ScalaLife
Scalable Software Services
for Life Science

Workpackage 4
Application Validation

Ferdinand Jamitzky, LRZ

Contents

- Introduction WP4
- Summary Deliverable D4.1
- Software Installation Progress
- Alpha user groups: test cases received
- ScaSCC hardware updates
- Open Issues
- Winterschool 2012
- Application Validation Suite
- Demo pre-alpha version (Momme)

WP4: Application Validation

The objective of WP4 is to verify and validate the enhanced applications and assess their performance on various resources. This **service** enables developers and users to rely on a common set of data, software and hardware, which has been thoroughly tested and can be used as a **"gold-standard"**. After validating the modified applications from WP3 against selected test examples a set of important and current test problems will be chosen in collaboration with a group of selected test users which serves as an **alpha user community** which will eventually be the seed for the user community of the **competence centre** in WP5.

Summary Deliverable D4.1

The project partners agreed on the following items:

- Communication between the developers and the ScaSCC is performed via the technical **mailing list** (scalalife-tech).
- The project releases internal versions of the ScaPSP (Software Packages) **twice** a year.
- These releases are installed at the ScaSCC for **at least two** hardware architectures each.
- The ScaSCC **validate** the ScaPSP with respect to scaling behavior for varying numbers of cores on their respective hardware.
- Each computing center of the ScaSCC chooses **at least one** Alpha User Group (ScaAUG) that gets access to the installed software (ScaPSP) on the respective computing centers of ScalaLife (ScaSCC) and can store the results in a dedicated server. This dedicated server will be set-up as part of the available resources

Mailing List scalalife-tech

- Mailing list is not used at the moment, why?
- Please register! Please use!

or:

- Shall we use a different means of communication?
- Forums?
- Ticketing System?

Software Installation Progress

- GROMACS 4.6 (new version) installed on:
  - LRZ Linux Cluster
  - BSC Mare Nostrum
  - OeRC yes
  - KTH
- DISCRETE 0.1.1a (new version) installed on:
  - LRZ Linux Cluster
  - BSC Mare Nostrum
  - OeRC yes
  - KTH

Software Installation Progress

- DALTON
    - requires registration for usage and will only be distributed to ScaCC (internal release password protected)

Alpha user groups: test cases received

So far we have received the following test cases from alpha users:

GROMACS

- Large membrane protein system (Guillem Portella, BSC)
- SNAP: predict the functional change of SNPs on a protein's function (Marc Offman, rostlab, LRZ)
- Stretching the envelope: large scale coarse-grained simulations of an influenza virion (Philip Fowler, OeRC)
- Modeling of mixed lipid bilayers (Alexander Lyubartsev, SU, KTH)

DALTON

- Anharmonic vibrational calculations for larger molecules (Johannes Weber, TUM, LRZ)
- Porphyrin Ring System (Clemes Woywod, TUM, LRZ)

DISCRETE

First Alpha User Group Meeting 2011

- VC with all alpha users planned within the next weeks
- Topics:
    - State of ScaPSP
    - Additional tools for ScaPSP?
    - Access to computing time at ScaCCs
    - Access to produced data at ScaCCs
    - Assistance with project proposals
    - Interest in joint publications?
    - Additional test and validation cases?
    - Conferences and Workshops?
    - International Scalalife User Conference in 2012?

ScaSCC validation hardware

- New Hardware at LRZ:
  - Fermi GPGPU Cluster (8GPGUs, 64cores) (April 2011)
  - IBM DataPlex (8000 cores) (August 2011)
    former: LRZ SGI UV, old GPGPU Cluster
- OeRC
  - SGI Altix 4700
  - SGI ICE
- BSC
  - MareNostrum (Power)
  - Intel Xeon Cluster
- KTH
  - Lindgren (Cray XE6)
  - Povel Cluster (Opteron)

Open Issues

- User accounts for alpha users for all ScaCC?
  - at the moment alpha users have to apply for a project and then obtain computing time from the respective ScaCCs. Do we want all alpha users have accounts on all ScaCCs?
- Shall we write a DECI Project for validation? Who will? All alpha users or scalalife consortium?
- Data transfer (together with WP7):
  - Where do we store the validation data sets?
  - Do we want a global file system (gpfs or such)?
- e-Infrastructure (together with WP5):
  - Who will make contact with PRACE, EGI, DEISA? PMB? EB?

Important Dates

- **Dec 2010 M21**
  Documentation and collaboration with the LS researchers about possible test cases available **(still need some input!)**
- **Feb 2011 MS13**
  Development version of GROMACS, Dalton, and DISCRETE available to WP4 **(GROMACS and DISCRETE uploaded)**
- **Aug 2011 MS14**
  Development version of GROMACS, Dalton, and DISCRETE available to WP4
- **Aug 2011 M22**
  First application validation suite ready and porting to the systems
- **Aug 2011 D4.2**
  Report on experience in preparing and porting the application examples to the system (work will start next month)

Winterschool 2012 at LRZ

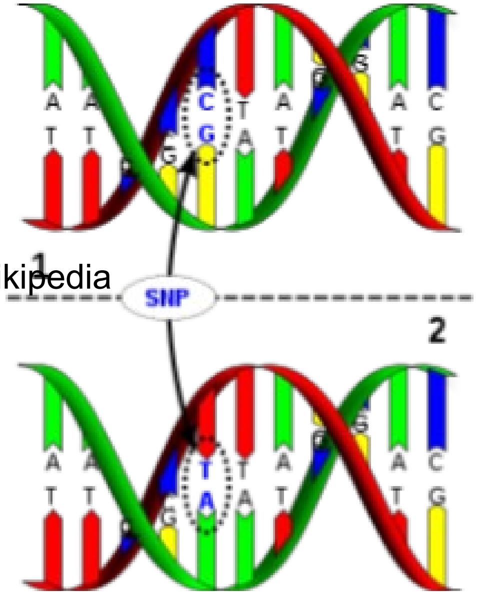
- First International Winterschool on Scalable Software Services in Life-Science 2012
- One week winterschool from 27.2. - 2.3.2012
- Programme:
  - Mon 27.2. Schrödinger (desmond) / nVidia
  - Tue 28.2. GROMACS / nVidia
  - Wed 29.2. DALTON
  - Thu 1.3. DISCRETE / NWChem?
  - Fri 2.3. Skiing Trip to "Wilder Kaiser"

Funding from nVidia and schrodinger possible?
How to approach international communities?

Validation Suite for ScaPSP

- Validation Suite Version ScaVS0.1 will be released soon after F2F meeting in BCN.
- It will contain the DEISA benchmark system for GROMACS and the validation case from rostlab.
- Input ca. 100 MB
- Output ca. 1 TB

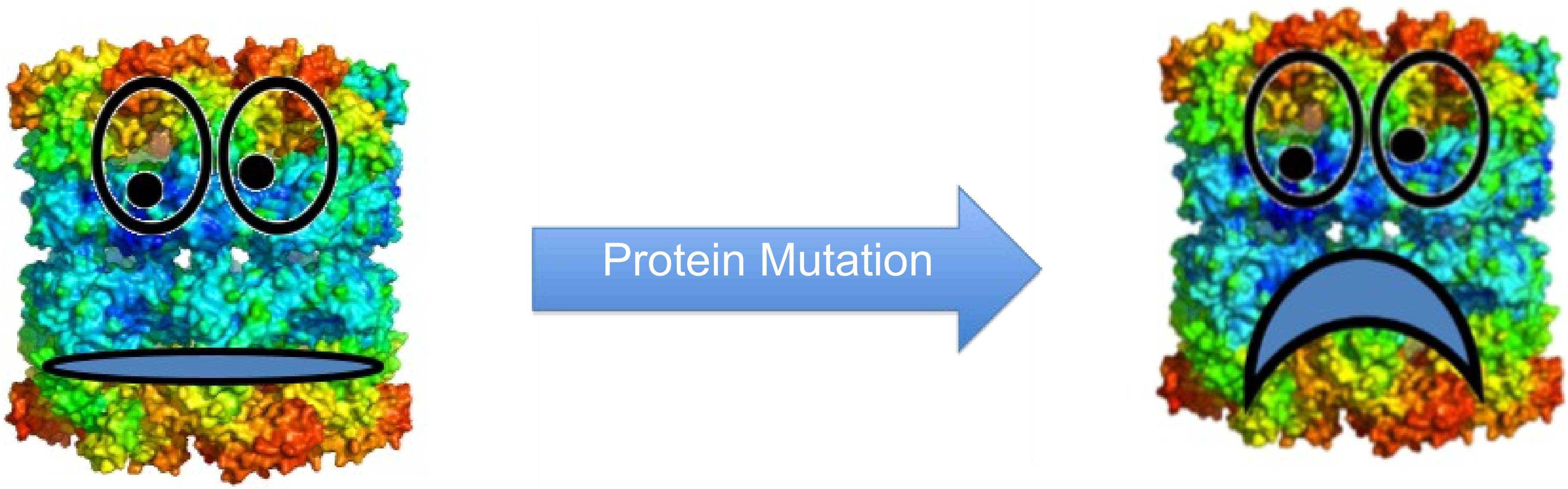
SNPs



m wikipedia

Change of a single nucleotide:

- synonymous □ no change in protein sequence
- non-synonymous □ change in protein sequence
- **change (mutation) in protein sequence can affect protein structure/function**



SNPs and MD

- Large scale analysis of SNPs (single nulceotide polymorphism) and point mutations on protein structure, flexibility and function using Molecular Dynamics (MD)
  - How detailed do the simulations need to be to get a binary answer to: Is protein function negatively affected? (or more)
  - Create “standard” protocol to investigate mutations
  - MD simulations + analysis (tools)
- Dataset:
  - Based on SNAP publication dataset: 6,245 proteins, 81,312 mutations
  - Correlated to PDB (1,781)
- Simulations:
  - Approx. 10ns length
  - Implicit/explicit solvent
  - Partially fixed backbone?

Dataset

	XRAY	XRAY < 2.0Å	XRAY < 1.6Å	XRAY < 1.0Å
HSSP 0	379(6573)	309(6138)	185(3490)	42(1664)
HSSP 40	1326(22692)	904(16406)	400(7815)	47(1941)
HSSP 60	1457(25528)	987(18501)	415(8722)	47(1941)
HSSP 90	1578(27492)	1061(19962)	431(9087)	47(1941)
HSSP 100	1615(27945)	1080(20239)	435(9149)	48(1955)

On average 1GB per simulation □ total 1.6 TB (1.1 TB)

Demo ScalaLife Benchmark Suite

DEMO TIME