# Scripting Workflows with the Application Hosting Environment

Stefan Zasada

University College London

Constructing workflows with the AHE

- By calling command line clients from Perl script complex workflows can be achieved

- Easily create chained or ensemble simulations

- For example jobs can be chained together:
  - ahe-prepare → prepare a new simulation for the first step
  - ahe-start → start the step
  - ahe-monitor → poll until step complete
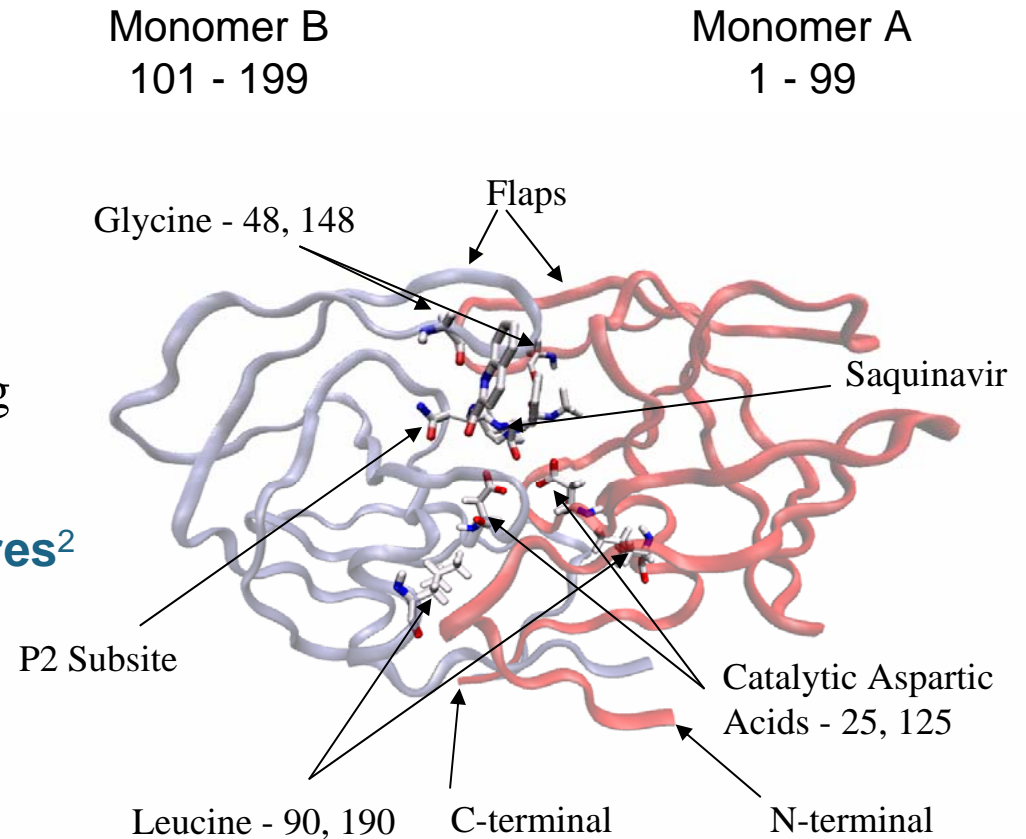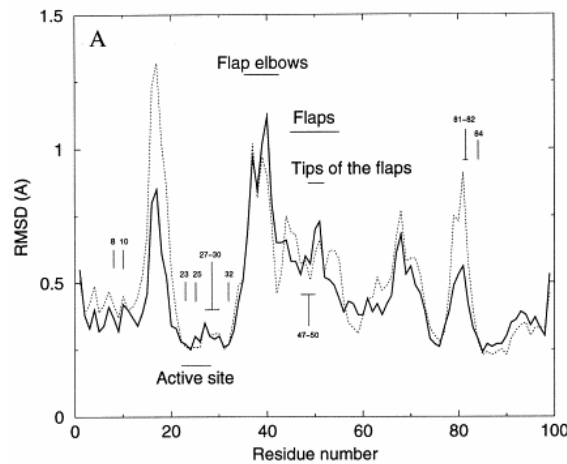 - ahe-getoutput → download output files
  - repeat for next step

Constructing a workflow

ahe-prepare

ahe-start

ahe-monitor

ahe-getoutput
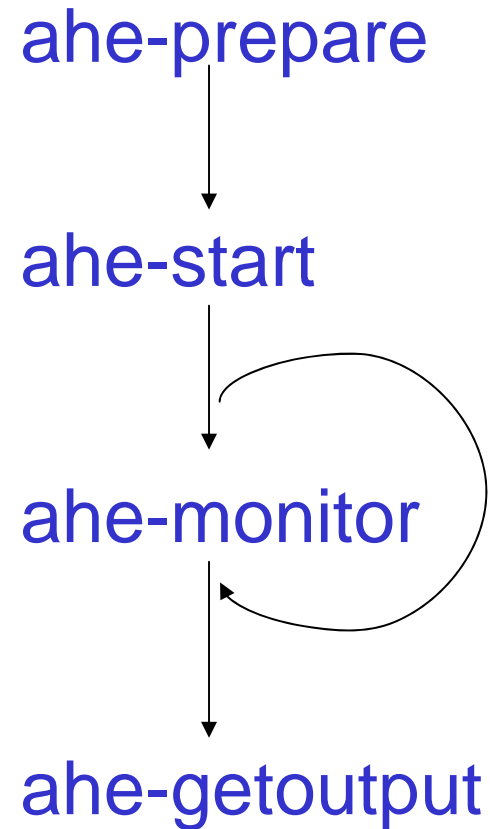
HIV-1 Protease

- Enzyme of HIV responsible for protein maturation
- Flexible Flaps open to provide access to active site[1]
- Catalytic Asp Dyad cleaves peptide bond
- Example of Structure Assisted Drug Design – 8 FDA inhibitors

RMSD of existing crystal structures[2]

Monomer B
101 - 199

Monomer A
1 - 99

Flaps

Glycine - 48, 148

Saquinavir

P2 Subsite

Catalytic Aspartic
Acids - 25, 125

Leucine - 90, 190   C-terminal   N-terminal

1. Hornak et al., PNAS, 2006, 103 4, 915-920
2. Zoete et al., J. Mol. Biol., 2002, 315, 21-52

Computational Techniques

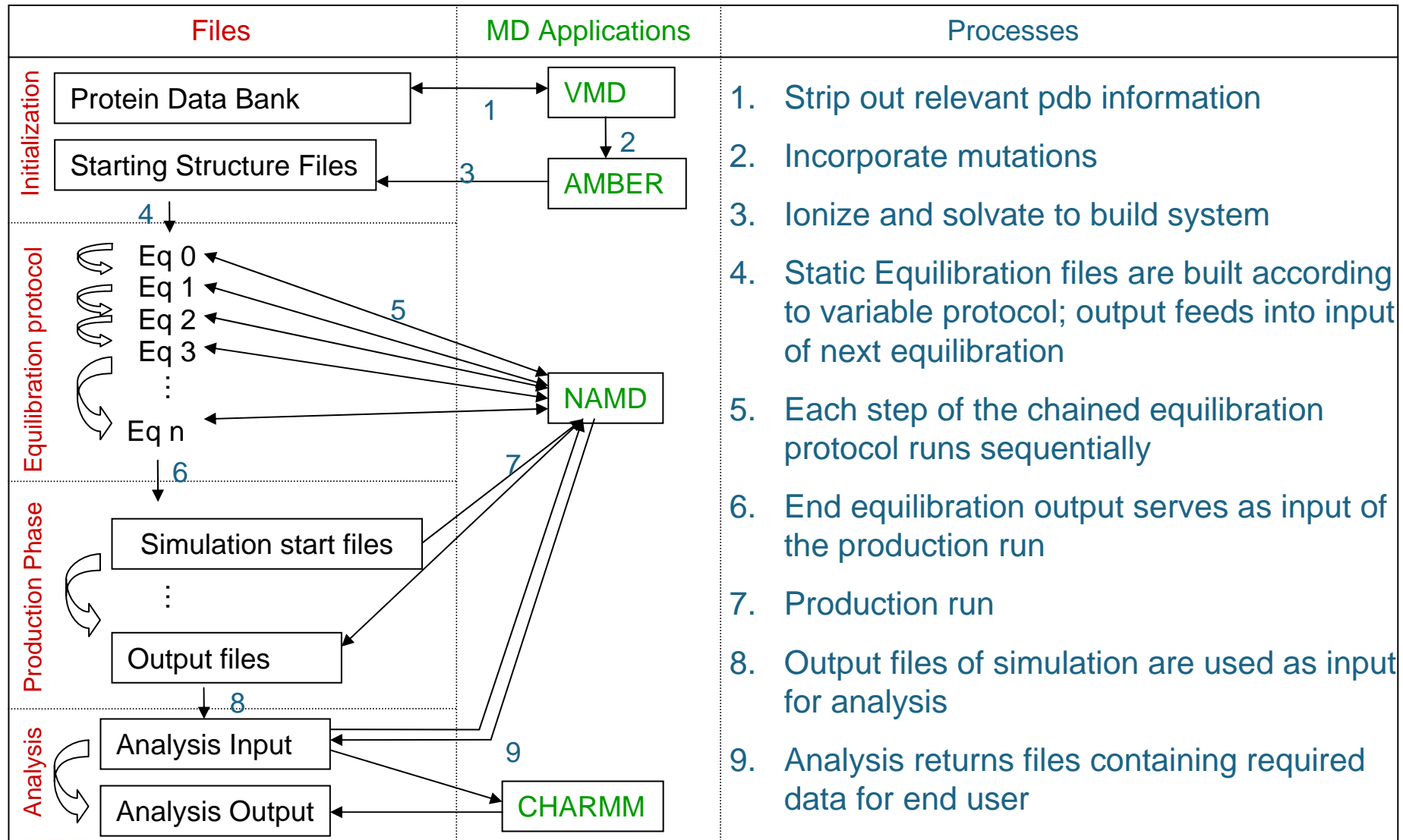## Ensemble MD is suited for HPC GRID

- Simulate each system many times from same starting position
- Each run has randomized atomic energies fitting a certain temperature
- Allows conformational sampling

Start Conformation   *Series of Runs*   *End Conformations*

Launch simultaneous runs
(60 sims, each 1.5 ns)

Cx

C1

C2

C3

C4

Equilibration Protocols

eq1 eq2 eq3 eq4 eq5 eq6 eq7 eq8

Simulation Workflow

Files	MD Applications	Processes

Initialization

Protein Data Bank ←→ VMD (1)

Starting Structure Files ← (3) ← AMBER (2)

Equilibration protocol

(4)

Eq 0
Eq 1
Eq 2
Eq 3
⋮
Eq n

(5) → NAMD

(6)

Production Phase

Simulation start files

⋮

Output files

(7)

(8)

Analysis

Analysis Input

Analysis Output ← CHARMM (9)

Processes:

1. Strip out relevant pdb information

2. Incorporate mutations

3. Ionize and solvate to build system

4. Static Equilibration files are built according to variable protocol; output feeds into input of next equilibration

5. Each step of the chained equilibration protocol runs sequentially

6. End equilibration output serves as input of the production run

7. Production run

8. Output files of simulation are used as input for analysis

9. Analysis returns files containing required data for end user

6

Practical 2

Using Perl:

- Write script to automate preparing and starting a job, and staging files

- Write script to distribute jobs around a number of resources

- Write script to chain jobs, one after the other

http://www.realitygrid.org/AHE/training/nesc/ex2/