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 systems

- Fluorescence decay
- Fluorescence correlation
- Immobilized traces
- Optical microscopy
- Super-resolution
- CryoEM
- Crystallography
- Meta-modelling
- Integrative modelling
- Fluorescence spectroscopy
- NMR
- PDA

Length:
- 1 µm
- 0.1 µm
- 1 nm
- 1 Å

Time:
- ps
- ns
- µs
- ms
- s
- h

- Local flexibility:
  - Bond vibration
  - Methyl rotation
  - Loop motion
  - Side-chain rotamers

- Collective motions:
  - Larger domain motions
  - Mediated function
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

**Chemistry:**
Purification..., but dynamic changes in time.

**Physics:**
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**

<table>
<thead>
<tr>
<th>Spectral changes (polarity probes, ratiometric probes)</th>
<th>Localization of the segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quenching</td>
<td>Proximity of other segments, accessibility of sites</td>
</tr>
<tr>
<td>Polarization, anisotropy</td>
<td>Order parameters (local flexibility) [1]</td>
</tr>
</tbody>
</table>

- **Global structure**

<table>
<thead>
<tr>
<th>Förster energy transfer (FRET)</th>
<th>Inter-dye distances [2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion (translation and rotation)</td>
<td>Shape</td>
</tr>
</tbody>
</table>

- **Dynamic exchange**

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

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


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 synchronizaion

- 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

1. Starting Model: rigid bodies, conformational ensemble
2. Network of DA-pairs
3. Quantitative FRET-measurements (MFD)
   - Distances (PDA, FRET-toolbox)
   - Errors ($\chi^2$, PDA)
4. - FRET-guided positioning, or
   - coarse grained model generation and screening by FPS
   - Atomistic model refinement (MD)
5. Model discrimination
6. Final unique model, Bootstrapping

Benchmark study: rSMD (hpFRET vs. X-ray) 0.5 Å
Software available: www.mpc.hhu.de/software

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

Standard deviation between the exp. FRET efficiencies:
\[ \Delta E = \pm 0.02 \text{ and } \pm 0.05 \]

Rel. deviations between mean exp. and model distances \( \langle R \rangle \):
\[ \Delta R/R_{\text{mod}} = 0 \text{ and } \pm 0.05 \]

→ well within the expected error

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
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

<table>
<thead>
<tr>
<th>Systems</th>
<th>States</th>
<th>Relaxation times</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proteins with 2 domains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNARE Protein Syntaxin 1</td>
<td>2</td>
<td>(700 µs)</td>
</tr>
<tr>
<td>PGK</td>
<td>&gt;2</td>
<td>(fast)</td>
</tr>
<tr>
<td>T4 Lysozyme</td>
<td>&gt;3</td>
<td>(3, 200 µs)</td>
</tr>
<tr>
<td>IF3 (ribosomal initiation factor)</td>
<td>&gt;3</td>
<td>(7) (2 µs, 300 µs)</td>
</tr>
</tbody>
</table>

| **NA binding proteins**                      |        |                                         |
| 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)                                  |

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

| **Intrinsically disordered proteins**        |        |                                         |
|                                              | ....... |                                         |
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:
     \( R_{\text{free}} \) 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

Sequence and chemical repositories
UniProt GenBank CCDC

Structural data repositories
PDB PMP

Experimental data repositories
PDB BMRB EMDB SASDB

Molecular Components
Proteins Nucleic acids Ligands

Starting structural models obtained from
X-ray crystallography NMR spectroscopy In silico methods

Spatial restraints obtained from
EM CX-MS FRET EPR NMR SAS AFM Proteomics

Integrative Modeling
Multi-scale Multi-state Time-ordered

Ensembles of macromolecular assemblies

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: $\chi_n^2$

$\chi_n^2 = \chi^2 / \chi_p^{68\%}$

$\chi_p^2 = \text{Inv.} \chi^2 (p, N_{dof})$

$= \frac{2^{-N_{dof}/2}}{\Gamma(N_{dof}/2)} p^{-N_{dof}/2-1} e^{-1/(2p)}$
Accuracy of FRET-guided structural models

CASP11 target T0806: YaaA (PDB ID 5CAJ)
Accuracy of FRET-guided structural models

CASP11 target T0806: YaaA (PDB ID 5CAJ)
Accuracy of FRET-guided structural models

### FRET-assisted modelling workflow

1. **Prior knowledge**
2. **Generation of initial structural ensemble**
3. **Informative FRET pair selection**
4. **FRET experiments and analysis**
5. **FRET-screening:** $\chi^2_n$

- $\chi^2_n < 1$: yes
- $\chi^2_n \geq 1$: no

6. **FRET-guided NMSim/MD**
   - $\chi^2_n < 1$: one conformer family
   - $\chi^2_n \geq 1$: seed structure does not fit

#### Clustering FRET-selection
- One conformer family: yes
- Seed structure does not fit: no

#### Precision estimation

### CASP11 target T0806: YaaA (PDB ID 5CAJ)

- NMSim: Seed "red" Seed "blue" Seed "cyan"
- MD: Seed "cyan"
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:**
Model Archive: DOI: 10.5452/ma-a2hbq
CASP13 webpage: ProteinDynamics-FRET webinar

**hybridFRET** : ERC Advanced Grant 2014
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 Lysozyme first structure of a new hidden state (200 µs)**

- **hGBP1 (new conformational state)**
Benchmark systems: Nucleic acids

DNA: bends (16°) and kinks

Mononucleosomes: disassembly pathway

DNA + RNA: 3W + 4W-junctions

in preparation:
• > 260 FRET pairs
• 3 structures solved in parallel
• Precision 2 – 4 Å

PNAS 105 p18773 (2008)

PNAS 106 p15308 (2009)
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
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>
<thead>
<tr>
<th>NMR</th>
<th>EPR</th>
<th>Fluorescence</th>
</tr>
</thead>
<tbody>
<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, τ</td>
</tr>
<tr>
<td>order parameter S2</td>
<td>line shape analysis (High field EPR)</td>
<td>fluorescence anisotropy, r rotational correlation time ρ 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>
</tbody>
</table>

**Problems + advantages with the label**

+ Selectivity
- Labelling strategies
- Label position: (AV, rotamer libraries, MD simulations)
- Orientation factor (κ²)

<table>
<thead>
<tr>
<th>Line width analysis</th>
<th>line width analysis: ns-dynamics</th>
<th>PDA (Photon distribution analysis)</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>