Method development in computational biology

René Staritzbichler
AG Hildebrand
Institut für Biophysik und Medizinische Physik (IMBP)
Universität Leipzig

CV

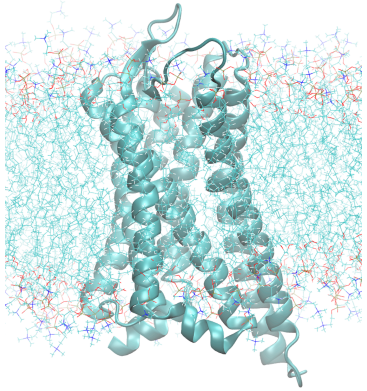
- Diploma in Physics at University Hamburg / Deutsches Elektronen Synchrotron (DESY)
- PhD in Physics at Max Planck Institute of Biophysics
- Postdoc
    - with Jens Meiler at Vanderbilt University
    - Max Planck Institute of Biophysics
    - with Ugur Sahin at BIONTECH
- CEO / Founder at IMMUTHERA GmbH
- Postdoc with Peter Hildebrand at Institute of Biophysics and Medical Physics

http://staritzbichler.com/rene.html

Molecular biology

Dynamics | Binding | Design

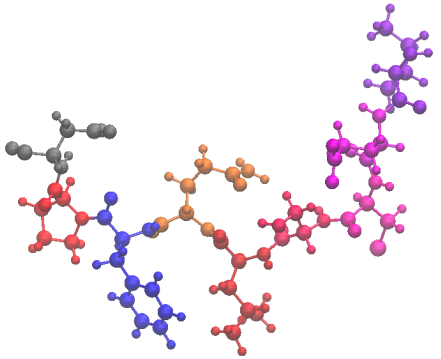
Methods



Machines of the cell

Proteins

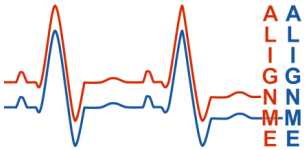
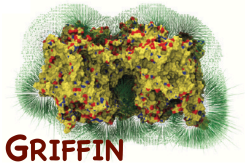
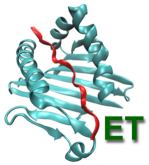
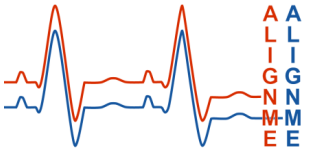
Receptors:
- GPCR
- Immune receptors

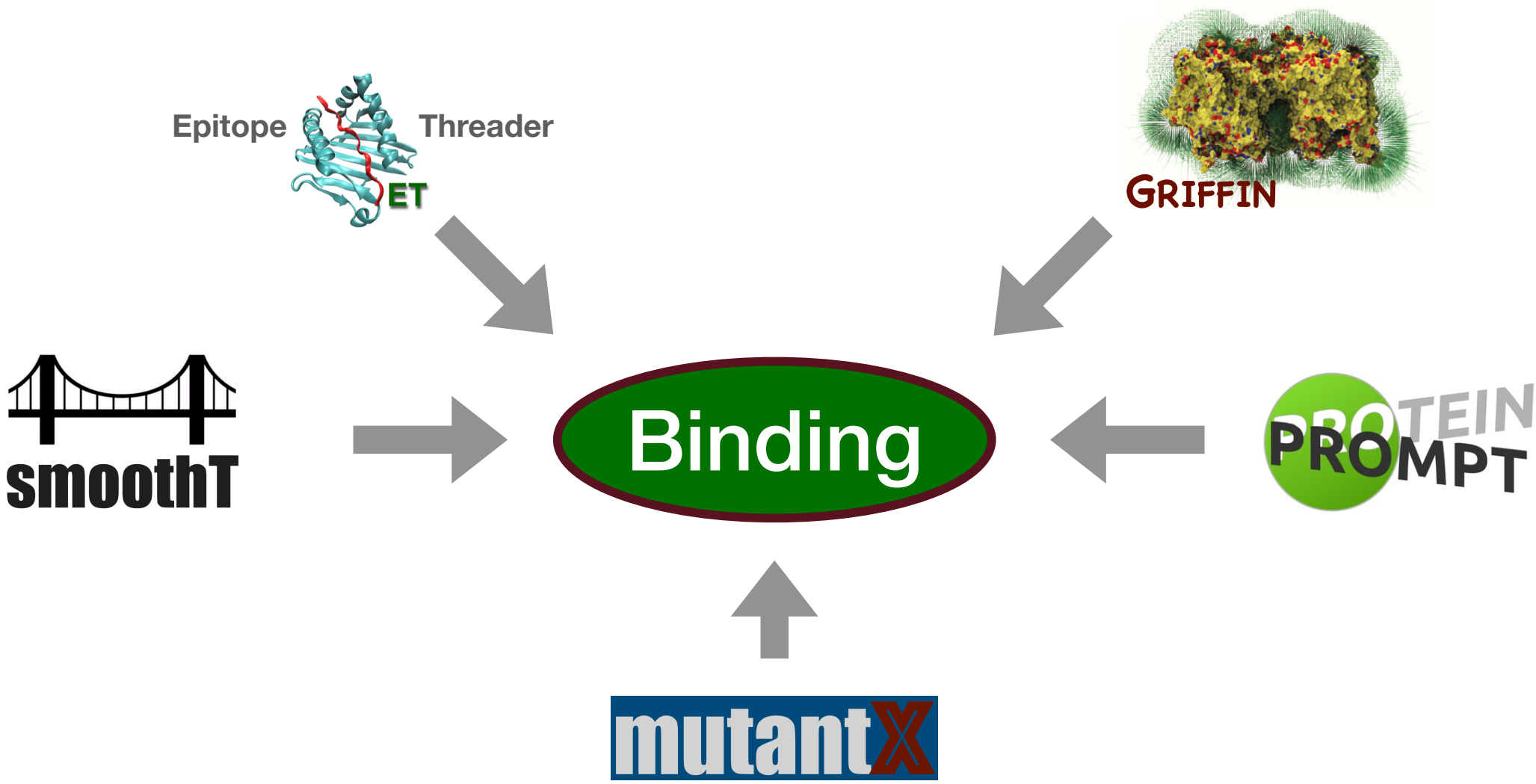


Side effect reduction | Marker | Therapeutics

Biomedical application

Key achievements

Medical	Software (C++)	Webserver (flask,js)
• New laboratory diagnostic method using Raman spectroscopy (patent pending) • Design of an epilepsy drug (currently tested in mice) collaboration with Charité, Uni Innsbruck and Company	 smoothT  PROTEIN PROMPT (python) Epitope Threader  **bcl::fold / bcl::em-fold**	 (cgi/perl) mutantX MDorv PROTEIN PROMPT **Voronoia**


Epitope Threader
ET
GRIFFIN
smoothT
Binding
PROTEIN PROMPT
mutantX

Outline

Webserver

#1 MDsrv
#2 AlignMe
#3 MutantX
#4 ProteinPrompt

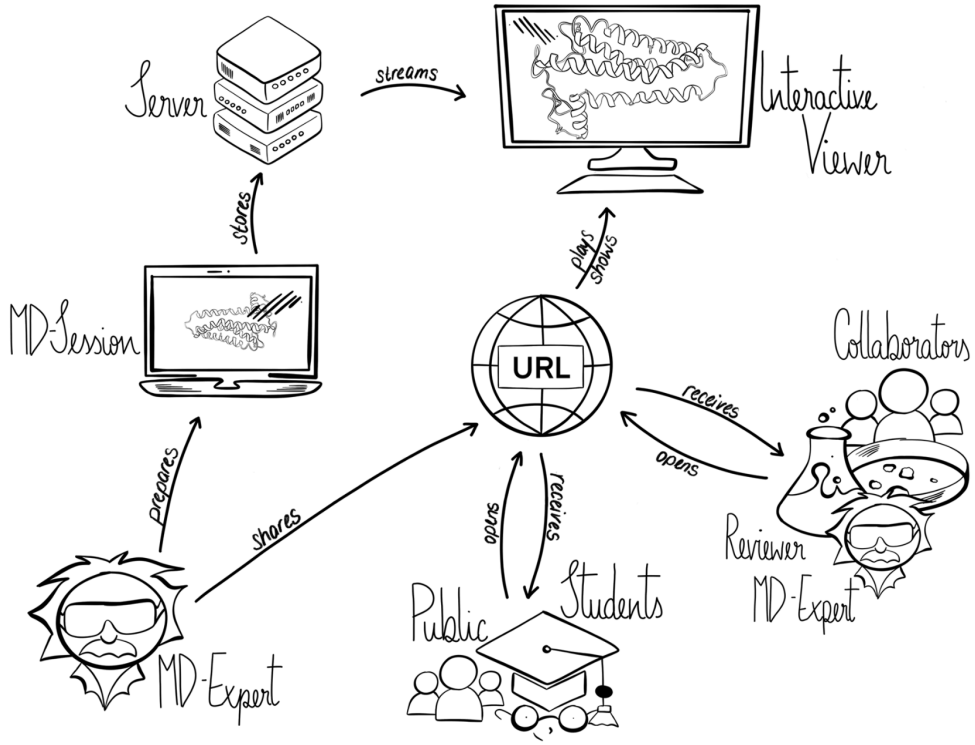
Software

#5 bcl::fold & bcl::em-fold
#6 EpitopeThreader
#7 SmoothT
#8 Griffin

#1 Webserver

@Uni Leipzig

Sharing simulations

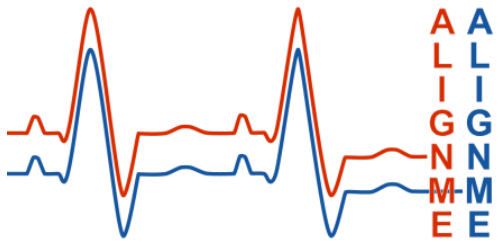


MDsrv: visual sharing and analysis of molecular dynamics simulations.
Kampfrath M, Staritzbichler R, Hernández GP, Rose AS, Tiemann JKS, Scheuermann G, Wiegreffe D, Hildebrand PW.
Nucleic Acids Res. 2022

#2 Webserver / software

@MPI Biophysik

Alignment of proteins lacking sequence similarity



- Similarity of biochemical and predicted profiles
- Advanced gap penalties for e.g. non-breaking of helices
- Expert knowledge: user defined anchors (multiple, varying strength)
- Forward alignment for 3D visualisation to MutantX

AlignMe--a membrane protein sequence alignment web server.
Stamm M, Staritzbichler R, Khafizov K, Forrest LR.
Nucleic Acids Res. 2014

AlignMe: an update of the web server for alignment of membrane protein sequences.
Staritzbichler R, Yaklich E, Sarti E, Ristic N, Hildebrand PW, Forrest LR.
Nucleic Acids Res. 2022

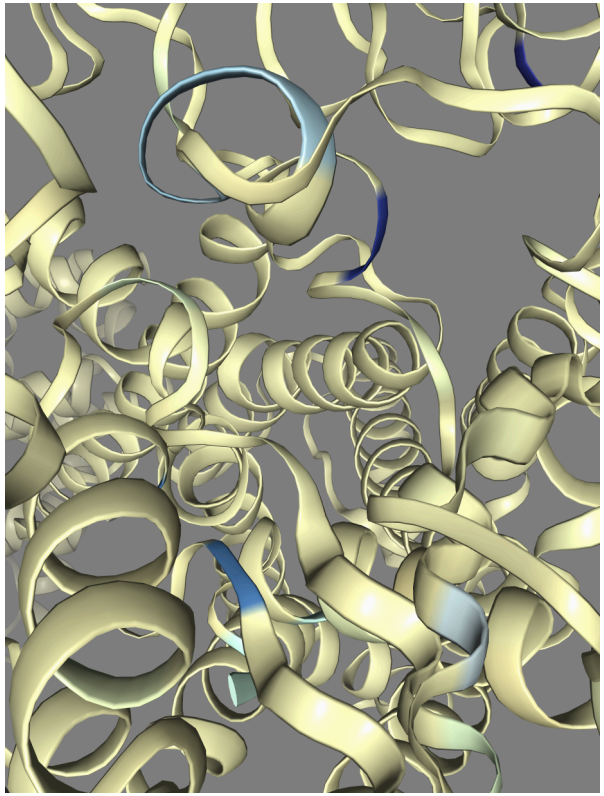
#3 Webserver

@Uni Leipzig

The mutation explorer

mutantX

- Mutations :: Rosetta design
  - User interface: „A:G13L“
  - Sequencing data
  - Alignments



https://proteinformatics.uni-leipzig.de/mutation_explorer_beta/mutantX

#4

Webserver / software

@IMMUTHERA GmbH

Predicting protein binding partners from sequence

- Unexpected binding may cause side-effects
- Scans entire human proteome in <1 minute (27223 sequences)
- 2 implementations (similar accuracy):
  - Random forest (scikit)
  - Graph neural networks (pytorch)

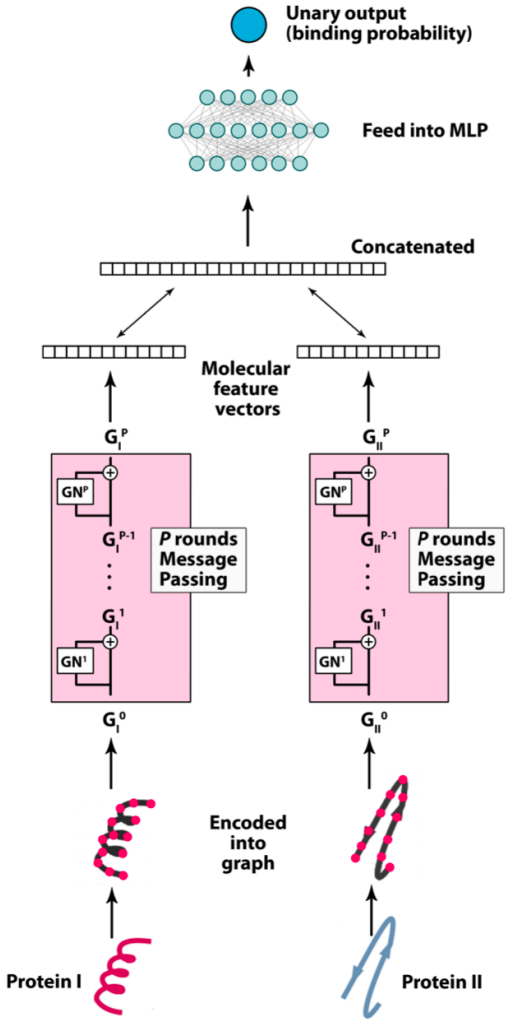
ProteinPrompt: a webserver for predicting protein-protein interactions
Sebastian Canzler, Markus Fischer, David Ulbricht, Nikola Ristic, Peter W. Hildebrand, René Staritzbichler
bioRxiv 2021.09.03.458859; doi: https://doi.org/10.1101/2021.09.03.458859

Random forest

- Rather simple approach, yet gold standard
- Ensemble decision trees with bootstrap aggregation
- Sequences must be transformed to feature vector of constant size ‚manually‘
    - Autocorrelation of 7 amino acid scales

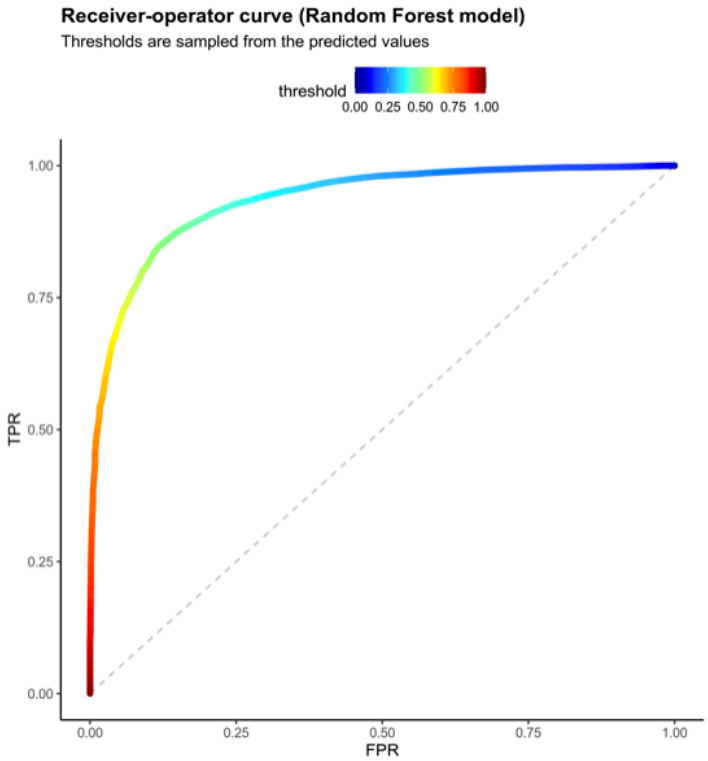
$$AC_{lag,j} = \frac{\frac{1}{n-lag}\sum_{i=1}^{n-lag}(S_{i,j} - \overline{S_j})(S_{i+lag,j} - \overline{S_j})}{\sigma_{S_j} * \sigma_{S_j}}$$

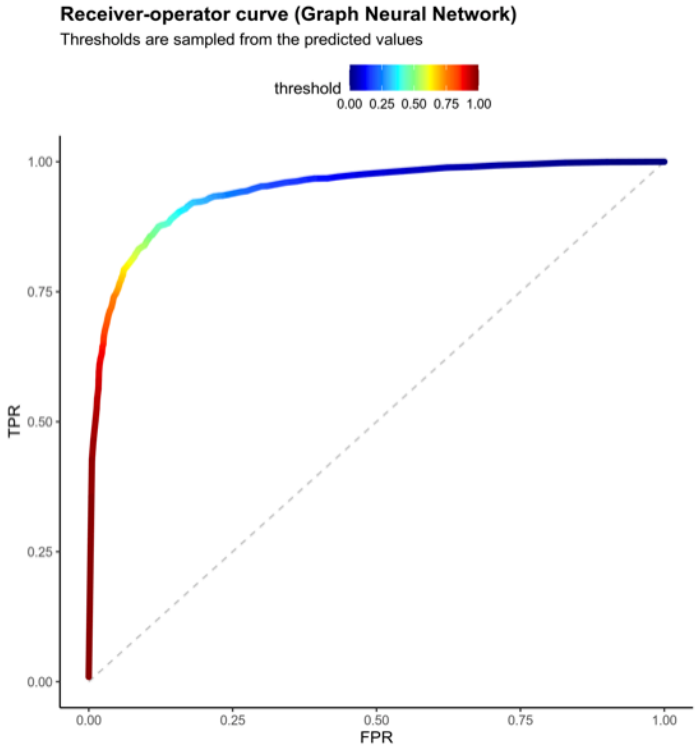
Graph neural network



- Can handle input data of varying size!
- Optimises feature vector transformation and prediction MLP commonly

Performance





Tool	AUC	Spec.	Sens.	Acc.
`ProteinPrompt`	0.94	0.88	0.84	0.86
`SPPS`	0.77	0.34	0.96	0.66
`TRI_tool`	0.73	0.95	0.32	0.63
`LR_PPI`	0.60	0.16	0.91	0.53

Plots were made with prettyPROC R-package

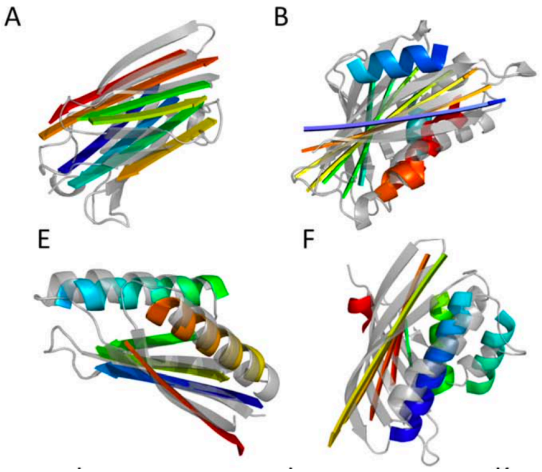
#5 Software

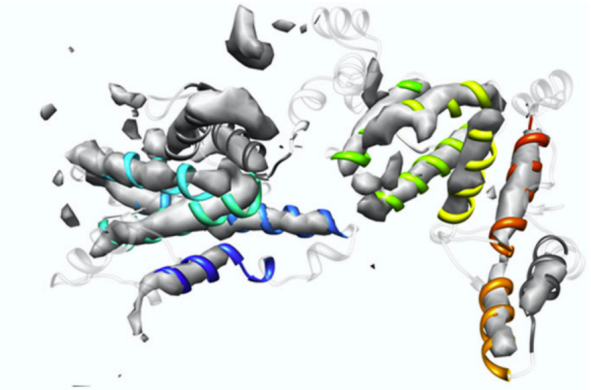
@Meilerlab

Structure elucidation

bcl::fold

bcl::em-fold



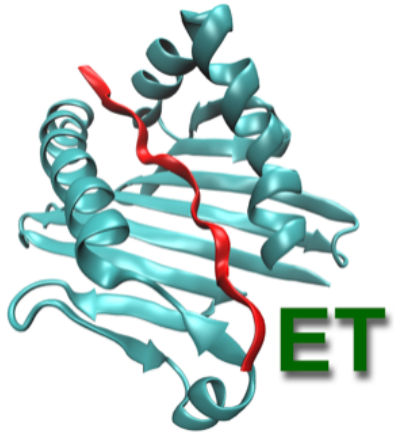


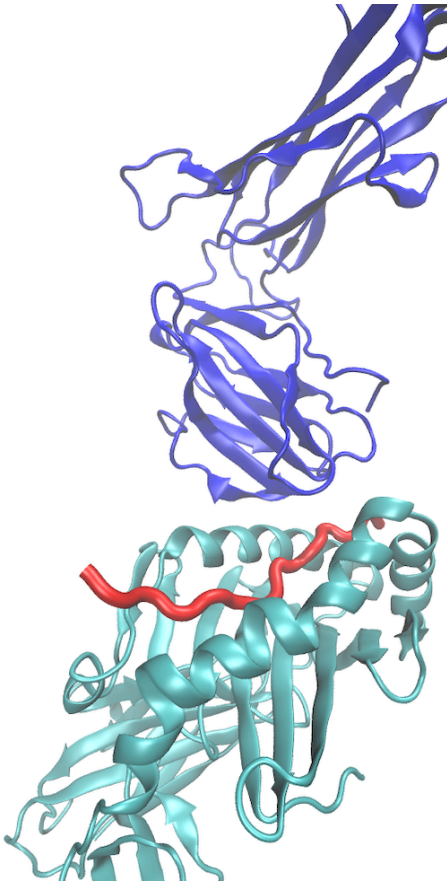
- Similar approach as Rosetta
  - Monte Carlo
  - Statistical potentials
- Secondary structure elements
- Overcoming Rosetta’s size limitation
- Fold into cryo EM density map

#6 Software

@BIONTECH

Predicting immune reactions





- Predicting vaccines
- Highly conserved structures
  - MHC: cyan
  - Epitope: red
  - TCR: blue
- Ideal for threading (extremely fast)

#7 **Software**

@Uni Leipzig

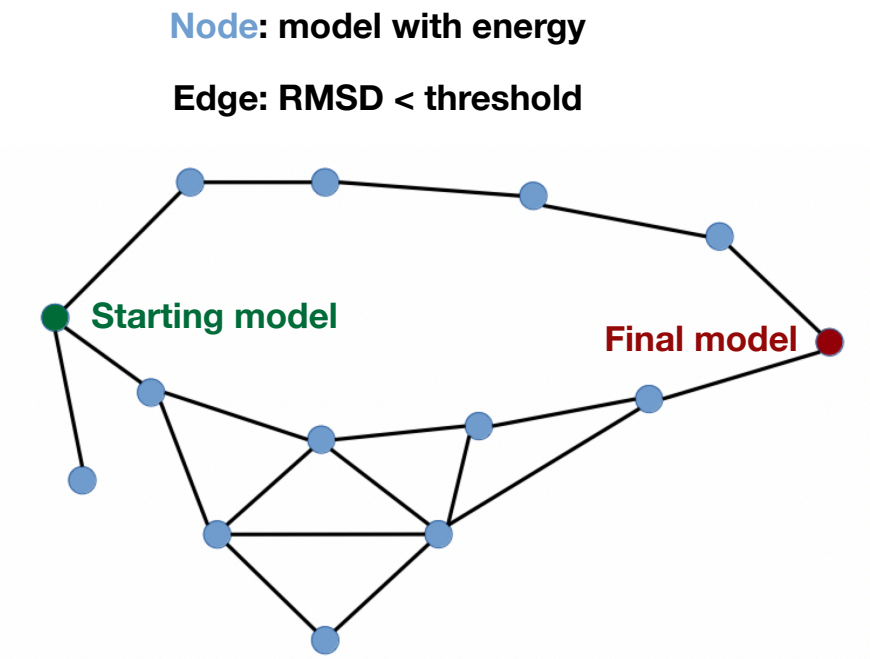
Transition pathways

- Realistic movies
- Simple transition pathway construction
    - Input: ensemble of models & energies
    - Output: pathway with lowest energy barrier (subset of ensemble)
- Enhanced sampling: start MD from each node of path
    - Link between MC and MD

Path construction

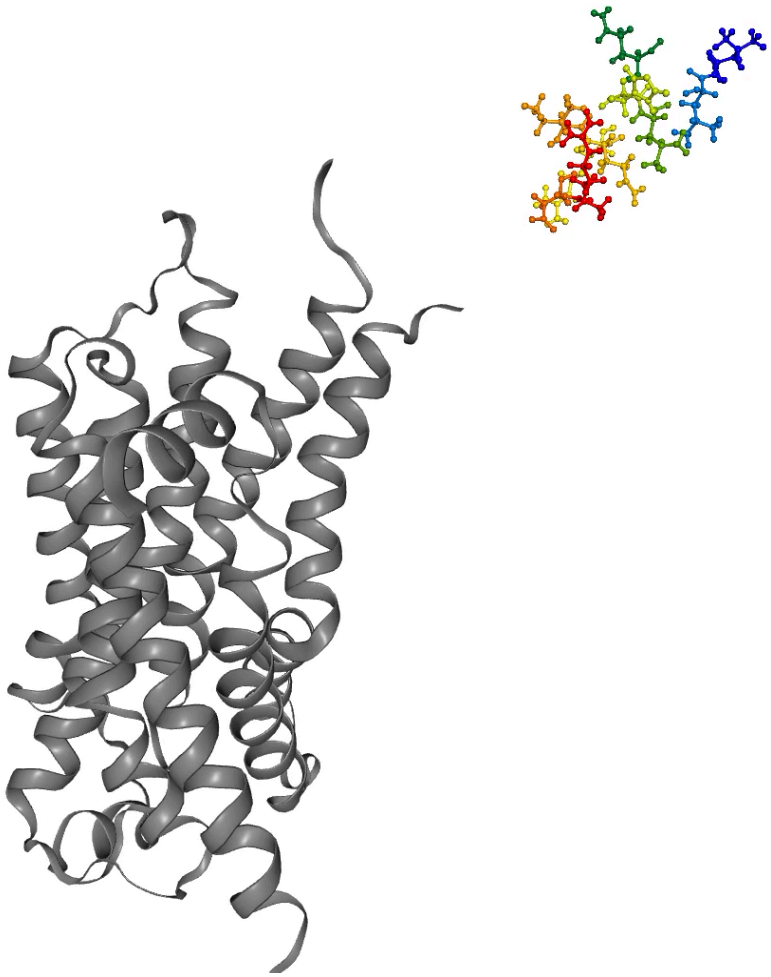
smoothT

- Forward graph construction
  - Find all pairs of similar models
- Backward path selection
  - Minimum energy barrier
  - Minimum energy integral



Live show

Path from MD

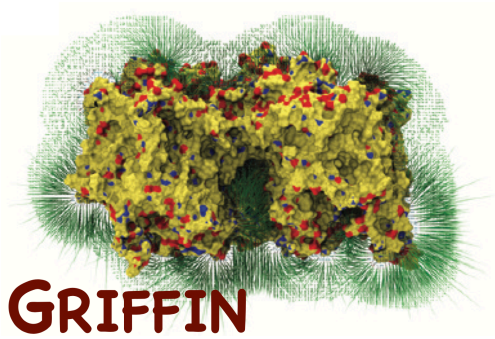


- MD spends most time with local fluctuations
- The movie summarises many long simulations

#8 Software

@MPI Biophysik

The implicit protein force field

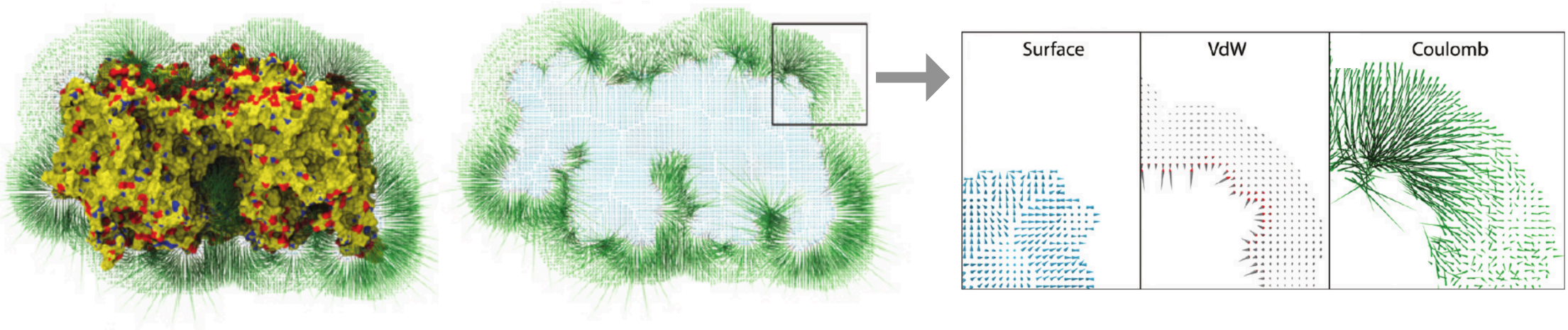


- Molecular dynamics is very accurate yet limited in sampling
- Griffin is precalulating a potential protein force field
  - Atomic detail (highly accurate)
  - Precalculated on grid (fast but static)

GRIFFIN: A versatile methodology for optimization of protein-lipid interfaces for membrane protein simulations.
Staritzbichler R, Anselmi C, Forrest LR, Faraldo-Gómez JD.
J Chem Theory Comput. 2011

Force grid

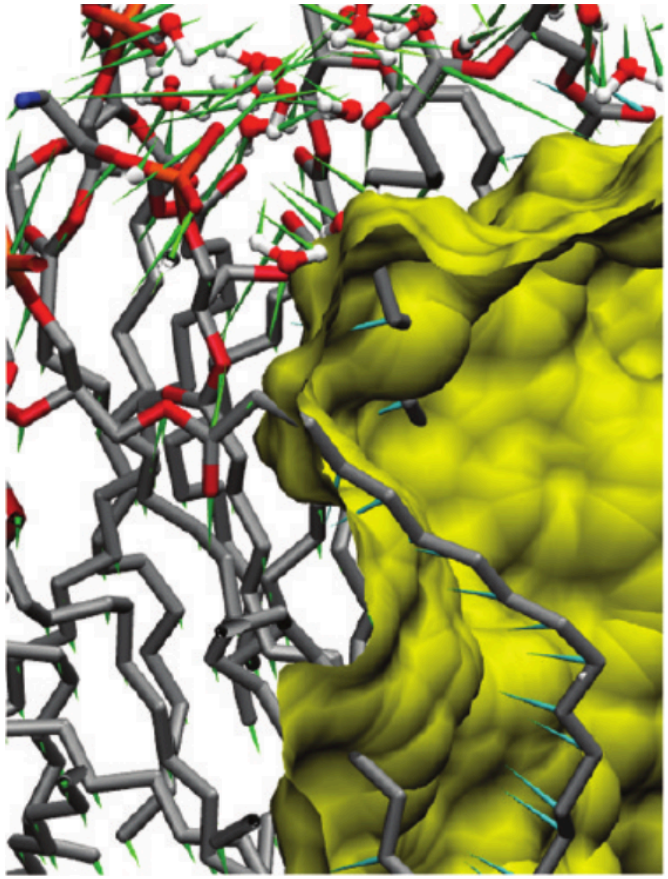
$$\vec{F} = q_{dummy} \cdot \sum_{i=1}^{atoms} \frac{1}{4\pi\epsilon_0} \frac{q_i}{r_{i,dummy}^2} \frac{\vec{r}}{r}$$



Precalculate forces acting on dummy atom located at grid point

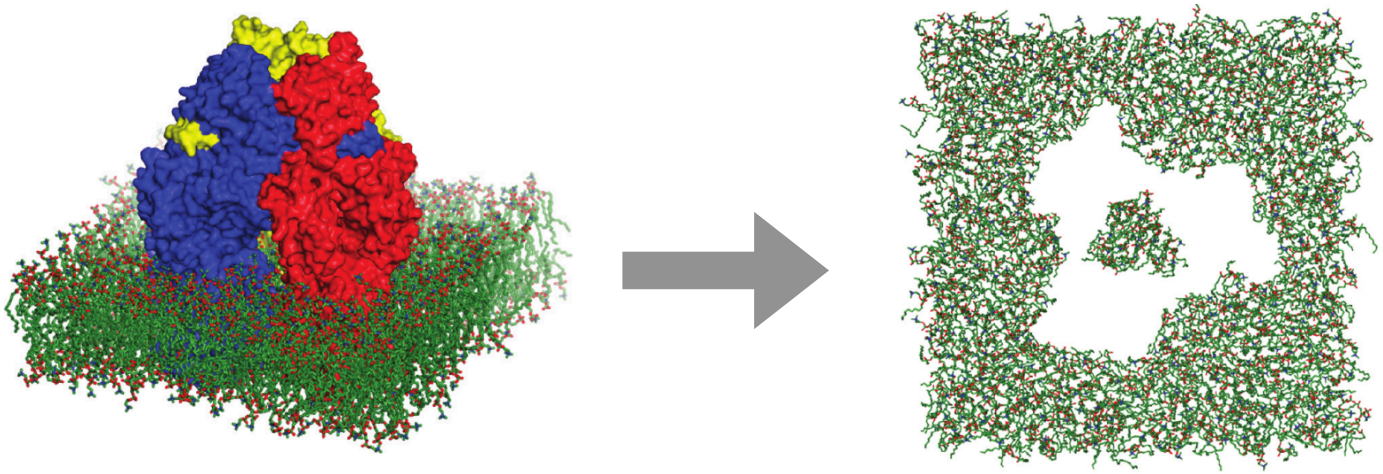
Resulting forces

- Inside molecule volume: surface force
- Outside: Coulomb and van der Waals

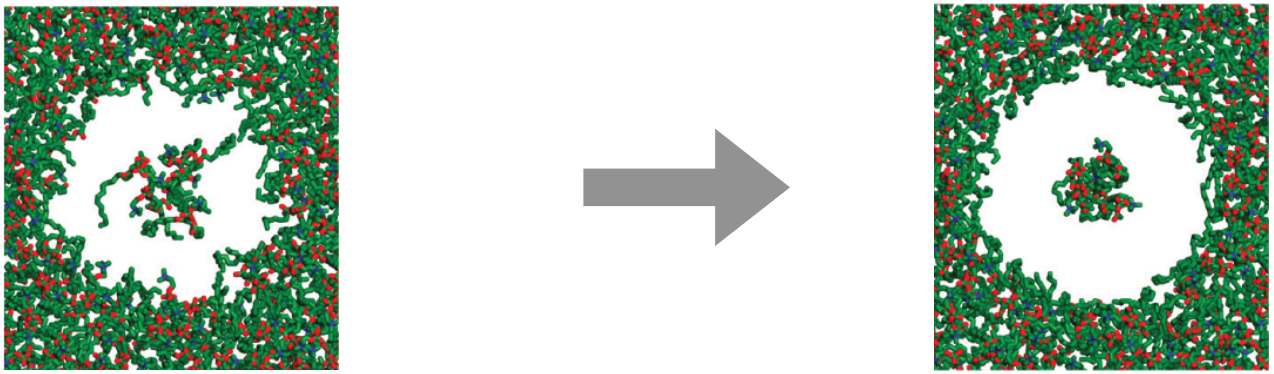


The ‚ghost‘ in action

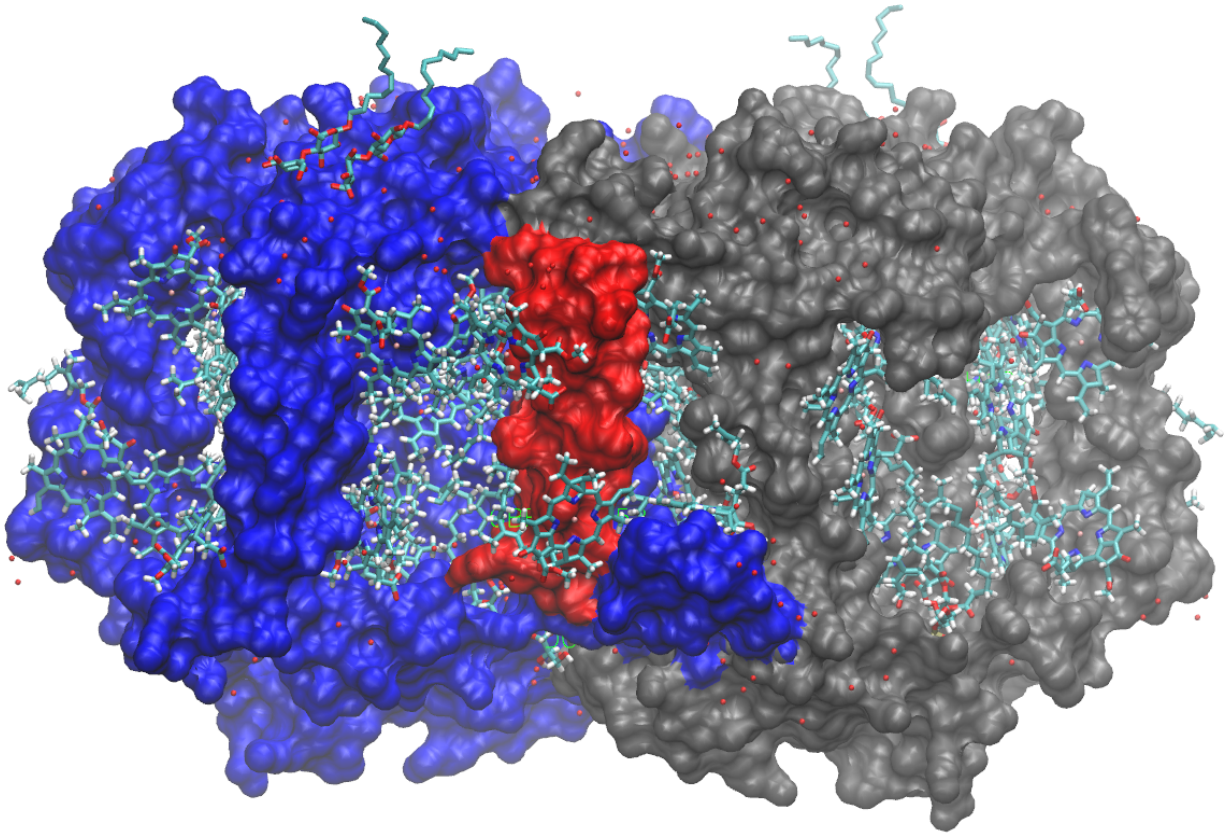
I:

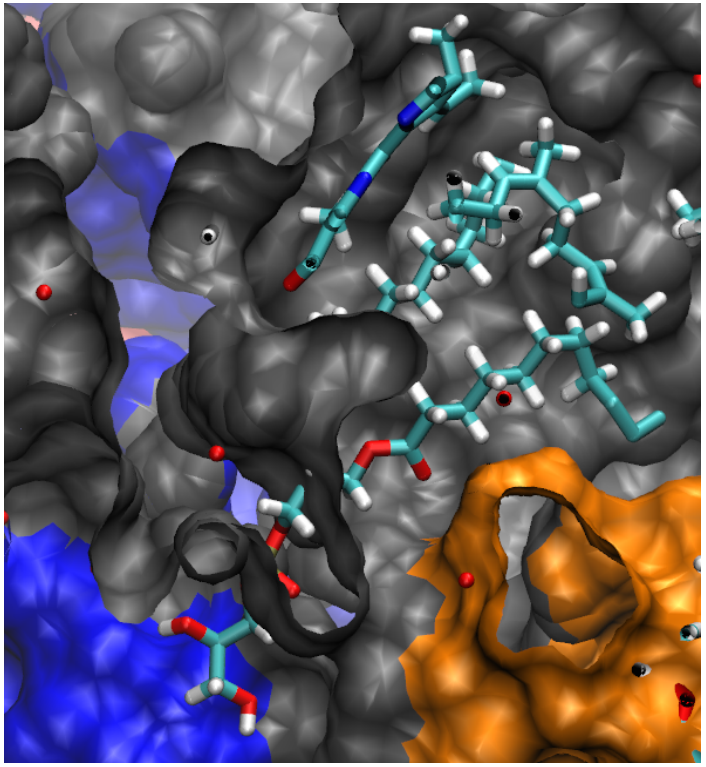


II:



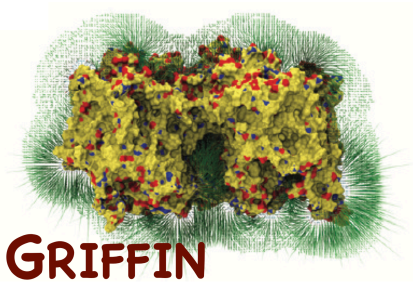
Bacterial photosystem





Modeling of complex topology with massive clashes.

C++ concepts



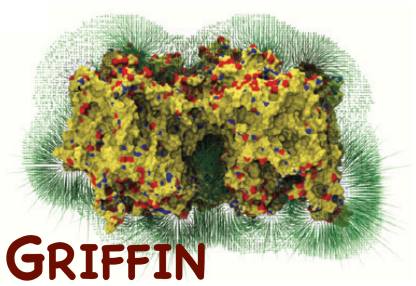
- Grid points: polymorphism
- Link to MD:
  - High RAM demand => fork daemon
  - Inter process communication (ipc.h)
  - Parallelization (mpi.h)

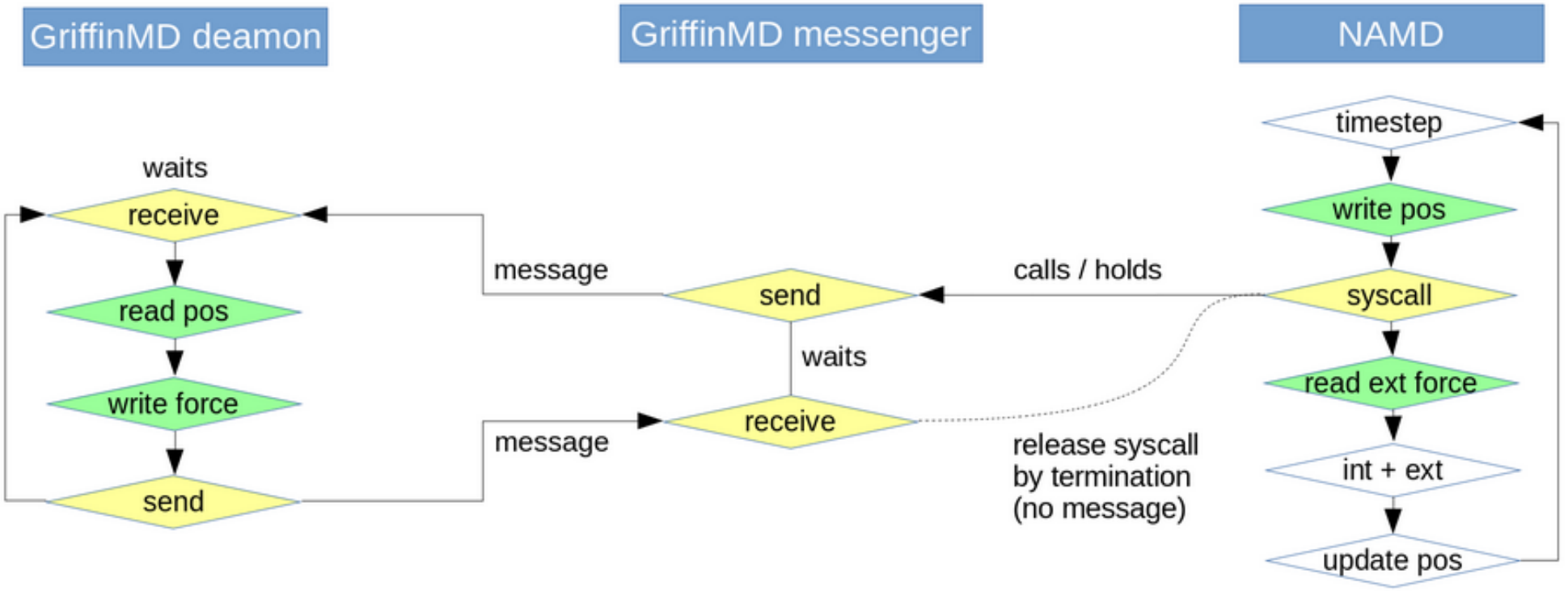
Molecule-atom force-field

```
shared_ptr<GridPoint> gp = new SurfaceForce(implicit_mols, pos);

vector3N force = gp->Force(explicit_atom);
```

Workflow

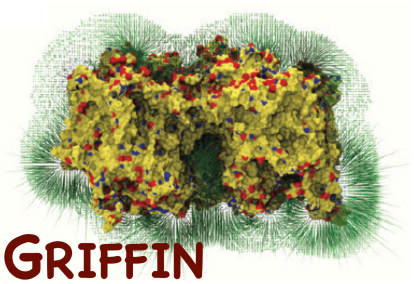


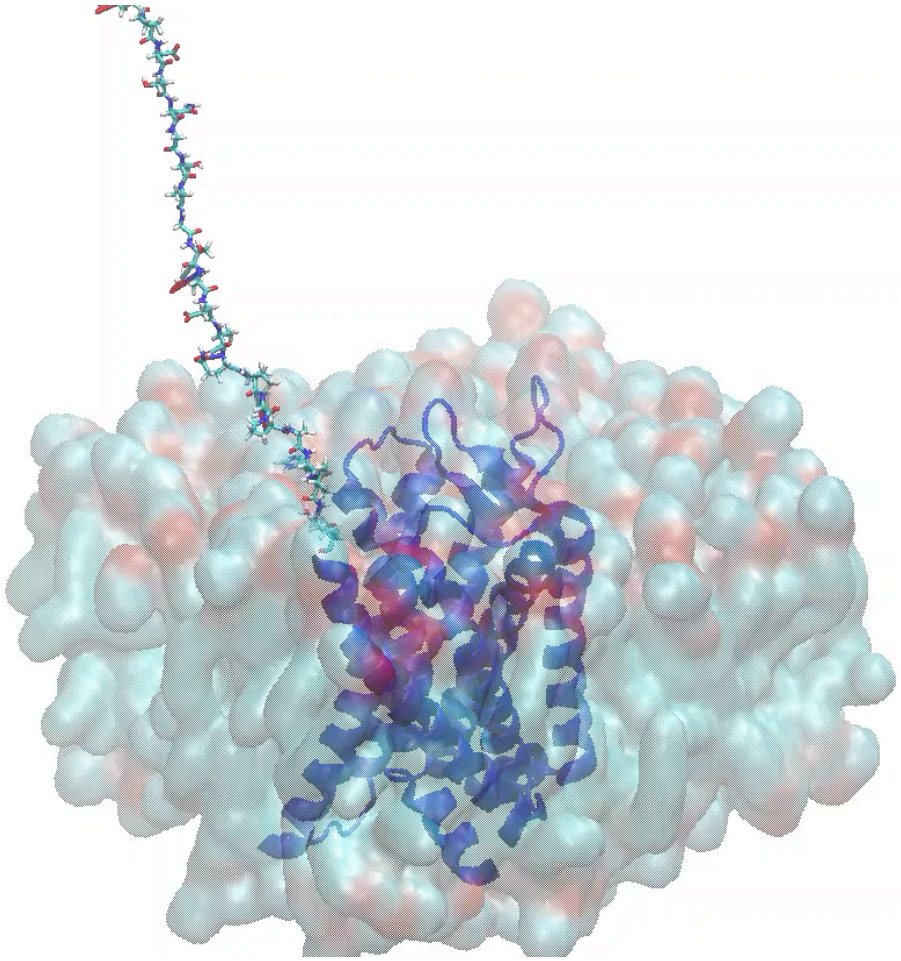


Implicit

Explicit

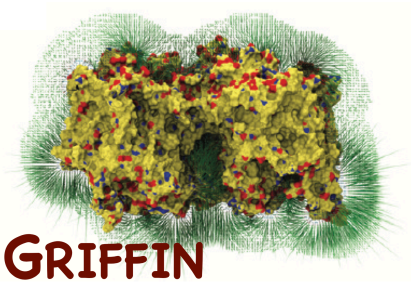
Explicit vs implicit

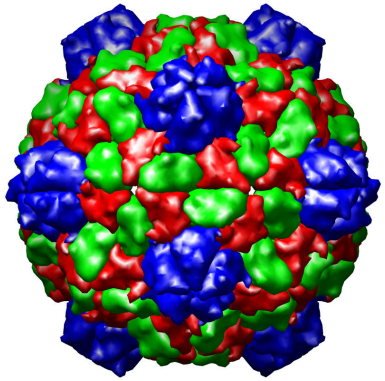


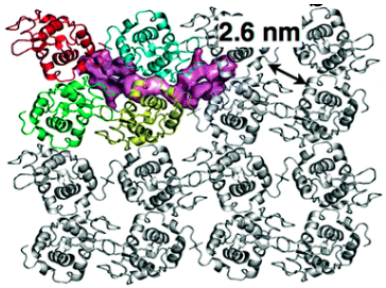


- Split
  - static: implicit
  - dynamic: explicit
- Flexible choice
  - even within one molecule

Griffin outlook







- Rigid body dynamics
  - Dynamics of large assemblies (e.g. viral capsids)
  - Protein-protein force-field (5D/6D grid)
  - SQL instead of RAM
- Monte Carlo
  - Binding of small molecules (Docking)

Acknowledgments

AlignMe

- Kamil Khafizov
- Markus Stamm
- Edoardo Sarti
- Lucy Forrest

Griffin

- Lucy Forrest
- José Faraldo-Gómez

ProteinPrompt

- Sebastian Canzler
- Markus Fischer
- Nikola Ristic
- Peter Hildebrand

EpitopeThreader

- Ugur Sahin

MDsrv

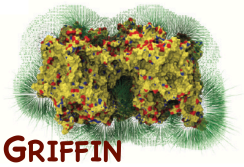
- Michelle Kampfrath
- Daniel Wiegreffe
- Guillermo Pérez Hernández
- Johanna Tiemann
- Alex Rose
- Gerik Scheuermann
- Peter Hildebrand

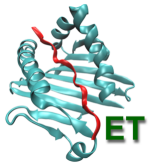
bcl::fold & bcl::em-fold

- Nils Wötzel
- Mert Karaks
- Jens Meiler

***'Lone wolf'* projects**

smoothT

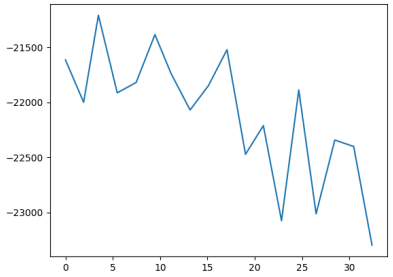


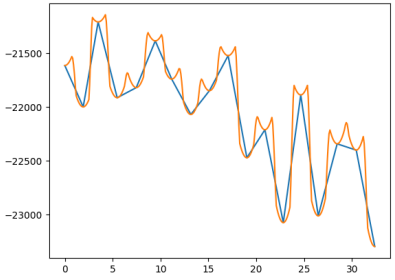
Epitope  Threader

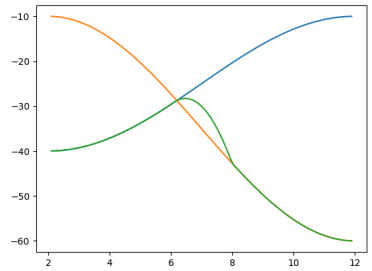
mutantX (Prototype)

Energy profile

- Path is basic energy profile (blue)
- Each model represents local minimum
- Minimum can be estimated by simple well
- Superimposing wells returns simple estimation of energy profile (orange)







SmoothT outlook

- Webserver
    - Upload an ensemble of models
    - Define a RMSD threshold, start and end point
    - Displays movies of pathways

AlignMe
ProteinPrompt
ProteinPrompt3D
Griffin
MD

MutantX
MDsrv

LearnIT

Raman
IR
VIS
MS
NMR
GenSeq

Clinical
- Diagnosis
- Characterization
- History

Graphical Report

Adverse effects

Marker Diagnosis