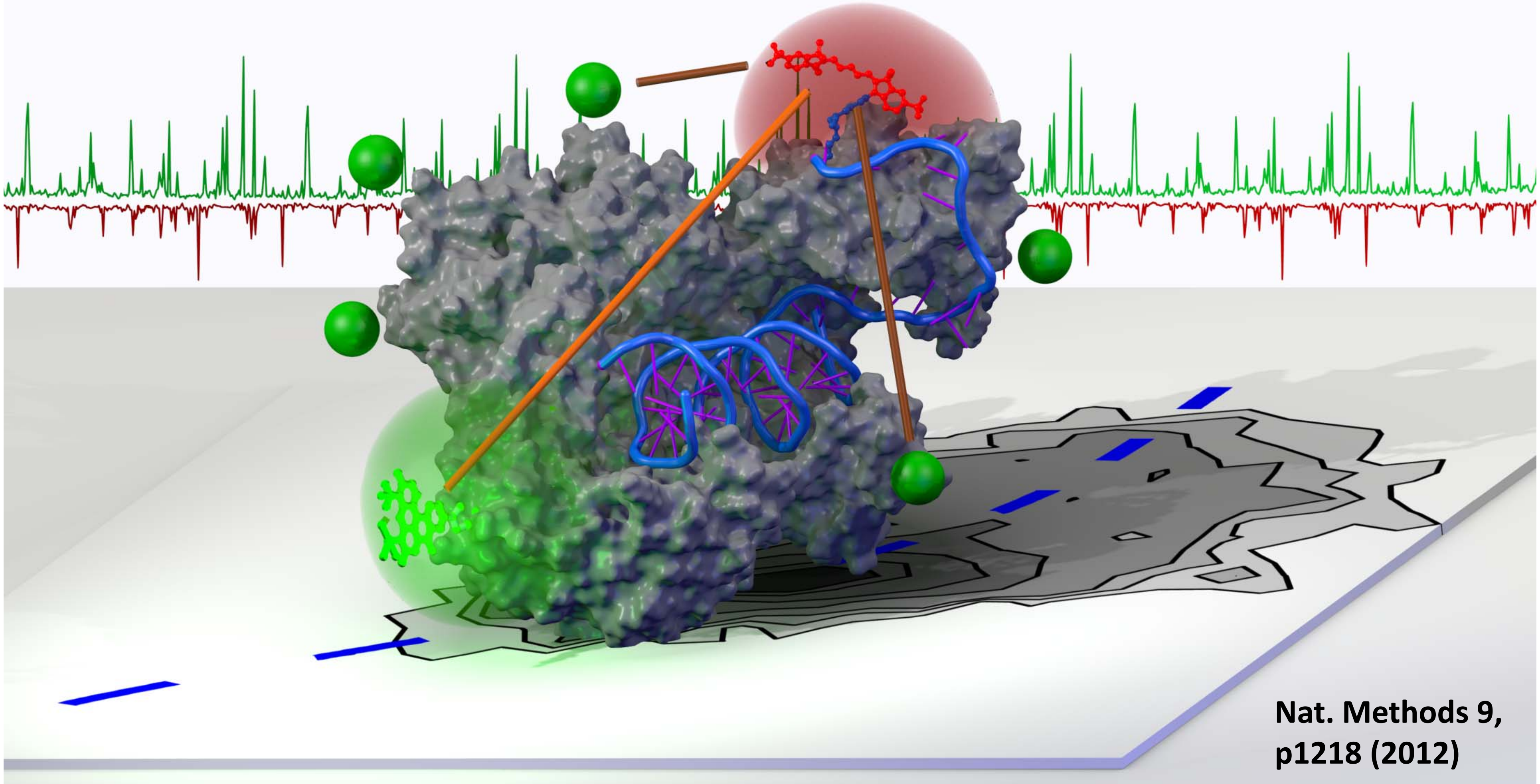
Integrative dynamic structural biology with fluorescence spectroscopy

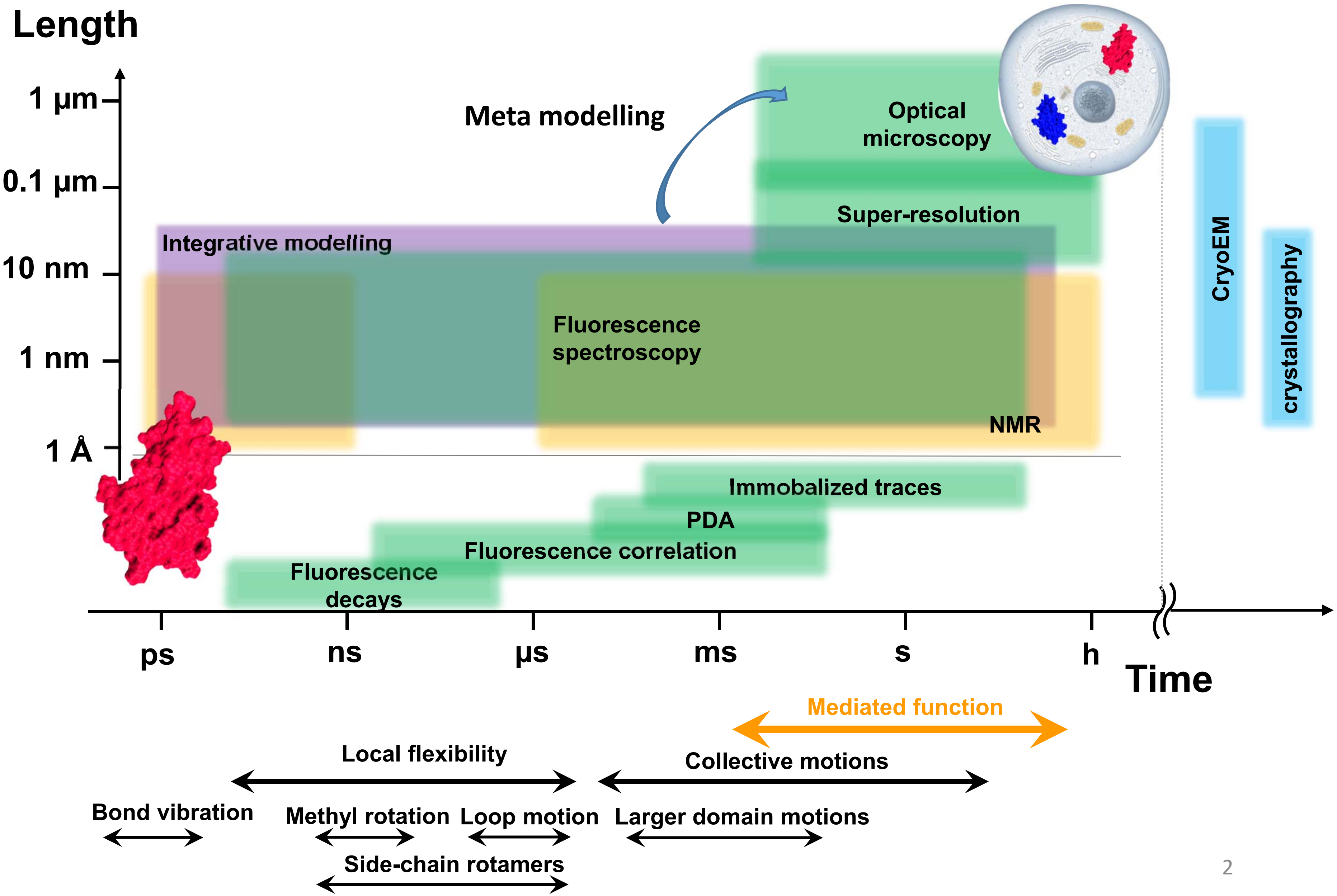
Claus A. M. Seidel

Chair for molecular physical chemistry, Heinrich-Heine-Universität Düsseldorf; cseidel@hhu.de



Nat. Methods 9, p1218 (2012)

Spatiotemporal models of living systmes

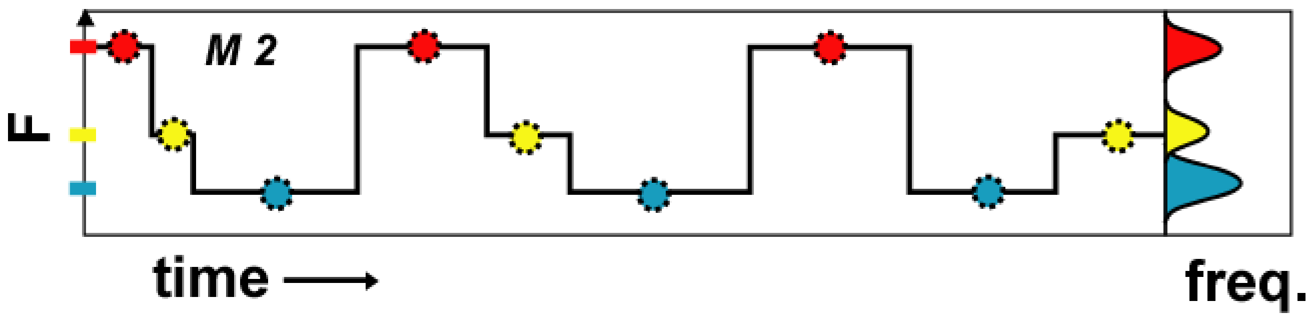


Fluorescence dimensions and protein information

Experiment types:

- single-molecule (variety of methods) to probe:
  - heterogeneity
  - structural information: proximities and distances
  - dynamics (from picoseconds to minutes): state connectivities

<u>Chemistry:</u>
Purification..., but dynamic changes in time.

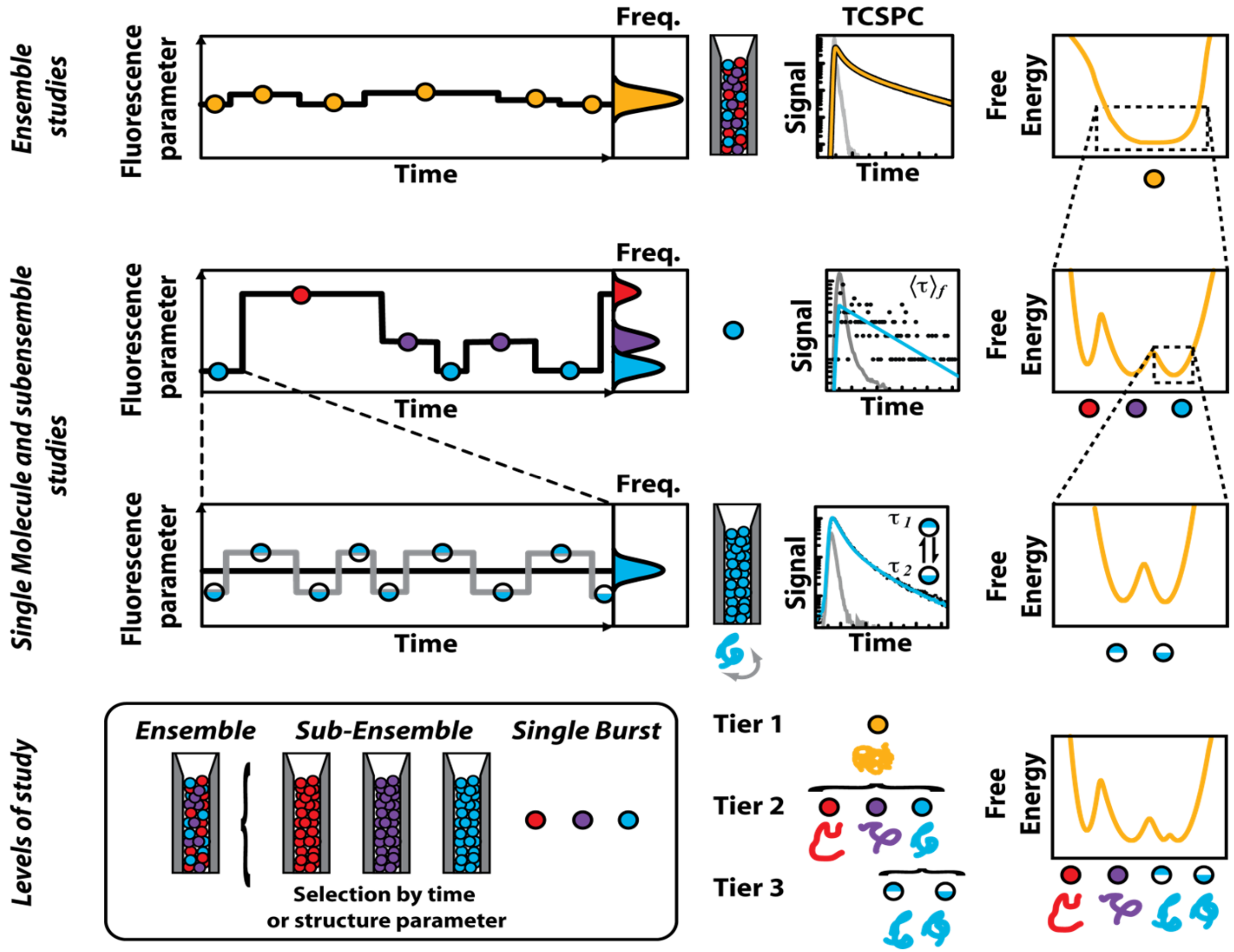


<u>Physics:</u>
Isolation in time via SMD.

Fluorescence dimensions and protein information

Experiment types (single photon counting (Poisson statistics!!):

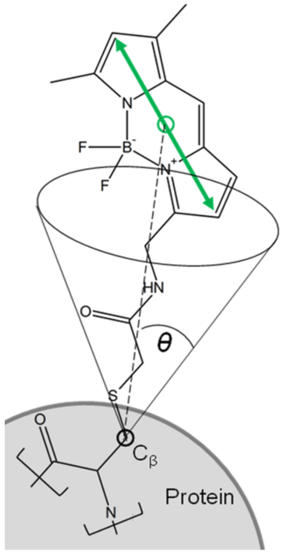
- single-molecule (variety of methods)
- sub-ensemble (like particle averaging in EM)
- ensemble (ultimate precision): 60 million photons

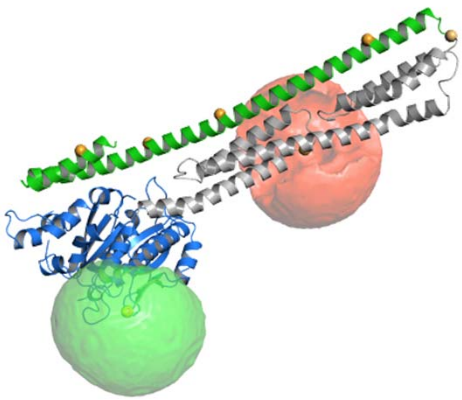


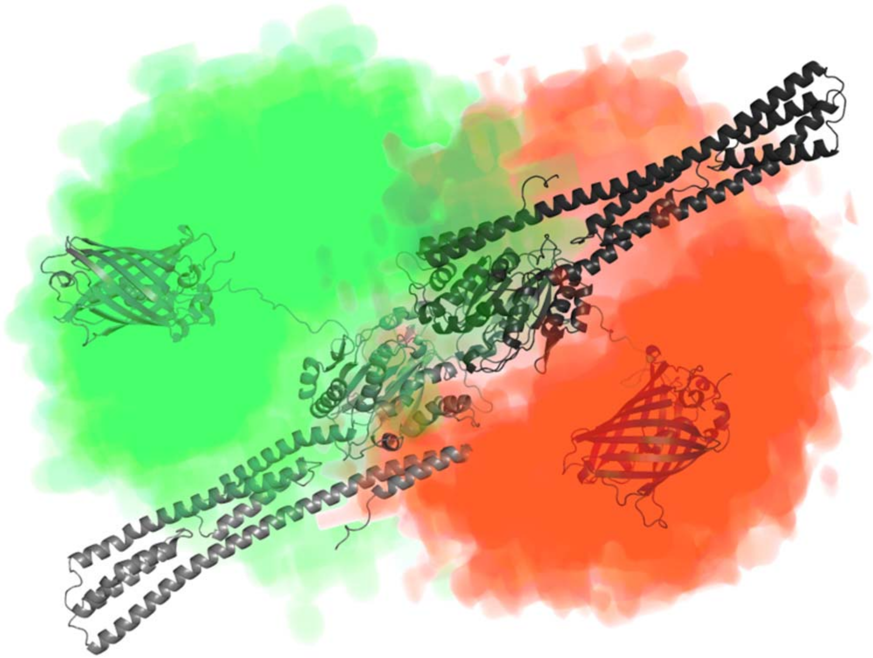
Fluorescence dimensions and protein information

Probes:

- Intrinsic, extrinsic, genetic:
  - aromatic amino acids (TRP, TYR)
  - selective labeling by small organic dyes,
  - fluorescent proteins







Consider linkage: Influence on spatial dye distribution

Fluorescence dimensions and protein information

- **Local structure**

Spectral changes (polarity probes, ratiometric probes)	Localization of the segment
Quenching	Proximity of other segments, accessibility of sites
Polarization, anisotropy	Order parameters (local flexibility) [1]

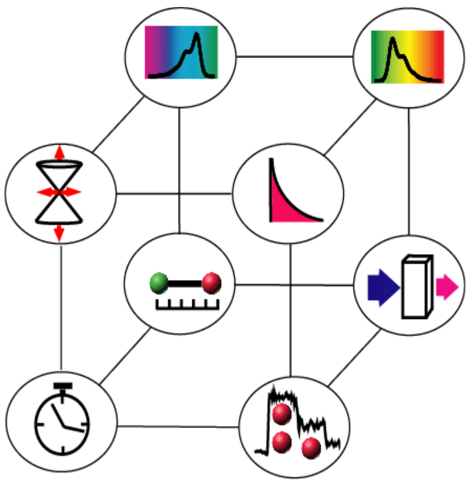
- **Global structure**

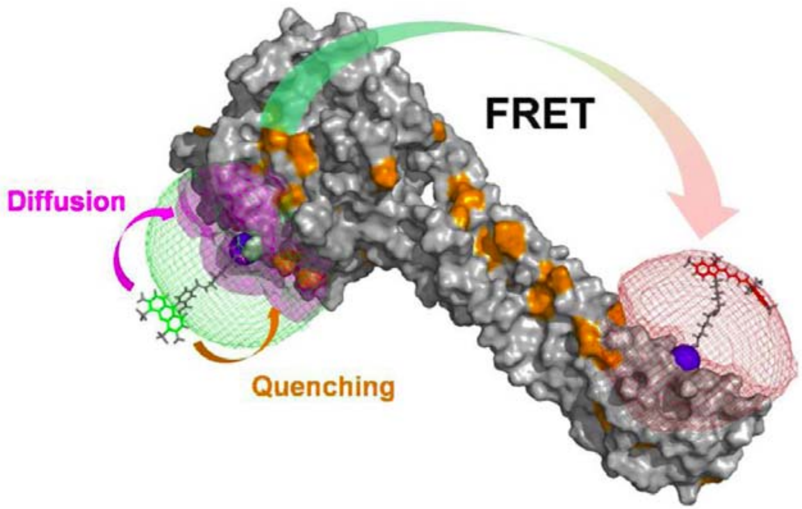
Förster energy transfer (FRET)	Inter-dye distances [2]
Diffusion (translation and rotation)	Shape

- **Dynamic exchange**

Time-resolved detection (ps-ms/ min)	Kinetic exchange networks

A N-dimensional vector with all observables characterizes each protein state





[1] Möckel et. al. J. Phys. Chem. B 122, in press (2018)

[2] Peulen et. al. J. Phys. Chem. B 121, 8211 (2017)

Quantitative high-precision FRET (hpFRET) analysis

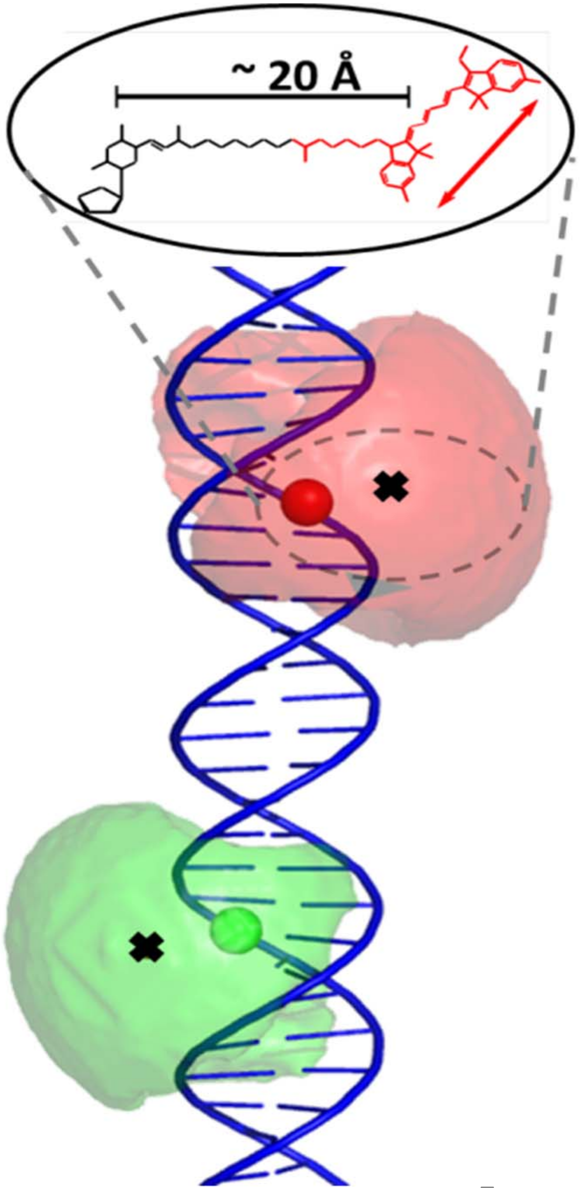
- **Multiparameter Fluorescence Detection (MFD)**
- **Detect dynamic averaging**
- **Rigorous error analysis**
- Dye properties: **Accessible volume (AV) simulations**
- Hybrid approach: **Combination with computer simulations**

Structural models in data bases (wwPDB)

- Precision
- Accuracy
- Confidence levels

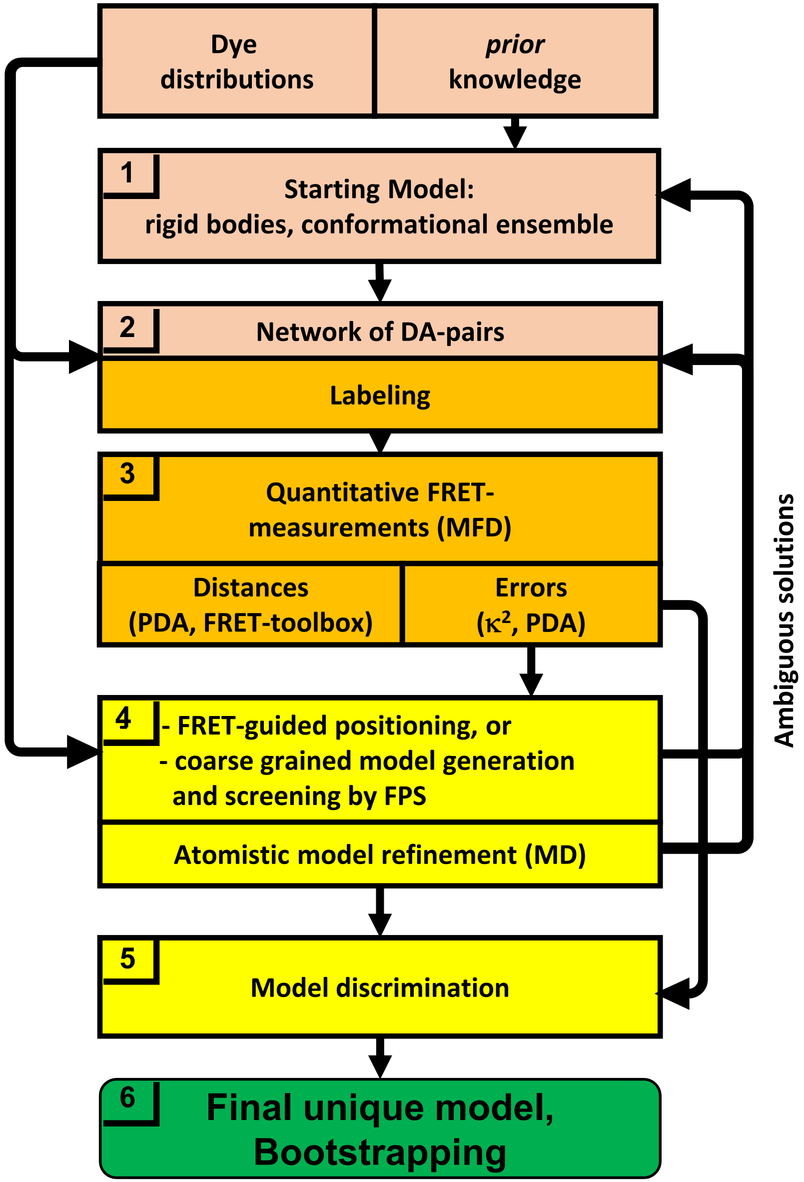
- **Dynamic / heterogeneous systems**:
  Use Single-molecule advantage:
  - multiple structures in parallel
  - kinetics: no need for sychronization

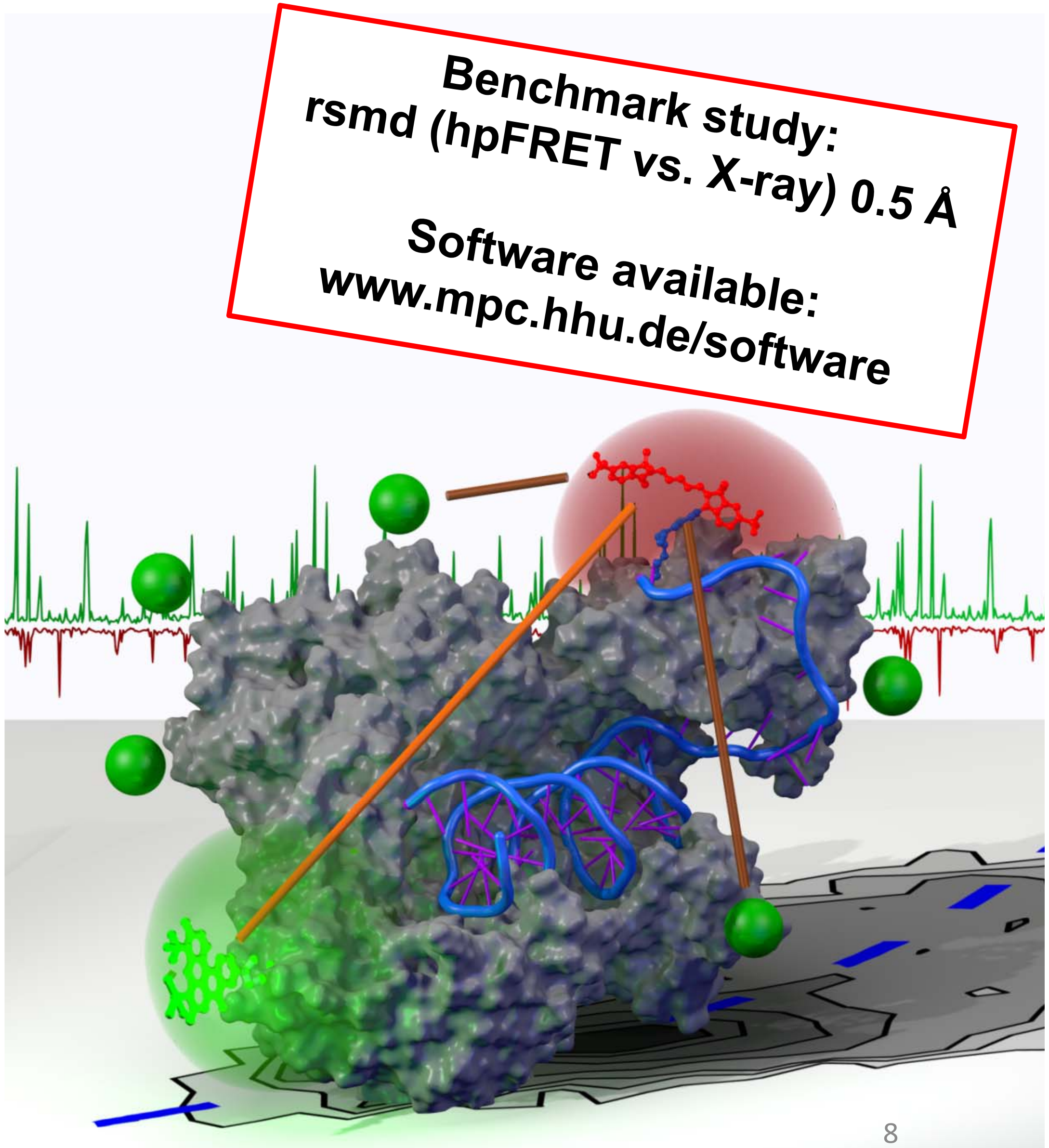
- Kinetics: reaction pathways
- Solution conditions, room temperature
- Large systems possible
- Combination with microscopy: *in vivo* option



Nat. Methods 9, p1218 (2012)

High-precision FRET (hpFRET):
6 steps to FRET-restrained structural biology

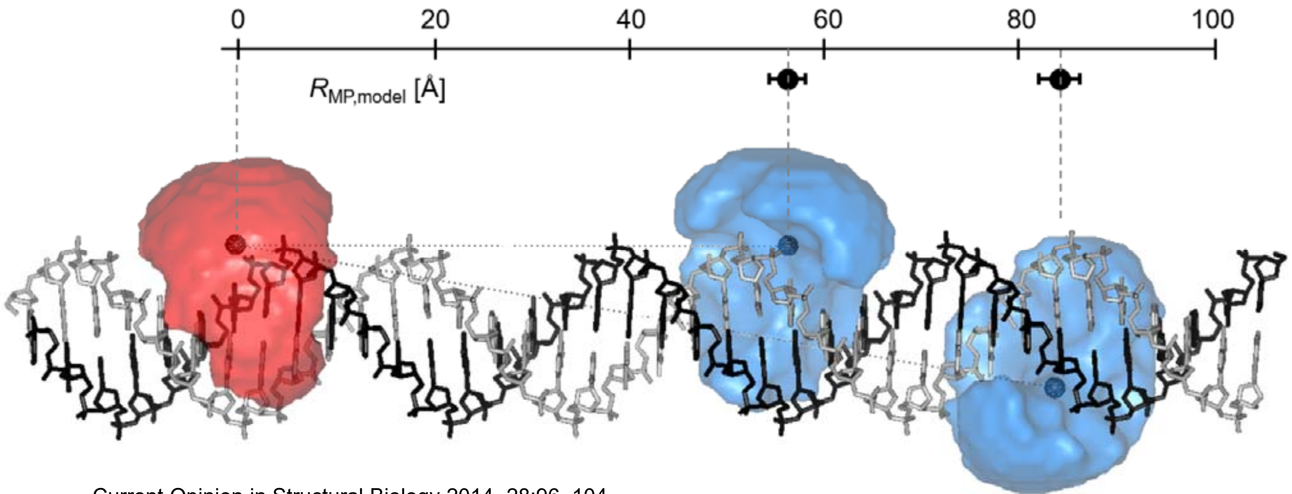




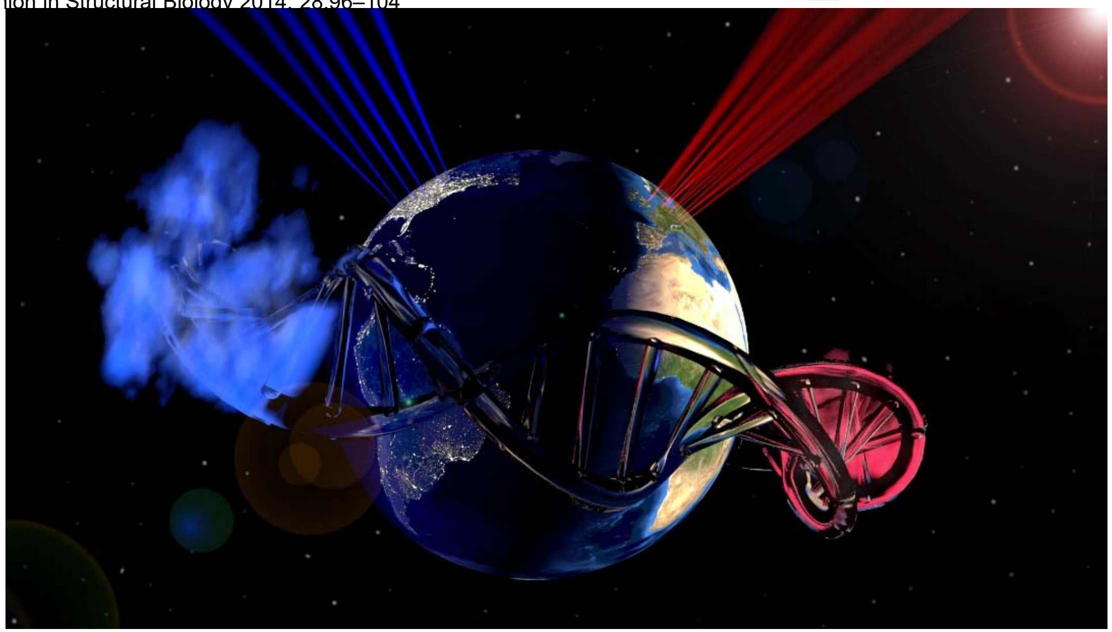
Nat. Methods 9, p1218 (2012)

Multi-laboratory challenge demonstrates the accuracy and precision of FRET

Blind study:
10 distinct DNA rulers were measured in 20 laboratories in the world



Current Opinion in Structural Biology 2014, 28:96–104



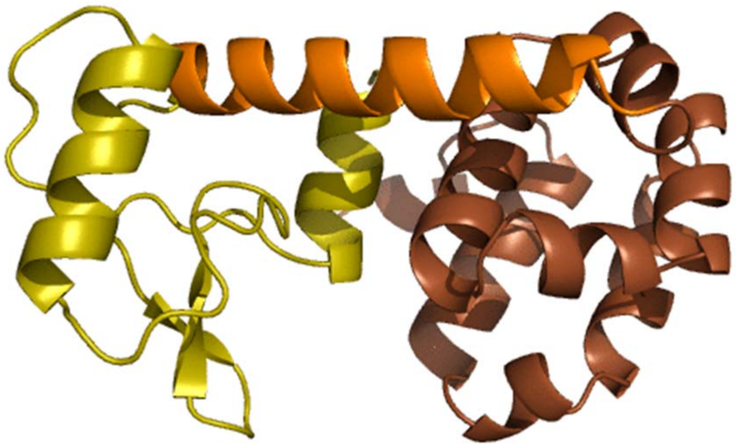
Standard deviation between the exp. FRET efficiencies:
$\Delta E = \pm 0.02$ and ± 0.05

Rel. deviations between mean exp. and model distances $\langle \boldsymbol{R} \rangle$:
$\Delta R/R_{mod} = 0$ and ± 0.05

→ well within the expected error

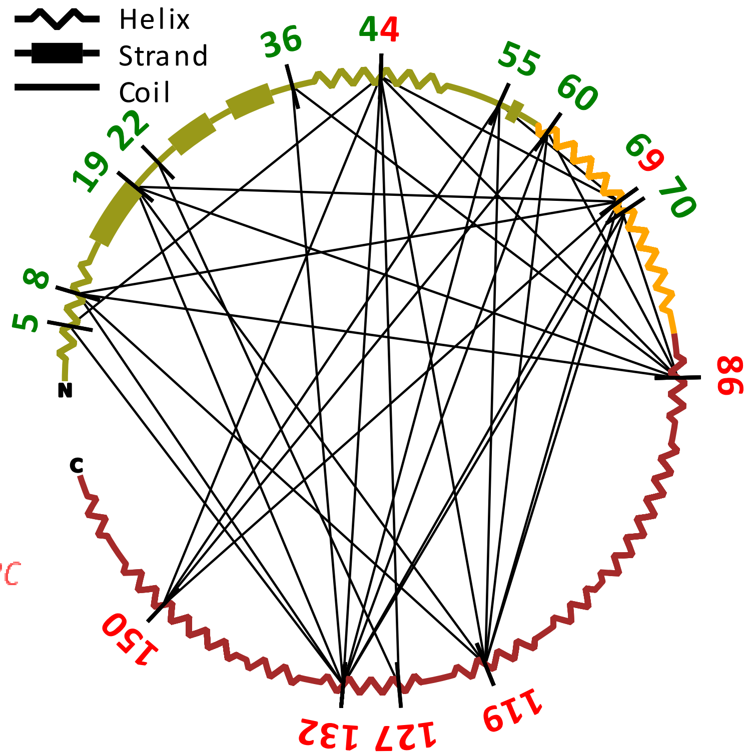
Nature Meth. **15**, 669 (2018) (coord. T. Hugel, T. Craggs, C. Seidel, J. Michaelis)

Proteins with more than one conformation



New challenge for CASP community:

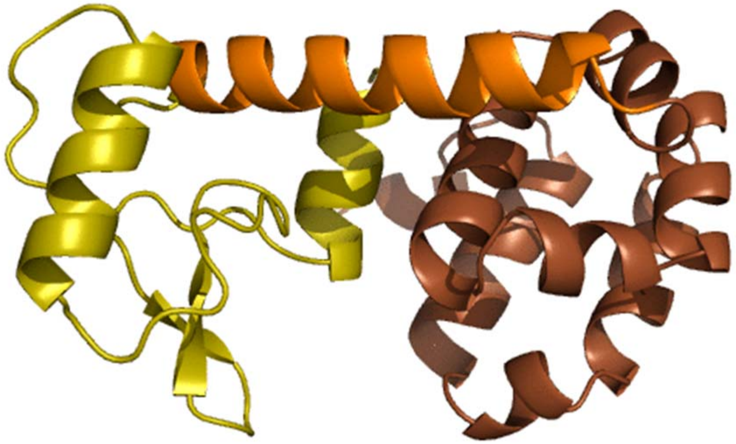
- Predict and describe proteins in distinct conformations
- Data-assisted modelling using fluorescence information



FRET measurements

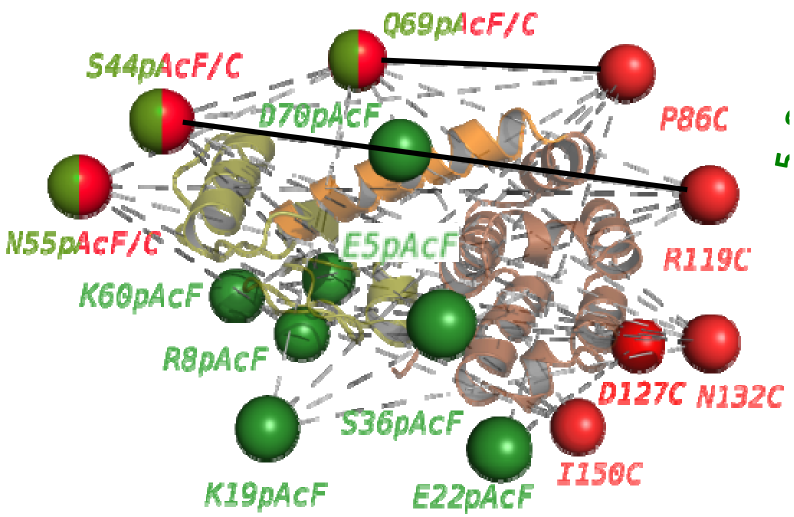
- Sparse data- combination with computer simulations
- Euclidean distance information between two points with upper and lower bound

Proteins with more than one conformation



New challenge for CASP community:

- Predict and describe proteins in distinct conformations
- Data-assisted modelling using fluorescence information

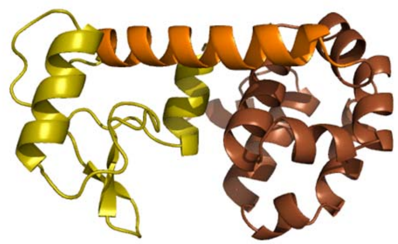


FRET measurements

- Sparse data- combination with computer simulations
- Euclidean distance information between two points with upper and lower bound
- Dynamics

Multiple structures and complexes: characterized on the fly

Systems | **States** | **Relaxation times**



<u>Proteins with 2 domains</u>

SNARE Protein Syntaxin 1 2 (700 μs)
PGK >2 (fast)
T4 Lysozyme >3 (3, 200 μs)
IF3 (ribosomal initiation factor) >3 (7) (2 μs, 300 μs)

<u>NA binding proteins</u>

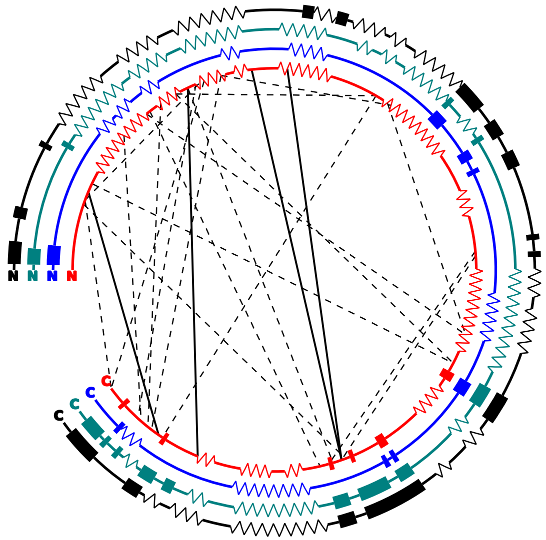
HIV-1 RT Primer/Template complexes: 3 (slow)
KlenTaq Polymerase >3(slow)
Klenow Fragment DNA Pol1 >5(< 100 μs)
BsoBI (restriction enzyme) 2-3 (10 μs)
MutS (mismatch repair) 2(slow)
Mononucleosomes >3(slow)

<u>Large GTPases</u>

hGBP1 >2 (2, 20, 300 μs)
Holliday junction (DNA 4W-Junction) 2(4) (100 μs)
Hairpin ribozyme (RNA 4W-Junction) 4 (100 μs)

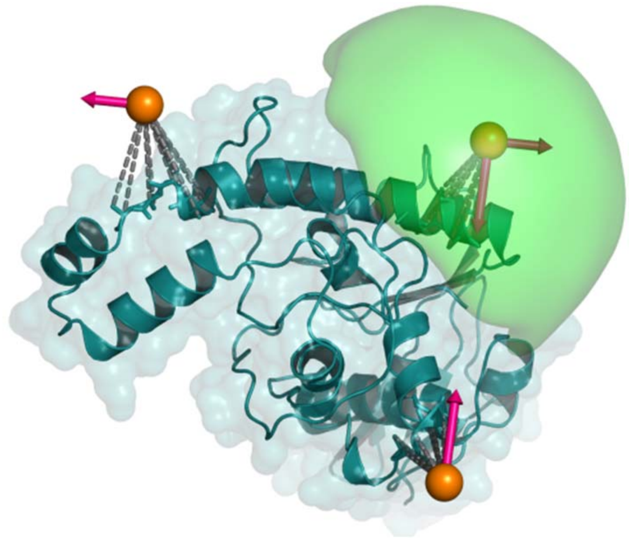
<u>Intrinsically disordered proteins</u> **..........**

Next generation of FRET toolkit: FPS 2.0 (tbd @ Poster)



1. **Finding most informative FRET-pairs**:
   FRET networks for least experimental work

2. **FRET-restrained optimization of structural models:**
   Beyond rigid body docking and simple model selection
   -> Targeted structural sampling by FRET guiding

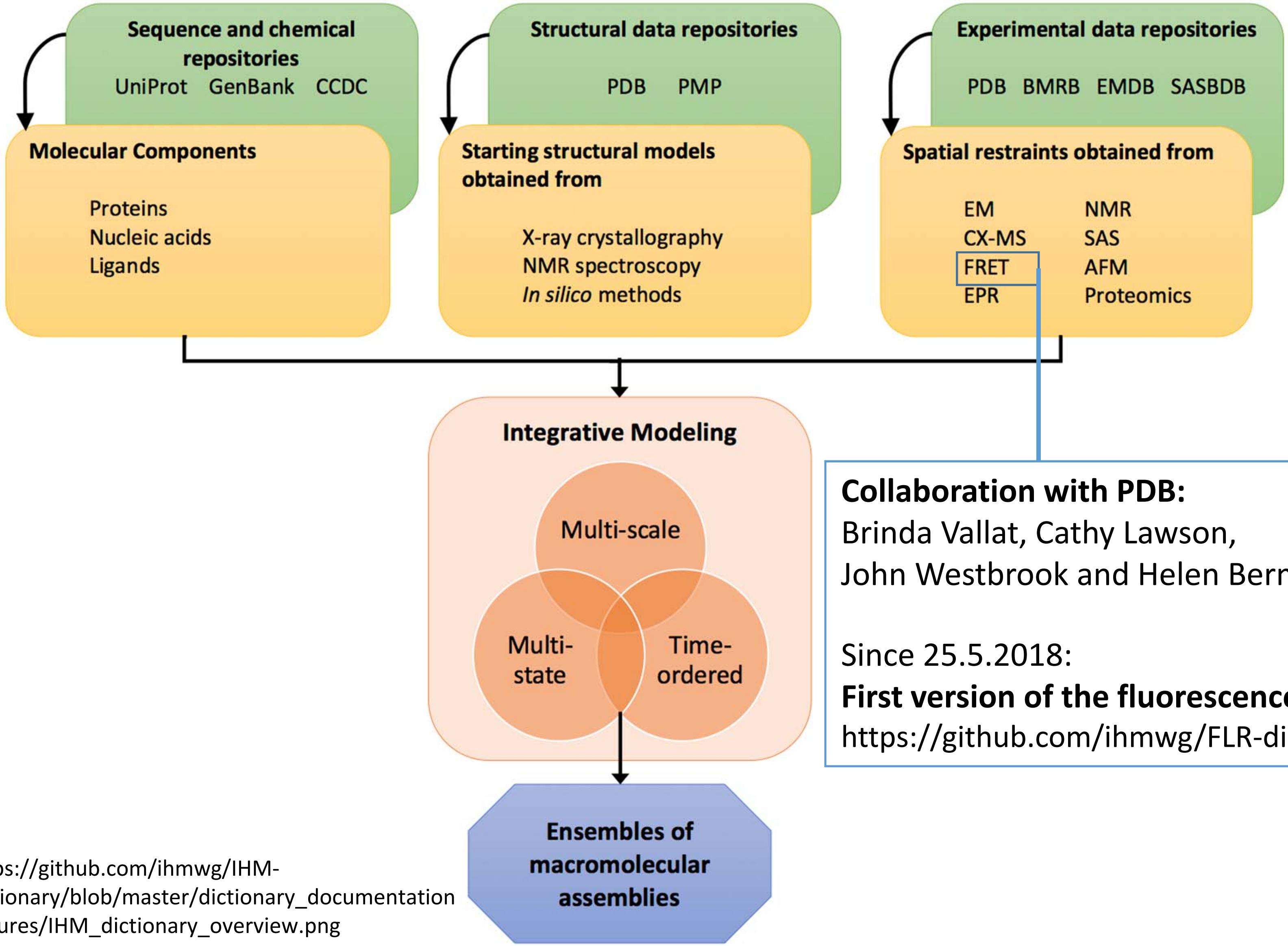


3. **Estimation of accuracy:**
   Can we trust the FRET-restrained structural model?
   -> **a crossvalidation approach analog to X-ray:**
   R_{free} and

4. **Archiving of the I/H models to share information:**
   Generation of documentation: Fluorescence dictionary

-> deposition in PDB-Dev

PDB-Dev: New archive of structural models obtained through integrative/hybrid (I/H) methods

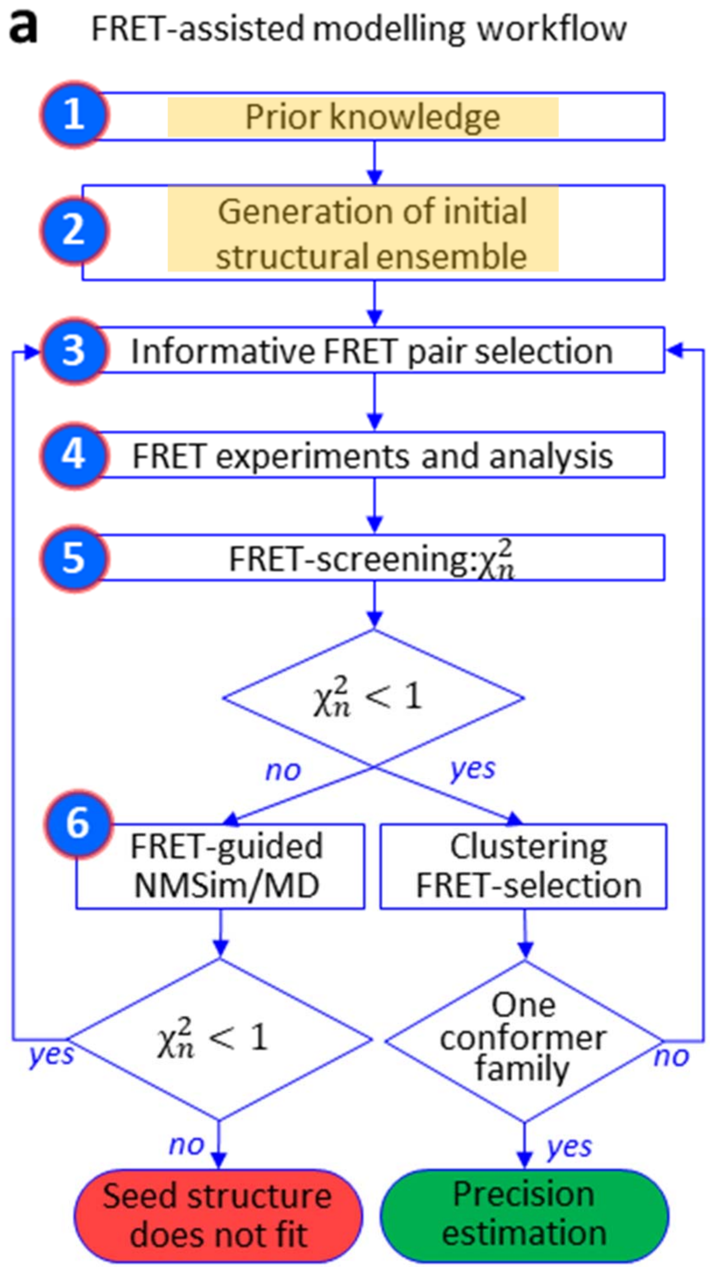


Collaboration with PDB:
Brinda Vallat, Cathy Lawson,
John Westbrook and Helen Berman.

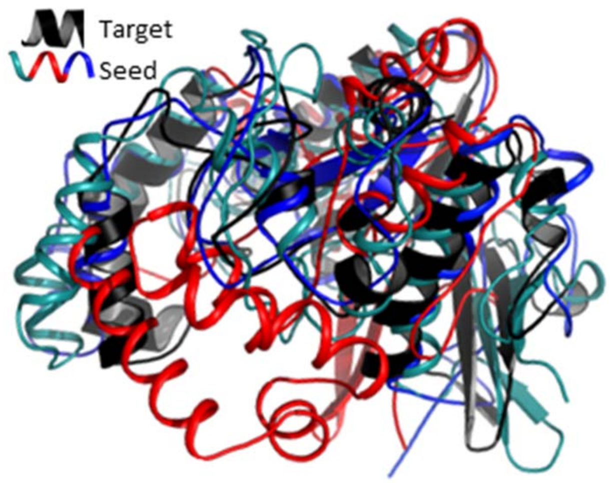
Since 25.5.2018:
First version of the fluorescence dictionary
https://github.com/ihmwg/FLR-dictionary

https://github.com/ihmwg/IHM-dictionary/blob/master/dictionary_documentation/figures/IHM_dictionary_overview.png

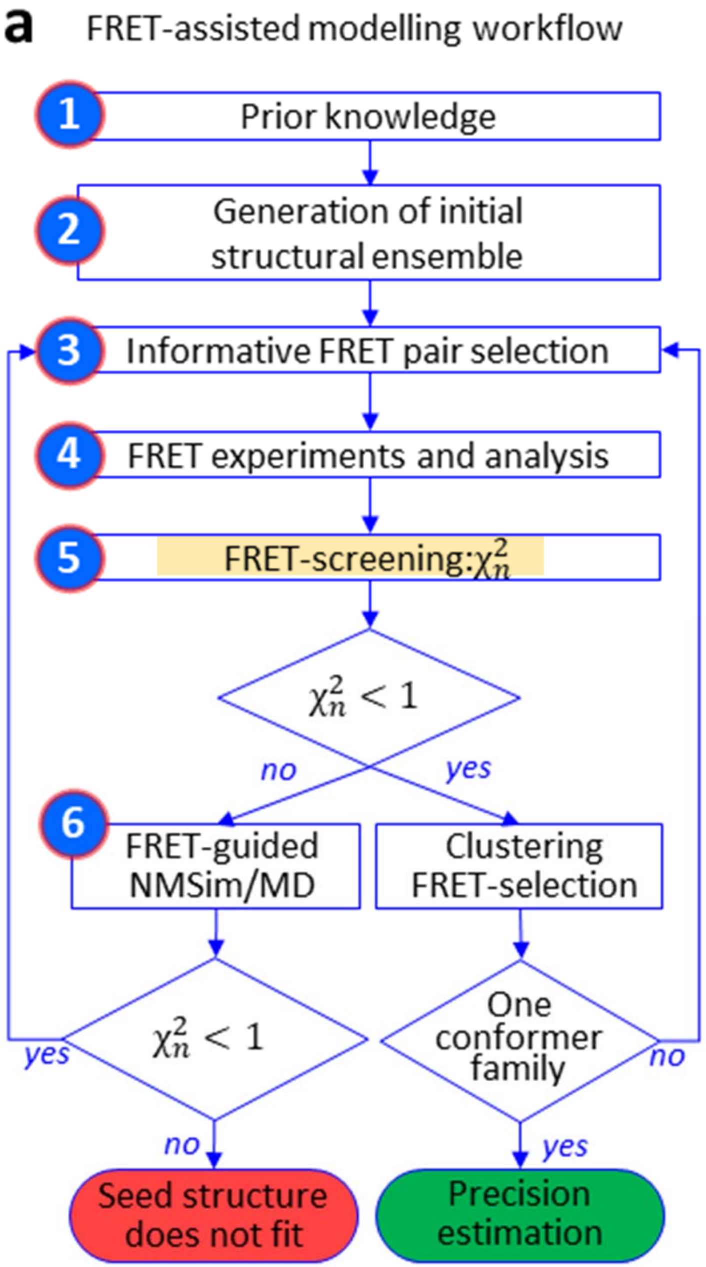
Selection of the most informative FRET pairs



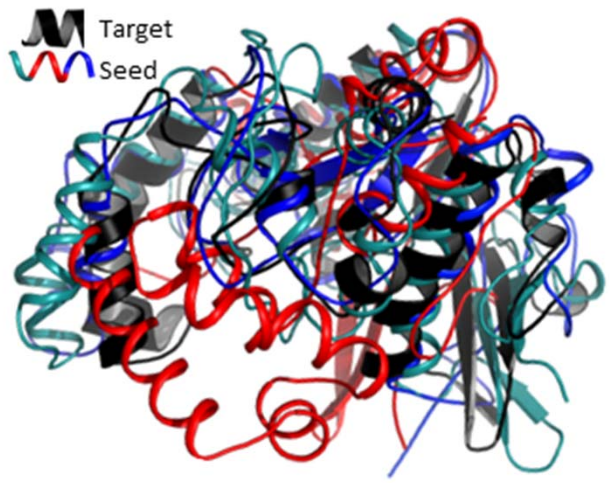
CASP11 target T0806: YaaA (PDB ID 5CAJ)



Probing the accuracy of FRET-restrained structural models



CASP11 target T0806: YaaA (PDB ID 5CAJ)

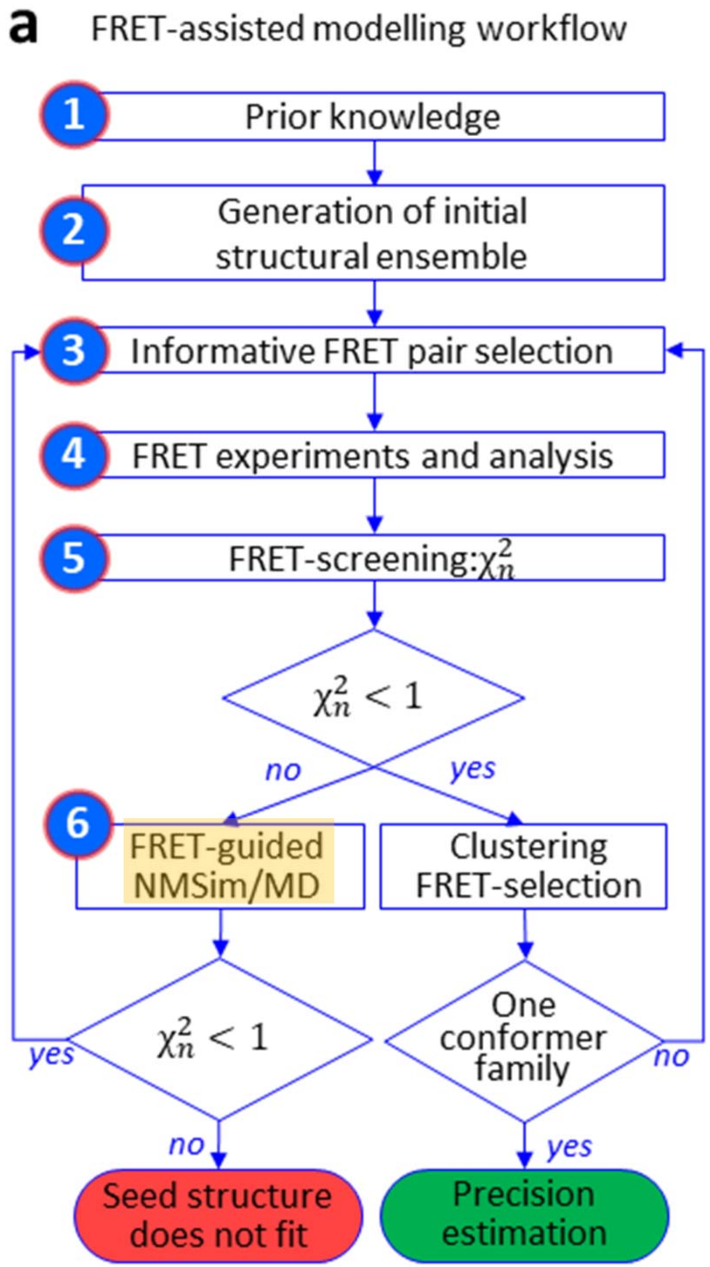


Quality parameter by cross validation: χ_n^2

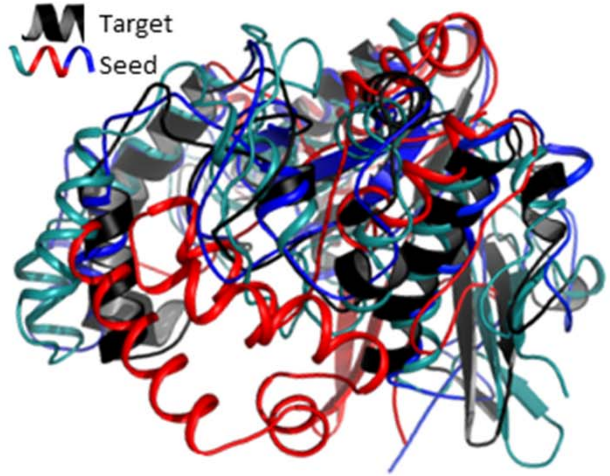
$$\chi_n^2 = \chi^2 / \chi^2_{\mathrm{p}=68\%}$$

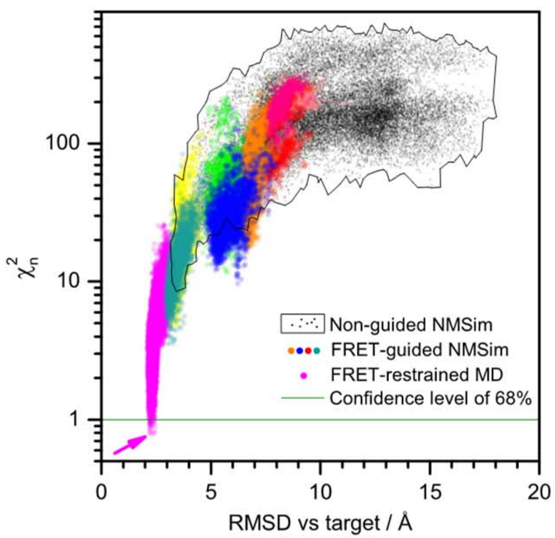
$$\chi_{\mathrm{p}}^2 = Inv.\chi^2\left(p, N_{dof}\right) = \frac{2^{-N_{dof}/2}}{\Gamma(N_{dof}/2)} p^{-N_{dof}/2-1} e^{-1/(2p)}$$

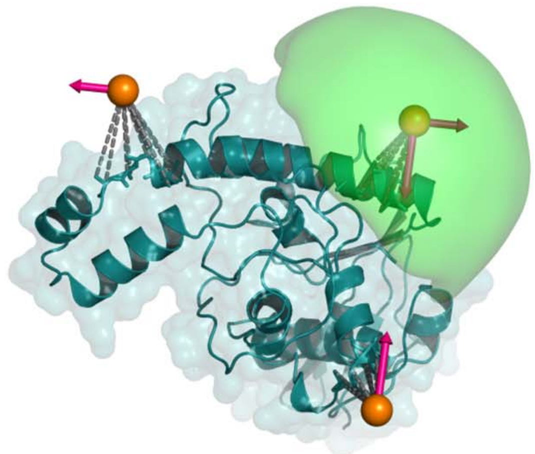
Accuracy of FRET-guided structural models



CASP11 target T0806: YaaA (PDB ID 5CAJ)

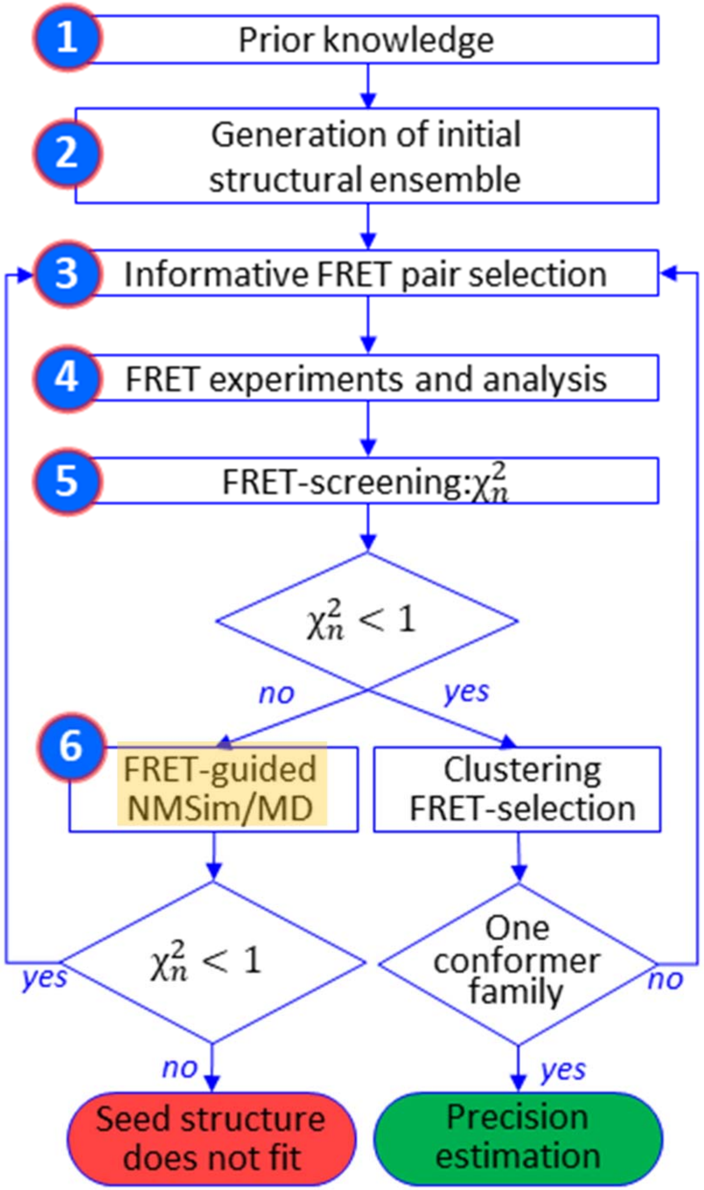




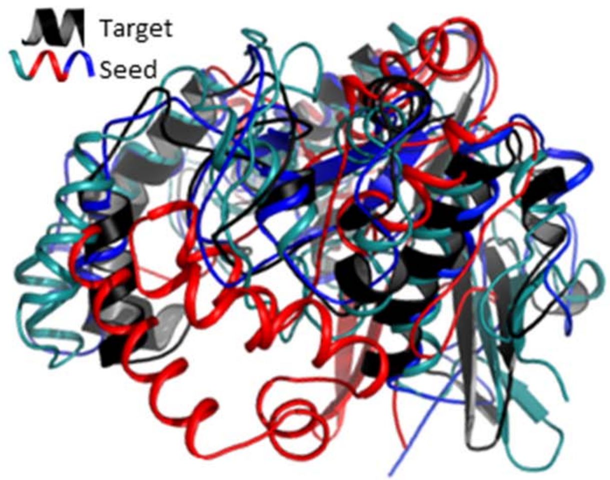


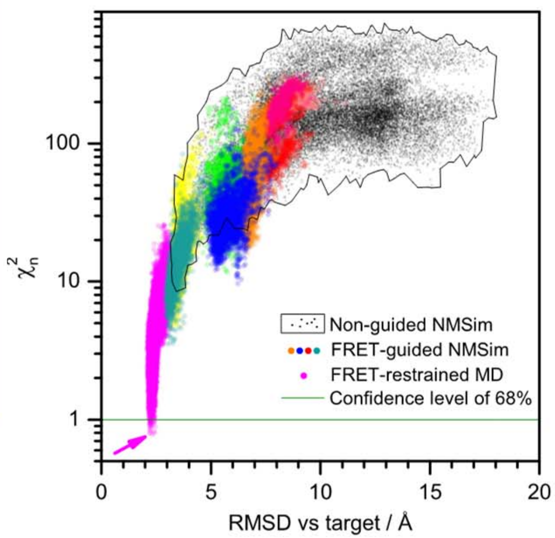
Accuracy of FRET-guided structural models

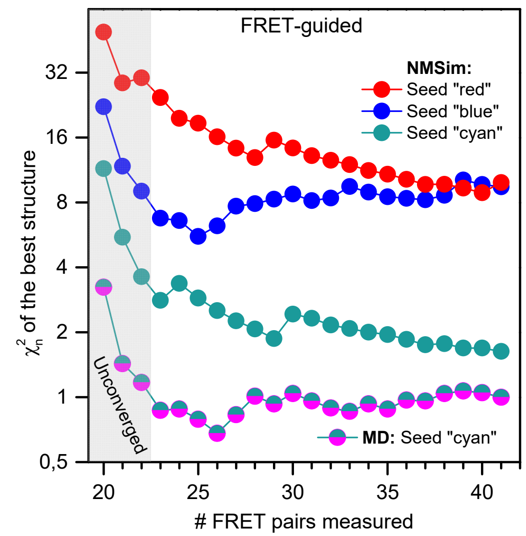
a FRET-assisted modelling workflow



CASP11 target T0806: YaaA (PDB ID 5CAJ)

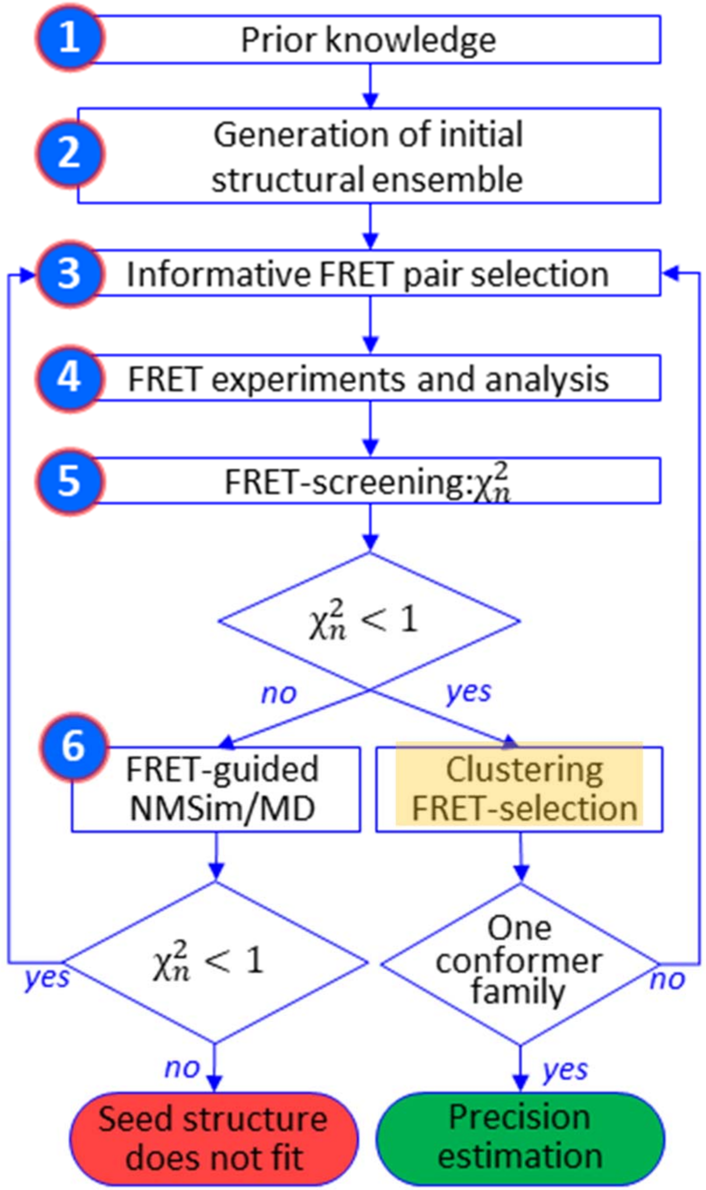




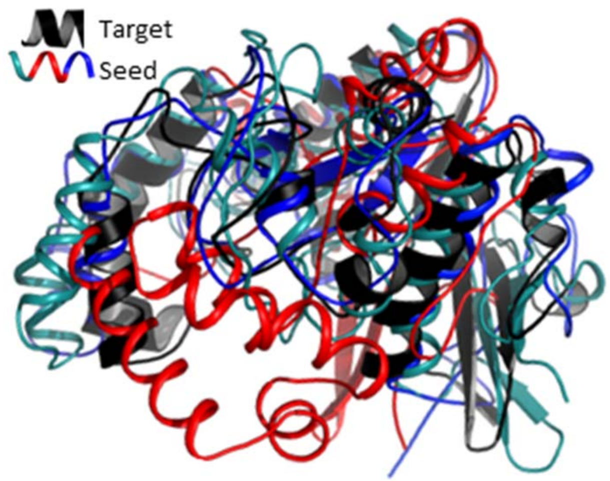


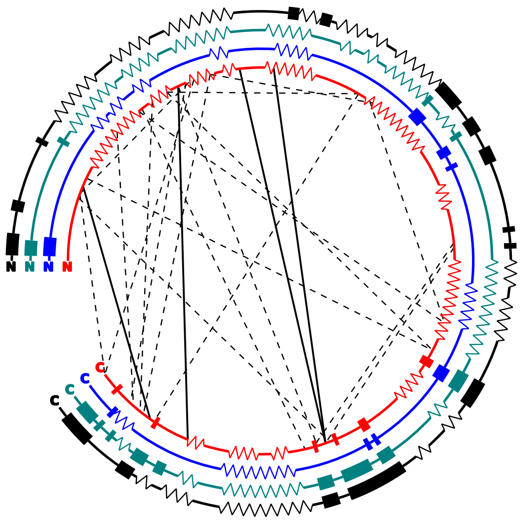
Accuracy of FRET-guided structural models

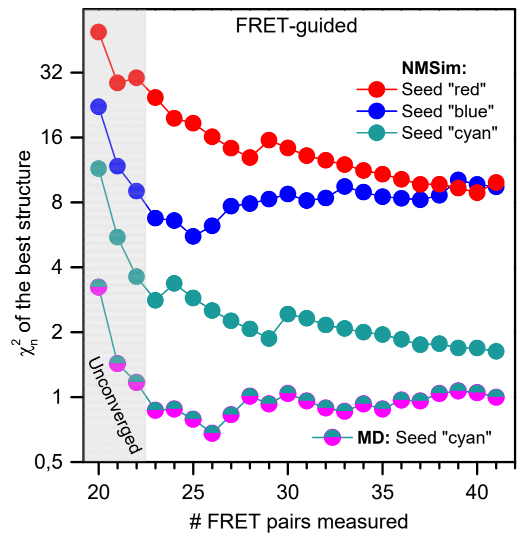
a FRET-assisted modelling workflow



CASP11 target T0806: YaaA (PDB ID 5CAJ)







Acknowledgements

hybrid-FRET team:

Mykola Dimura, Thomas-Otavio Peulen, Christian Hanke.

Soft- and Hardware:

www.mpc.hhu.de/software

https://github.com/Fluorescence-Tools

Examples:

Curr. Opin. Struct. Biol. 40, 163–185 (2016)

Model Archive: DOI: 10.5452/ma-a2hbq

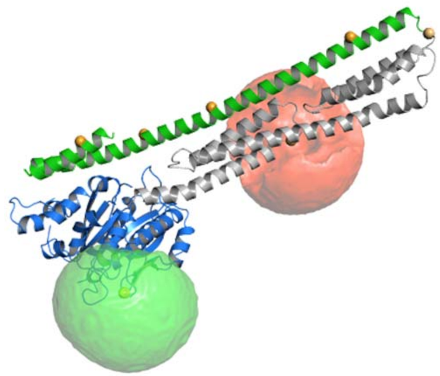
CASP13 webpage: ProteinDynamics-FRET webinar

European Research Council

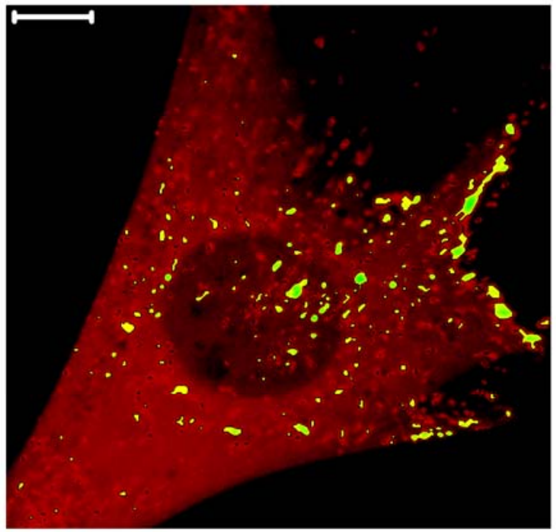
Established by the European Commission

hybridFRET :ERC Advanced Grant 2014

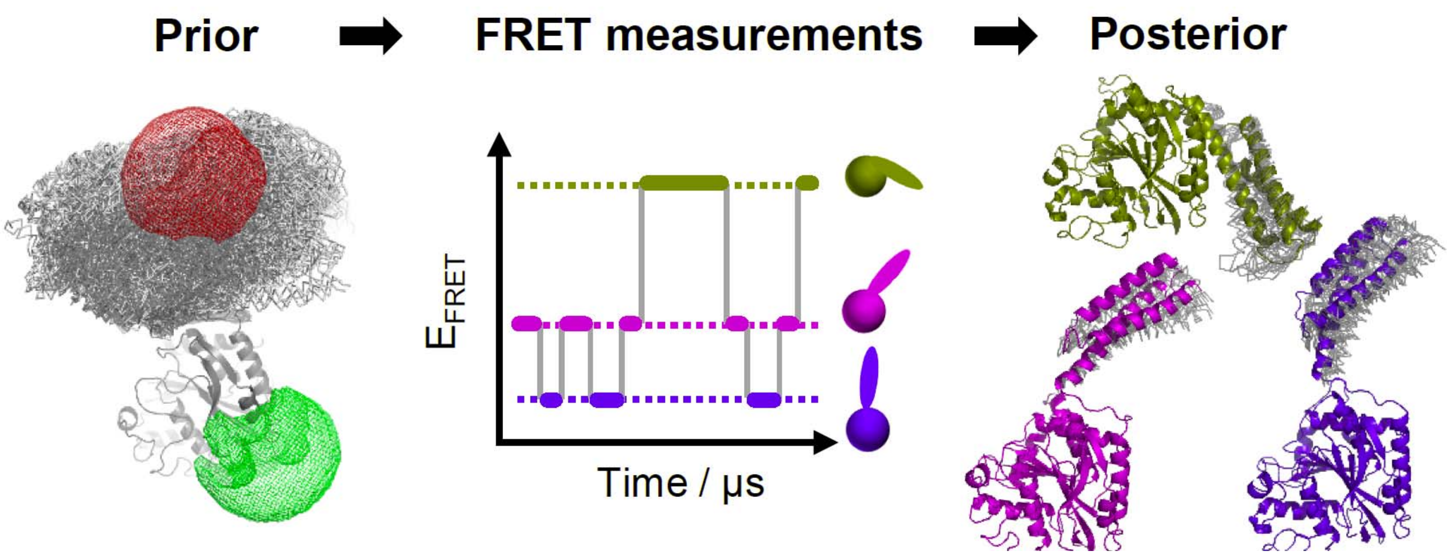
Watching ensembles



Live cell studies



Exploring the capabilities of hybrid-FRET modeling: An *in silico* experiment:



Curr. Opinion in Structural Biology 40, 163–185 (2016)

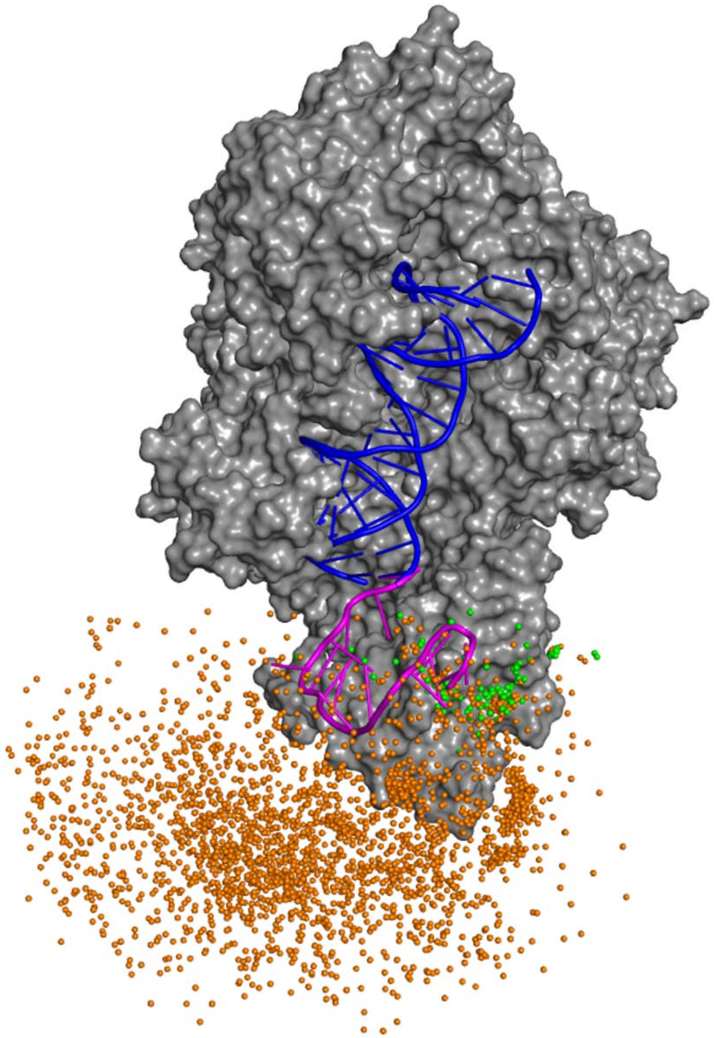
and

Model Archive: DOI: 10.5452/ma-a2hbq

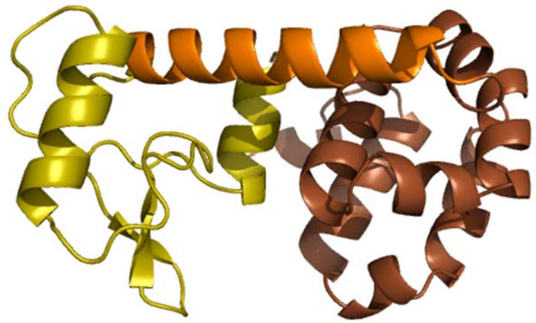
with a comprehensive description, all tools and data

Benchmark systems: Proteins

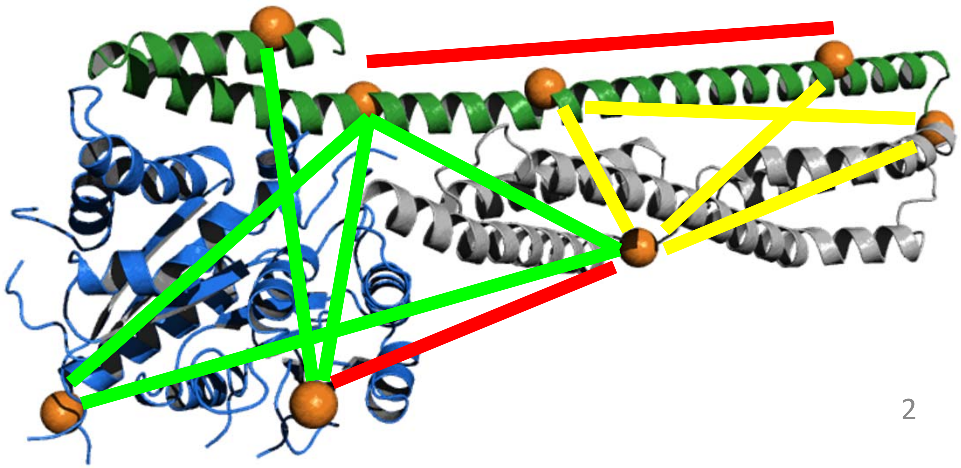
HIV- RT : p/t
(Nat. Methods 9 p.1218 (2012))



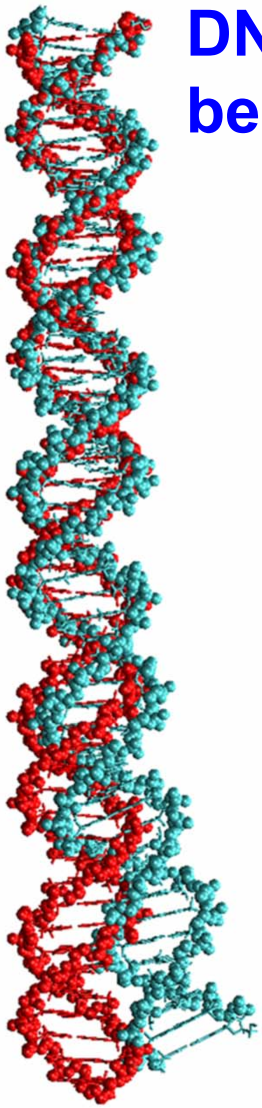
T4 Lysozmye first structure of a new hidden state (200 µs)



hGBP1 (new conformational state)



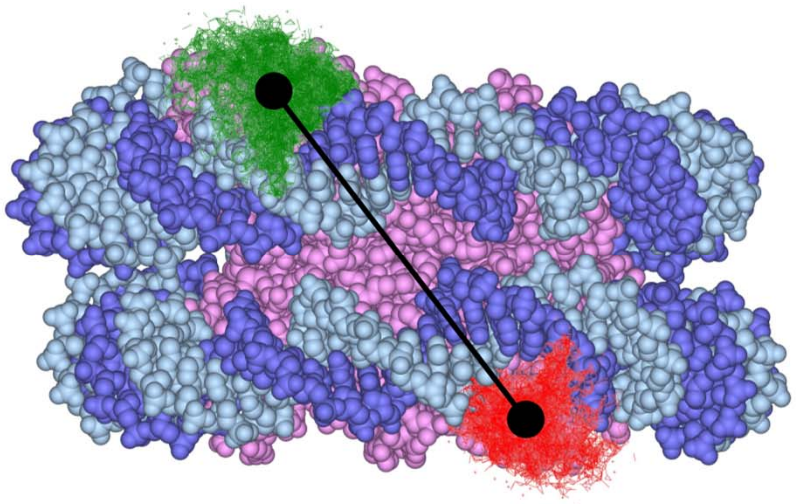
Benchmark systems: Nucleic acids



DNA:
bends (16°) and kinks

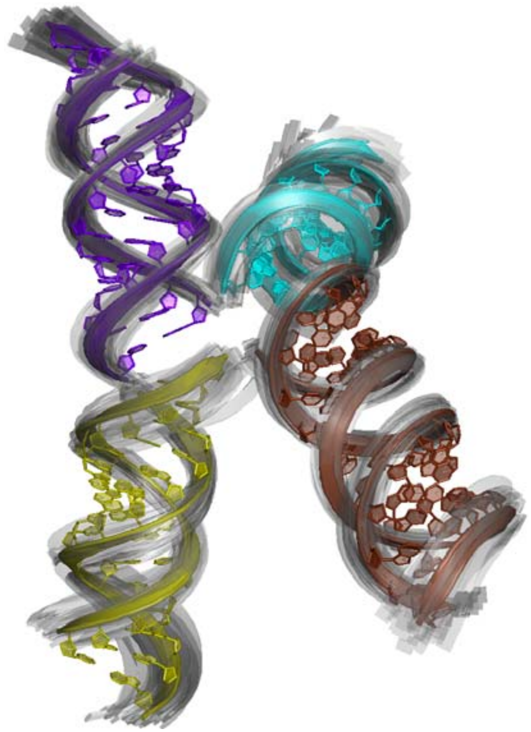
PNAS 105 p18773 (2008)

Mononucleosomes:
disassembly pathway



PNAS 106 p15308 (2009)

DNA + RNA: 3W
+ 4W-junctions



in preparation:

- > 260 FRET pairs
- 3 structures solved in parallel
- Precision 2 –^{3}4 Å

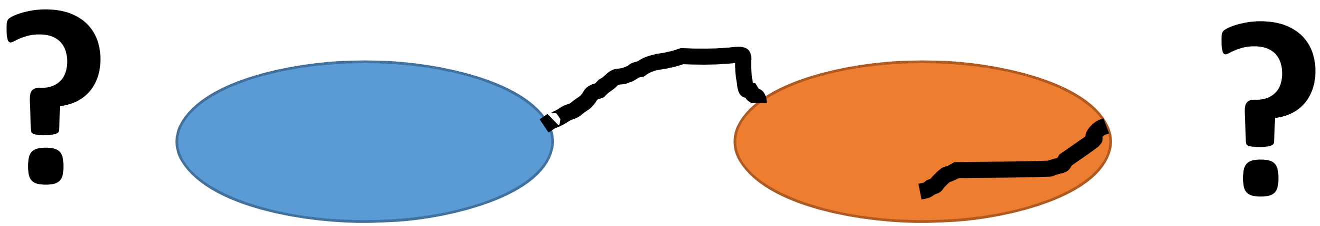
CASP target T0964 (CBM56) listed as F0964

- Protein name: CBM56
- Carbohydrate binding module from a β-1,3-glucanase (Bacillus circulans)
- Number of amino acids: 184 (694 - 877)
- Molecular weight: 18.936 kDa

```
                                                                                                694        700
                                                                                                  NVPVTGV
       710        720        730        740        750        760        770        780        790        800
TVNPTTAQVE VGQSVQLNAS VAPSNATNKQ VTWSVSGSSI ASVSPNGLVT GLAQGTTTVT ATTADGNKAA SATITVAPAP STVIVIGDEV KGLKKIGDDL
       810        820        830        840        850        860        870    877
LFYVNGATFA DLHYKVNNGG QLNVAMAPTG NGNYTYPVHN LKHGDTVEYF FTYNPGQGAL DTPWQTYVHG VTQGTPE
```

CASP target T0964 (CBM56) listed as F0964
Data-assisted modeling

- Protein name: CBM56
- Carbohydrate binding module from a β-1,3-glucanase (Bacillus circulans)
- Number of amino acids: 184 (694 - 877)
- Molecular weight: 18.936 kDa



12 of 14 variants worked

Efficiency <E>	**<RDA> / Å**	**± Δ<RDA> (Meta-analysis) / Å**	**± Δ<RDA> / Å**	**σDA (total) / Å**	**±ΔσDA (total) / Å**	**σDA (protein) / Å**

Dictionary for structural biologists

<table>
<tr><th>NMR</th><th>EPR</th><th>Fluorescence</th></tr>
<tr><td>2D NMR etc
τ (NOESY mixing time)</td><td>double resonator</td><td>multi-parameter detection</td></tr>
<tr><td>T1 time, 1/ ρ</td><td>T1 time</td><td>fluorescence lifetime, $\boldsymbol{\tau}$</td></tr>
<tr><td>order parameter S2</td><td>line shape analysis (High field EPR)</td><td>fluorescence anisotropy, r
rotational correlation time $\boldsymbol{\rho}$ and corresponding amplitudes</td></tr>
<tr><td>distance information:
NOE: short range
PRE: Paramagnetic Relaxation Enhancement (PRE) long-range</td><td>PELDOR/ DEER (distance r)</td><td>FRET (distance R)</td></tr>
<tr><td colspan="3">Problems + advantages with the label
+ Selectivity
- Labelling strategies
- Label position: (AV, rotamer libraries, MD simulations)
- Orientation factor (κ^2)</td></tr>
<tr><td>Line width analysis</td><td>line width analysis:
ns-dynamics</td><td>PDA
(Photon distribution analysis)</td></tr>
<tr><td>Relaxation dispersion analysis
(transverse spin-relaxation, CPMG relaxation dispersion NMR experiments)</td><td>(DEER: frozen sample)</td><td>dynamic PDA</td></tr>
<tr><td>Correlation methods</td><td>(DEER: frozen sample)</td><td>FCS (Fluorescence correlation spectroscopy)
no gaps in the time axis over 10 orders of magnitude
SCCF (Species cross correlation function) in the MFD space</td></tr>
</table>