chemistry
Boehringer Ingelheim Pharma GmbH & Co. KG
Biberach a.d. Riss

HPC User Forum October 6&7 2011, Stuttgart
Outline

- Why HPC in Pharmaceutical Research?
- Application Examples:
  - Virtual screening
  - High-level quantum mechanical calculations
  - Molecular dynamics simulations of proteins
  - In-silico profiling of molecules
- Summary
- Future Perspective
The Drug Discovery Process: Pharmaceutical Research

High Performance Computing in Pharmaceutical Research:
Support the identification and optimization of new molecular entities by computational methods.
High Performance Computing in Pharmaceutical Research:
Support the identification and optimization of new molecular entities by computational methods.
Virtual Screening
Find the needle in the haystack!

Which of these molecules could be biologically active?

N.B. Some estimates tell us that the chemical universe comprises more than $10^{50}$ molecules!
You already know that a certain molecule binds to your protein of interest. Look for other (similar?) molecules that might also bind but display different overall properties.

Comparison of descriptors, bit strings, cliques, ... ... or 3D shapes.

Lessel et al., 2011
Virtual Screening

You know the 3D structure of your protein of interest and you know where the molecule should bind. Look for molecules that might fit into the binding pocket.
The distribution of charges in the atomic shell gives rise to attractive or repulsive forces between atoms.
Small molecules need to adopt a certain conformation in order to fit into the binding site of the protein. Is this conformation accessible at body temperature?

Systematic conformational analyses: fragments of molecules

Example 1: 2D torsional scan
30° steps each
169 parallel jobs
Density functional theory, 6-31 G*
vacuum

Example 2: 1D torsional scan
~20 heavy atoms
Density functional theory, 6-31 G*
vacuum
10° steps, full optimization at each step per optimization: up to 30 h on 8 CPUs.
Molecular Dynamics Simulations: Flexibility of Biomolecules

Molecular dynamics simulations:

Follow the thermal movement of each atom within the force field of all other atoms up to several 100 nanoseconds.

- All atoms are treated as individual particles: $10^2-10^5$ atoms.
- The simulation obeys Newton’s equation of motion.

We want to assess:
- conformational substates of proteins
- the stability of ligand binding modes
- free energies of binding, ...
Molecular Dynamics Simulations: Flexibility of Biomolecules

Computation time: 2-3 days on 64 CPUs
Molecular Dynamics Simulations: Flexibility of Biomolecules

Tiotropium (Spiriva®) in the human M3 receptor (model)

- 66877 atoms
- 10093 water molecules
- 282 amino acids (hM3 receptor)
- 2 ns/day on 8 CPUs

- 238 phospholipids
- 24382 heavy atoms
HPC Setup

Distributed Computing

Queueing System Master
(Sun Grid Engine)

Linux Server
Execution-Hosts

Results

Workstations
Submit-Hosts

Queue

Linux Cluster
Execution-Hosts

Workstations
Execution-Hosts

Massive, trivial parallelization: chunks of compounds are processed in parallel. Multiple, simultaneous access to file system for I/O (high throughput storage).
Parameters like metabolic stability, solubility and many others are closely linked to molecular properties. In-silico surrogates are applied to get a first assessment of these parameters.

<table>
<thead>
<tr>
<th>molecule</th>
<th>ID</th>
<th>XlogP</th>
<th>pKa</th>
<th>H_don</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ DICLOFENAC STRUCTURE ]</td>
<td>Diclofenac</td>
<td>4.4</td>
<td>3.9</td>
<td>2</td>
</tr>
</tbody>
</table>
For a complete in-silico profile many different prediction engines must be managed in parallel. The results from all the individual models are then gathered and reported back to the sender.
HPC @ BI (Biberach)
High Performance Computing Cluster (4\textsuperscript{th} generation)

1664 CPUs:
- 13 Enclosures à 16 blades
- 8 CPU-Cores per blade
- Infiniband/Ethernet
- Lustre and EVA storage
An efficient HPC environment is key to CompChem in Pharma Research.

The applications are highly diverse, ranging from RAM and CPU demanding tasks to thousands of short tasks which are massively parallelized.

Short response times are indispensable to feed CompChem work into the iterative design cycles of drug discovery programs.

Execution of heterogeneous scientific software requires a flexible HPC environment.
A high need for enhanced compute power is foreseen because
- the complexity of the systems under investigation increases.
- more accurate, yet computationally more demanding methods like QM/MM or ultra-long MDs will continue to mature.

An increase of at least two orders of magnitude will be necessary to fully establish these methods in the drug discovery process, together with novel methods to analyze and aggregate the continuously growing data heap.

Investment and maintenance costs need to be kept at an affordable level.

Modern HPC requires a sustainable concept for energy consumption.
## Acknowledgements

### Computational Chemistry
- Bernd Beck
- Thomas Fox
- Uta Lessel
- Peter Haebel
- Sandra Handschuh
- Clara Christ
- Jasna Klicic
- Matthias Zentgraf
- Nils Weskamp
- Michael Bieler
- Christofer Tautermann
- Bernd Wellenzohn
- Alexander Weber
- Daniel Seeliger
- Herbert Köppen
- Ulrike Küfner-Mühl

### Scientific Computing Services
- Andreas Teckentrup
- Oliver Wissdorf
- Johannes Koppe
- Peter Stauffert
- Matthias Röhm

### Collaboration Partners (excerpt)
- Andreas Hildebrandt (University of Mainz)
- Benny Kneissl (University of Mainz)
- Klaus Liedl (University of Innsbruck)
- Hannes Wallnoefer (University of Innsbruck)
- Gisbert Schneider (ETH Zürich)
- Tim Clark (University of Erlangen-Nuremberg)
- Chris Skylaris (University of Southhampton)
- Steven Fox (University of Southhampton)
- Jürgen Bajorath (University of Bonn)
- Anne M. Wassermann, Matthias Waver (University of Bonn)