

Large Scale Molecular Dynamics Analysis of Single Nucleotide Polymorphisms

Marc Offman

Group retreat talk

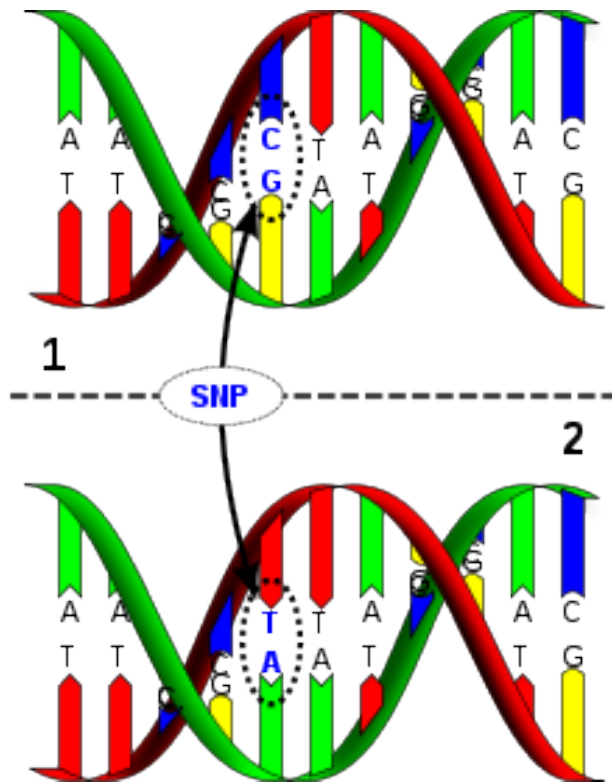
03/04/11



SNPs and MD

- Large scale analysis of single nucleotide polymorphisms (SNPs) 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?
 - What do we gain compared to sequence-based and other more simple structure-based methods?
 - Create “standard” protocol to investigate mutations
 - MD simulations + analysis (tools)

Single Nucleotide Polymorphisms



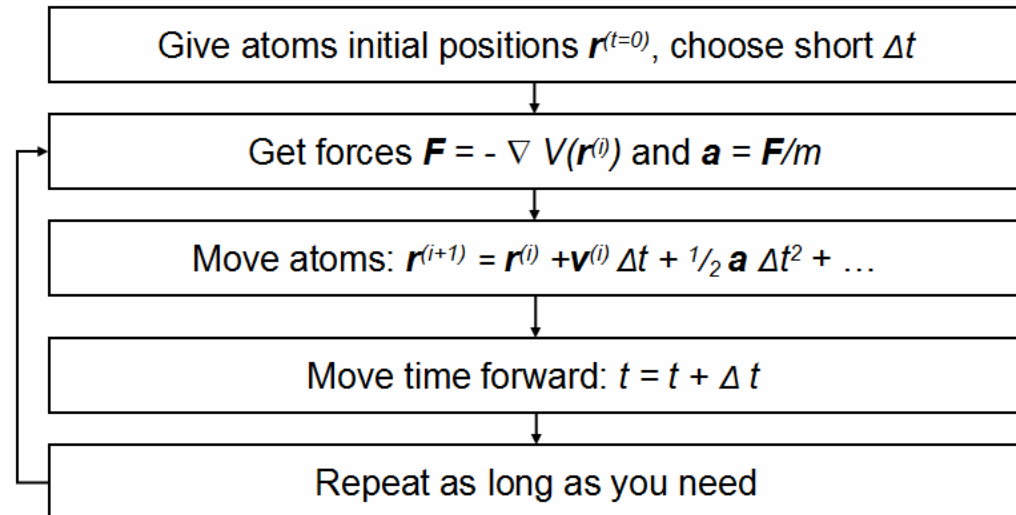
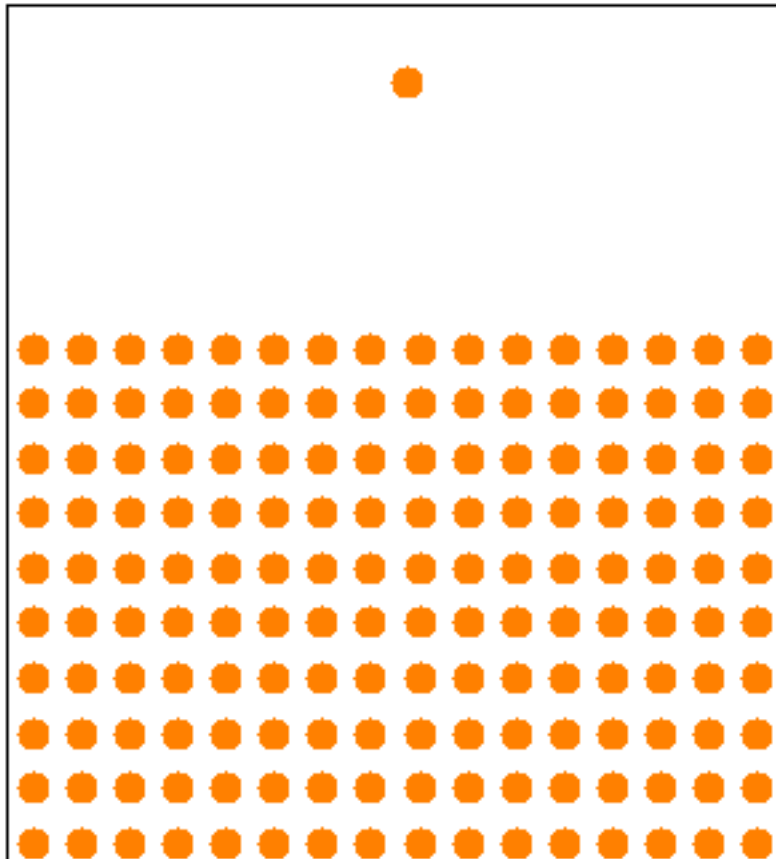
- Changes of a single nucleotide
- Recur in significant proportion of population
- Synonymous SNP: no change in protein sequence
- Non-synonymous SNP: change in protein sequence

Molecular Dynamics

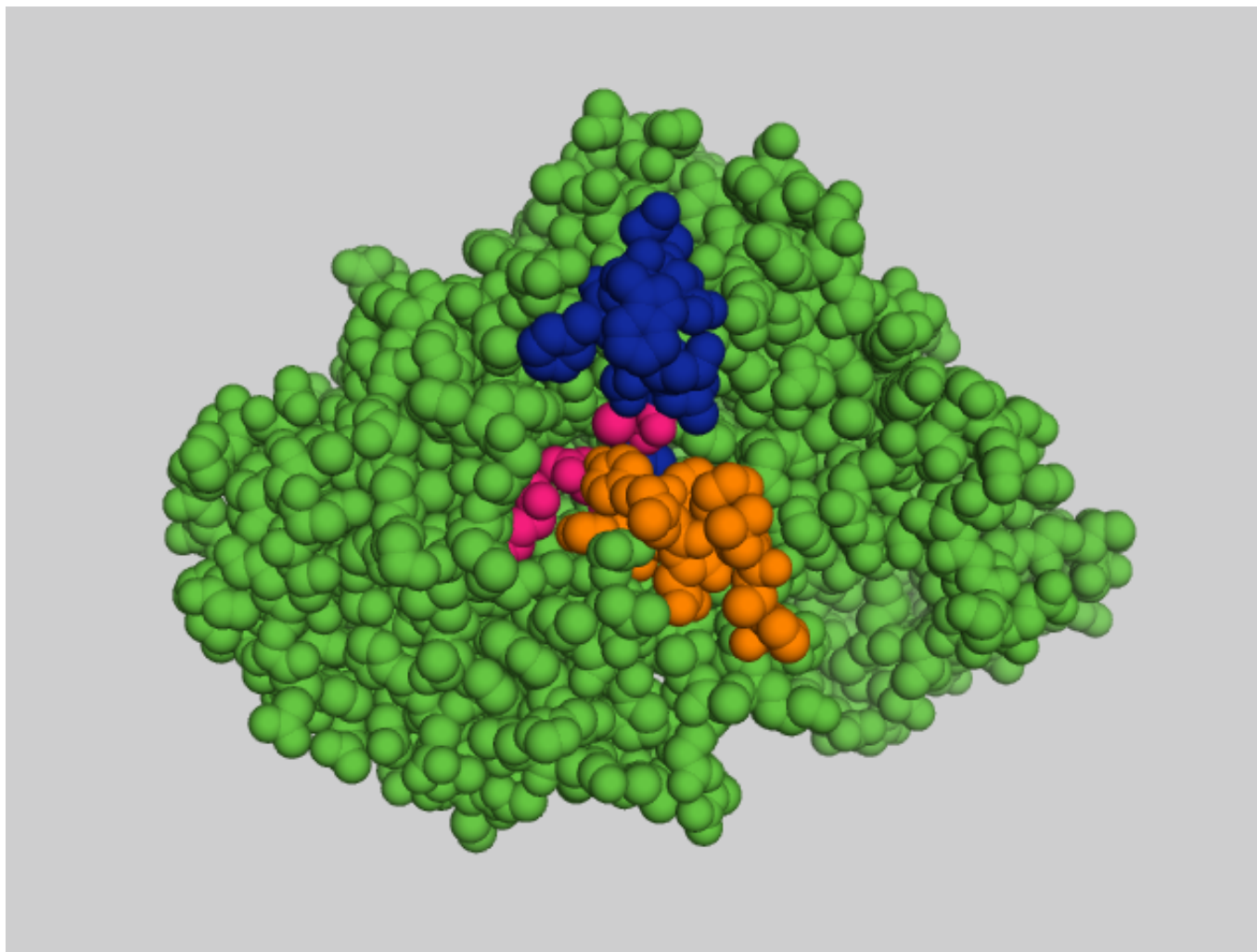
- Atoms move
- Approximation of known physics
- Based on statistical mechanics
- Statistical ensemble averages are equal to time averages
- Simulates interactions of protein and solvent over short time (10ns)

MD - Principles

time 0.0041 ps



MD Simulation



Offman et al., 2010

Material and Methods

- Dataset:
 - Based on SNAP publication dataset: 6,245 proteins, 81,312 mutations
 - Correlated to current PDB
- Simulations:
 - GROMACS package
 - Approx. 10 ns length
 - Implicit/explicit solvent
 - Partially fixed backbone
 - Different optimization techniques

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
 or up to → total 27.9 TB
 That's why it's done on the LRZ !

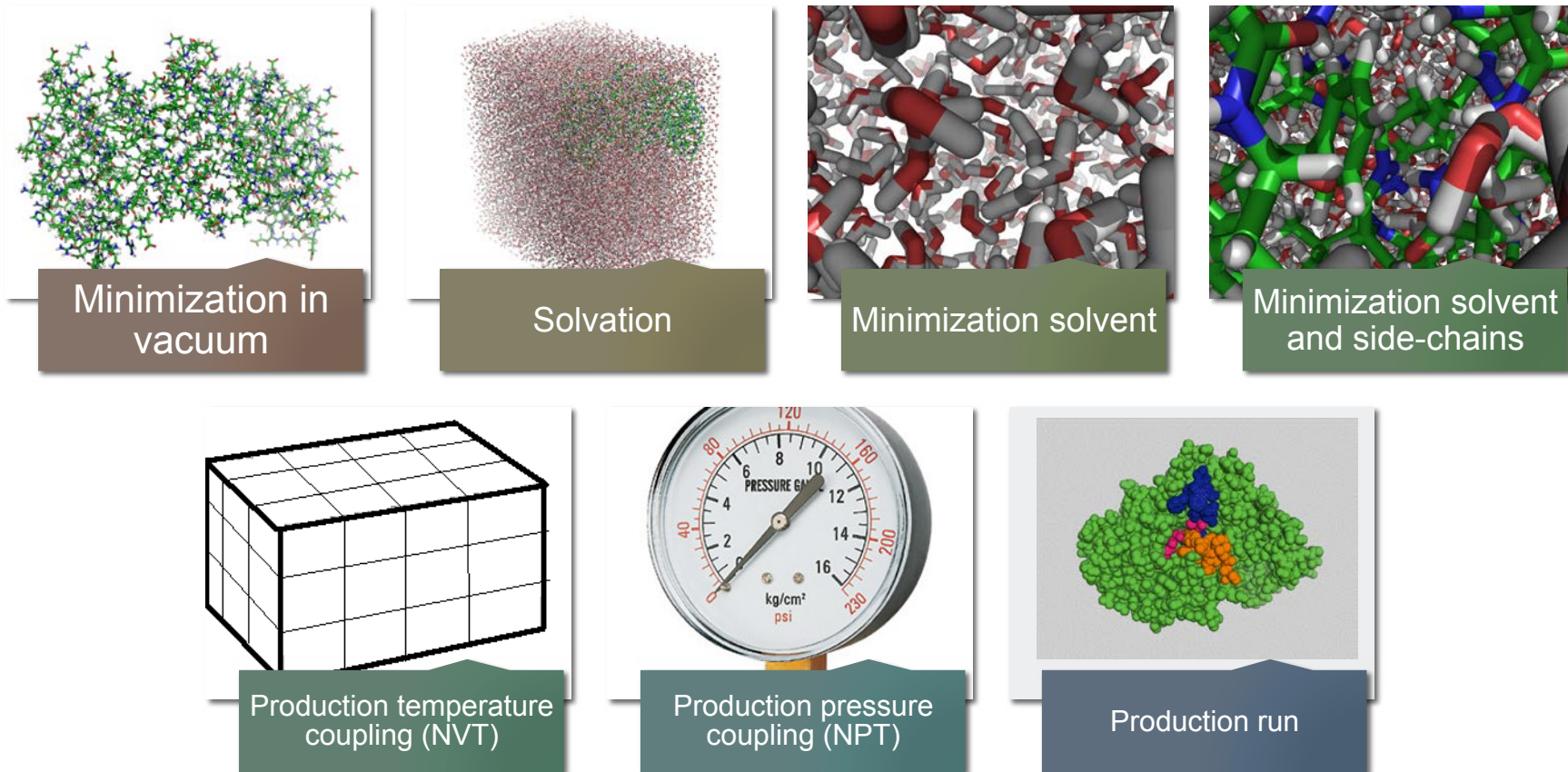


ScalaLife Scalable Software Services for Life Science

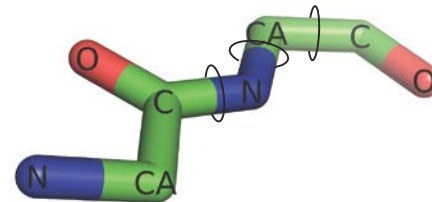
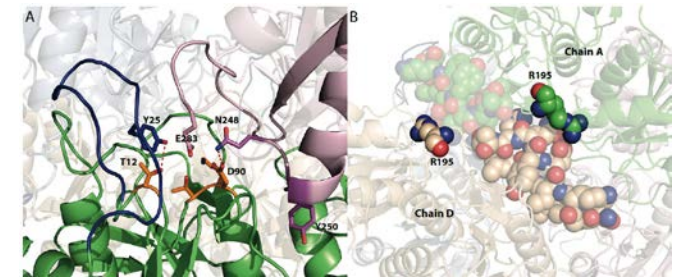
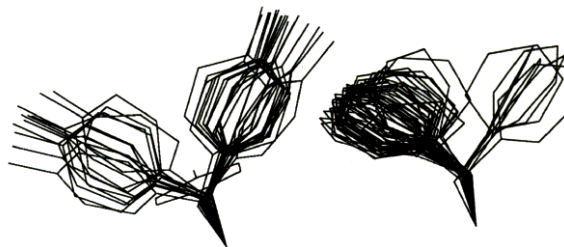
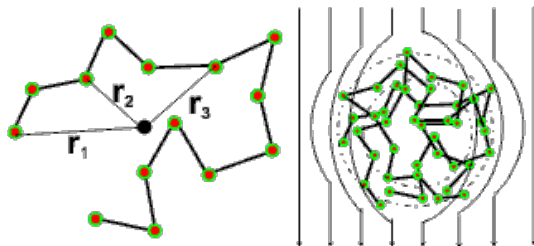
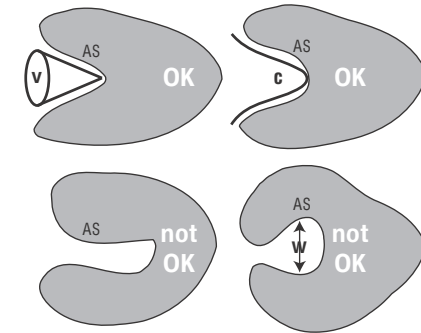
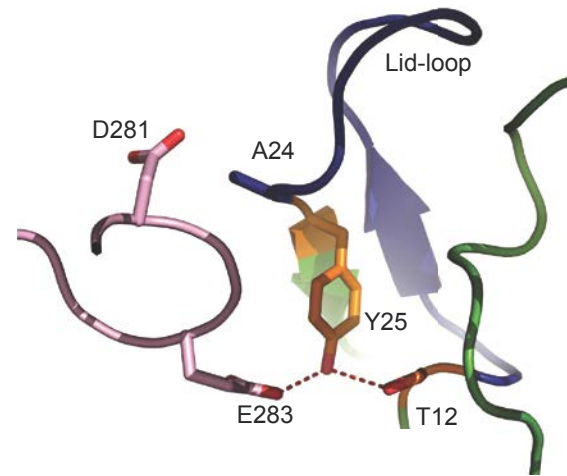
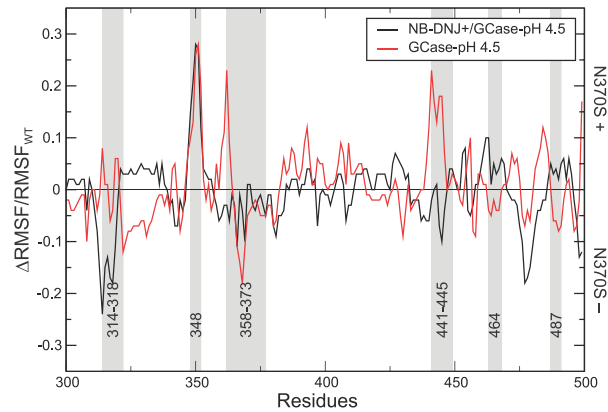


Simulation Pipeline

Simulations with GROMACS package



Structural Analysis



Comparison to...

- Sequence-based methods:
 - SNAP
 - PolyPhen
 - SIFT
- Side-chain configuration:
 - SCWRL
 - SCREAM
- Energy calculations:
 - Pair Potential (foldX)
 - Molecular Mechanics force fields (EEF, GB, PB)
- Normal Mode analysis (minimization)
- Principle Component analysis (MD)

So far...

- Dataset
- Analysis tools and scripts
- Working automatic MD script:
 - single proteins
 - protein complexes
- Theoretically infrastructure
- Some test cases

Acknowledgements

