Centre de
Montpelli

Combined Use of NMR and SAS for Flexible Systems

Pau Bernadó

Centre de Biologie Structurale - Montpellier

pau.bernado@cbs.cnrs.fr

Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS

Nathalie Sibille and Pau Bernadó[1]
Centre de Biochimie Structurale, INSERM U1054, CNRS UMR 5048, Université Montpellier 1 and 2, 29 rue de Navacelles, 34090 Montpellier, France

Biochem. Soc. Trans. (2012) 40, 955-962.

Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2012, **8**, 151–167

www.rsc.org/molecularbiosystems

REVIEW

Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering†

Pau Bernadó*[a] and Dmitri I. Svergun*[b]

Mol. BioSyst. (2012) 8, 151-167.

ELSEVIER

Available online at www.sciencedirect.com
SciVerse ScienceDirect

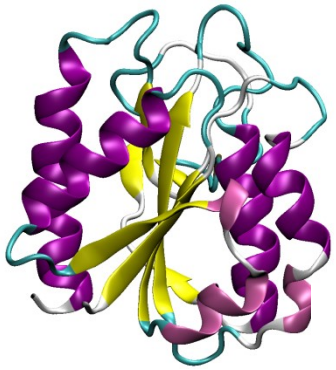
Current Opinion in Structural Biology

Small-angle scattering studies of intrinsically disordered proteins and their complexes

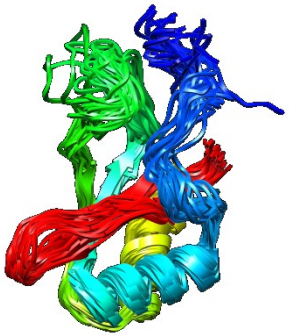
Tiago N Cordeiro[1], Fátima Herranz-Trillo[1,3], Annika Urbanek[1], Alejandro Estaña[1,2], Juan Cortés[2], Nathalie Sibille[1] and Pau Bernadó[1]

Curr. Opin. Struct. Biol. (2017) 42, 15-23.

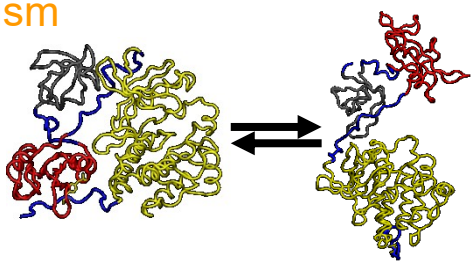
ORDER

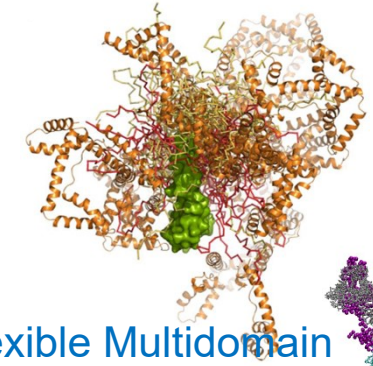


Protein vibrations
Catalysis
Partner Recognition

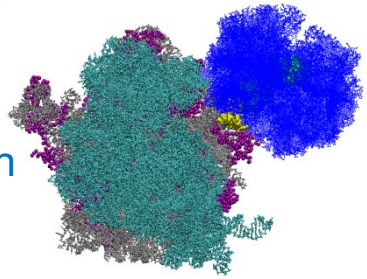


Concerted Motions
On/off mechanisms
Allosterism

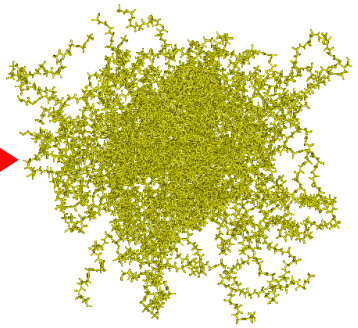




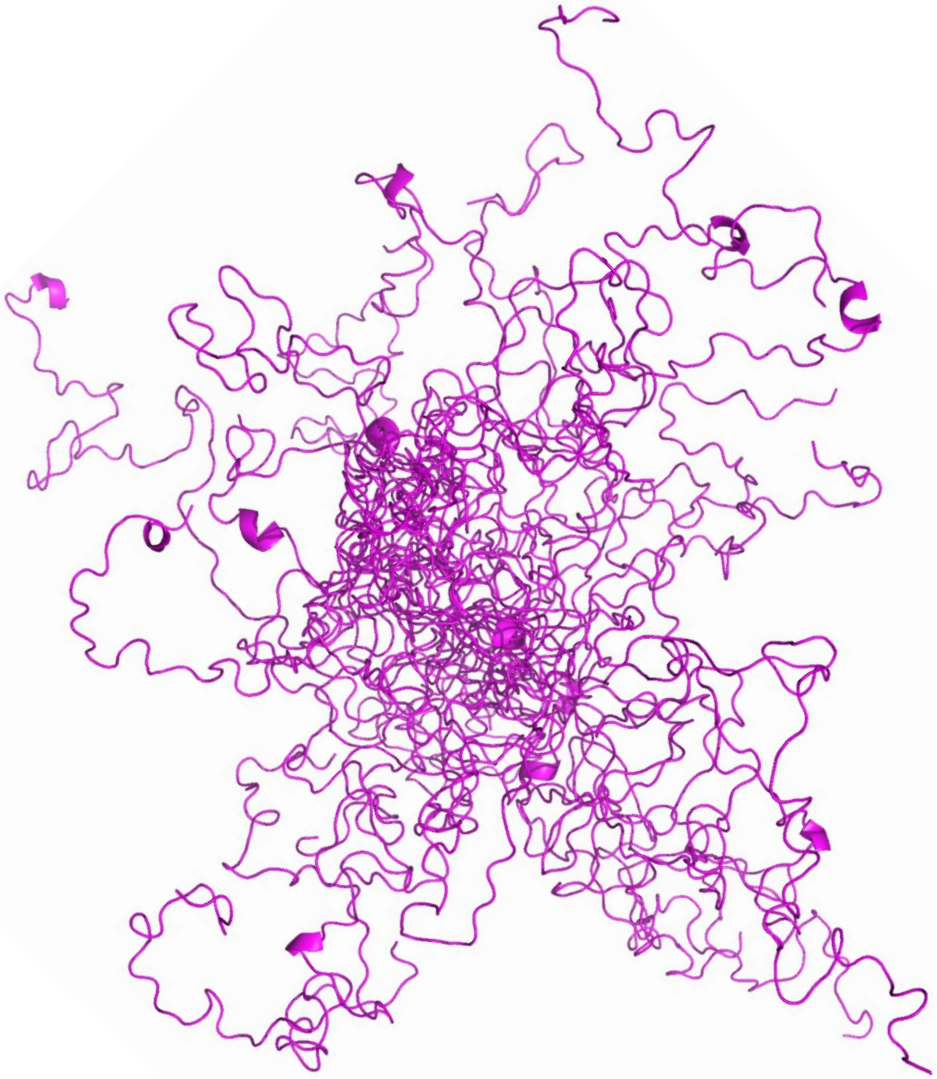
Flexible Multidomain
Proteins
Springs
Entropic Linkers
Recognition/Fishing



IDPs
Signalling and control
PTM
Recognition



DISORDER

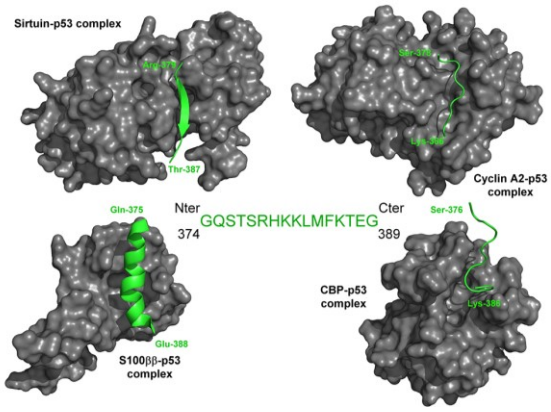


Intrinsically Disordered Proteins

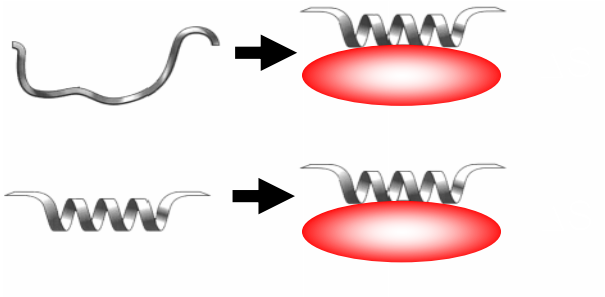
Intrinsically Disordered Proteins and Intrinsically Disordered Regions

- IDPs lack stable tertiary and/or secondary structure
- IDPs are more common in eukaryotes than in bacteria and archaea: Probably linked with their major biological complexity… Up to ~30-50% of genome
- Rich in charged and non-structuring residues while depleted in hydrophobic ones. Enriched in low complexity and compositionally biased sequences
- IDPs are the target of the majority of PTM: Phosphorylation, Glycosylation…
- IDPs are submitted to a strong evolutionary activity
- Disordered regions predominate in alternatively spliced regions
- Key elements in cancer, cardiovascular and neurodegenerative pathologies
- Interactome hubs are enriched in disordered regions

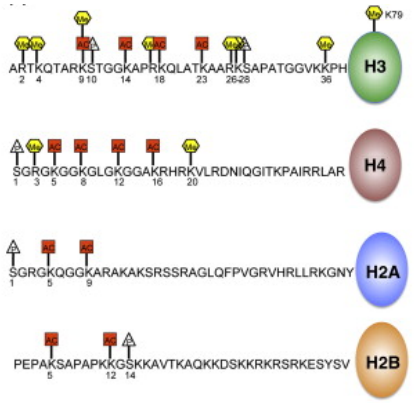
Why Disordered?



Specific but Promiscuous



Fine tuning of the thermodynamic properties



Accessible to PTM

ORDER

keywords
GMP biosynthesis
Amino-acid biosynthesis
Transport
Electron transport
Lipid A biosynthesis
Aromatic hydrocarbons catabolism
Glycolysis
Purine biosynthesis
Pyrimidine biosynthesis
Carbohydrate metabolism
Branched-chain amino acid biosynthesis
Lipopolysaccharide biosynthesis
Sugar transport
Antibiotic resistance
Lipid synthesis
Tricarboxylic acid cycle
Arginine biosynthesis
Ion transport
Rhamnose metabolism
Peptidoglycan synthesis

Both Ordered and Disordered regions are associated with distinct functions

Disordered Proteins complement the functions of ordered protein regions

Xie et al. *J Prot Res* 2007, 6, 1882.

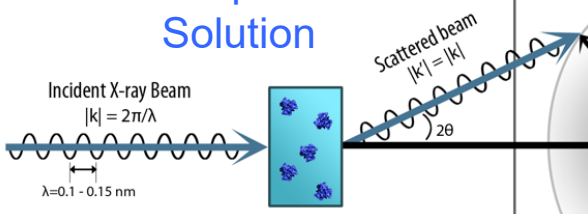
keywords
Differentiation
Transcription
Transcription regulation
Spermatogenesis
DNA condensation
Cell cycle
mRNA processing
mRNA splicing
Mitosis
Apoptosis
Protein transport
Meiosis
Cell division
Ubl conjugation pathway
Wnt signaling pathway
Neurogenesis
Chromosome partition
Ribosome biogenesis
Chondrogenesis
Growth regulation

DISORDER

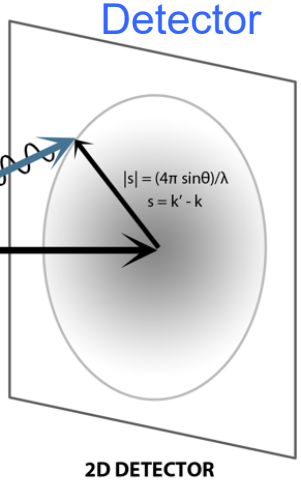
Small-Angle X-ray Scattering (SAXS)

Synchrotron

Sample in Solution

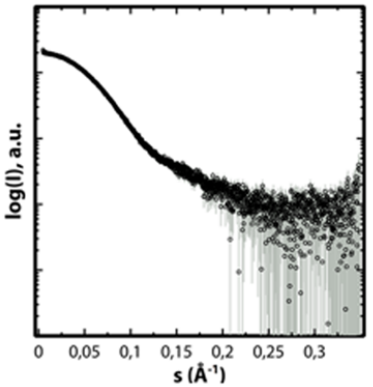


Detector

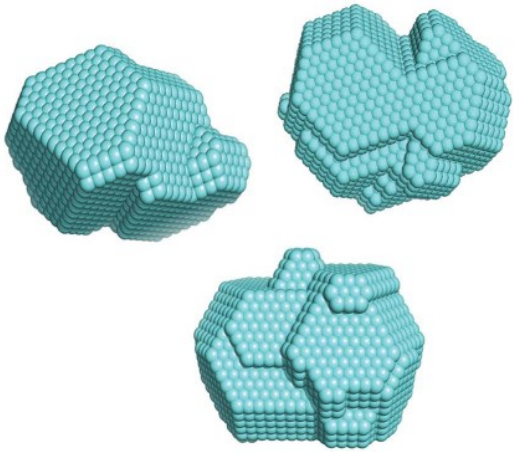


Integration

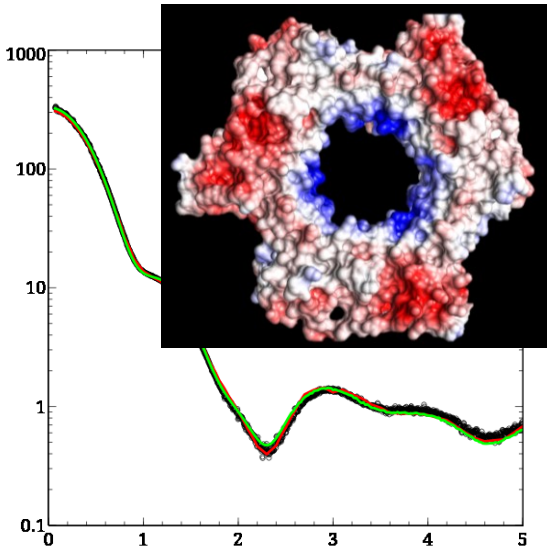
SAXS data



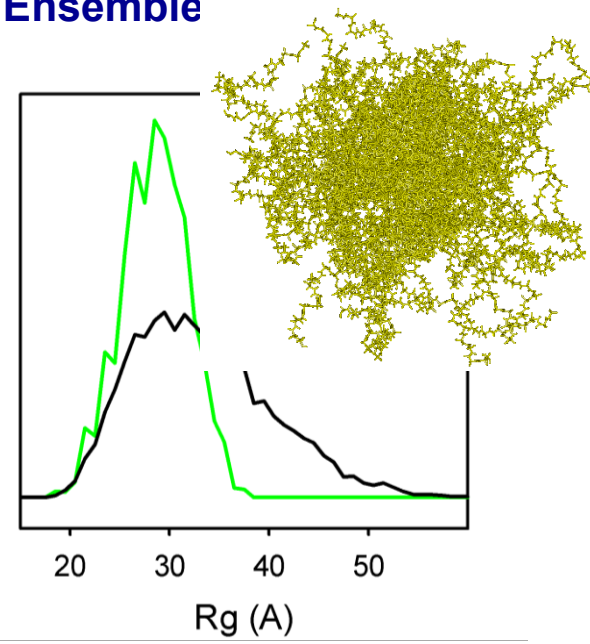
Low-Resolution Structure



Model Validation



Ensemble



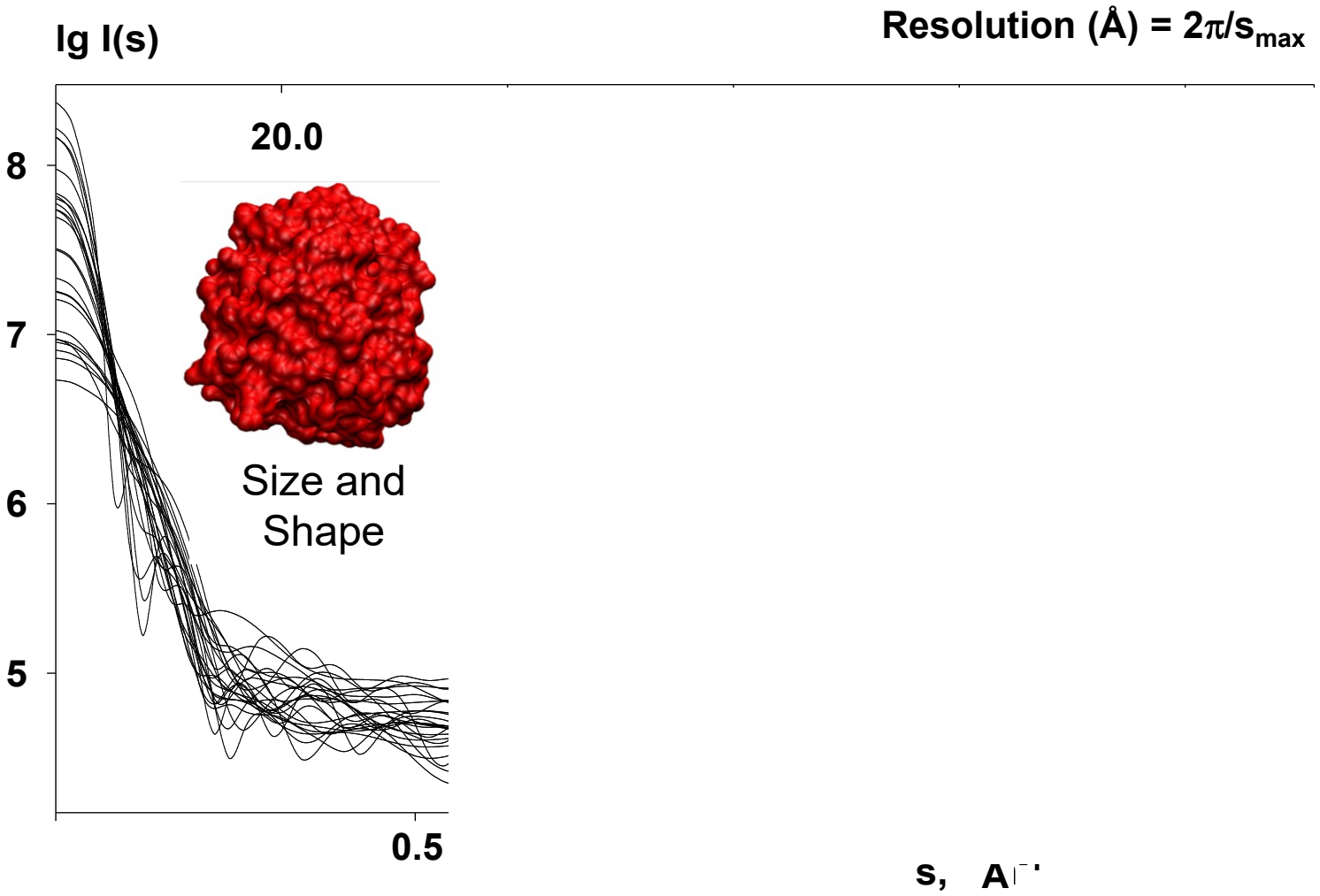
SAXS in Flexible Proteins

A Great Opportunity for SAS

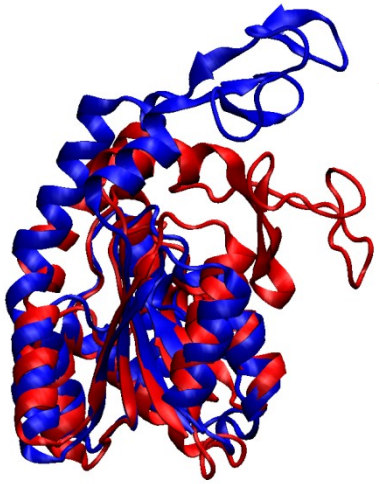
In flexible proteins the measured SAXS curve is the weighted average of all conformations present in solution

$$\langle I(s) \rangle = \Sigma v_i \, I_i(s)$$

Small Angle X-Ray Scattering Resolution

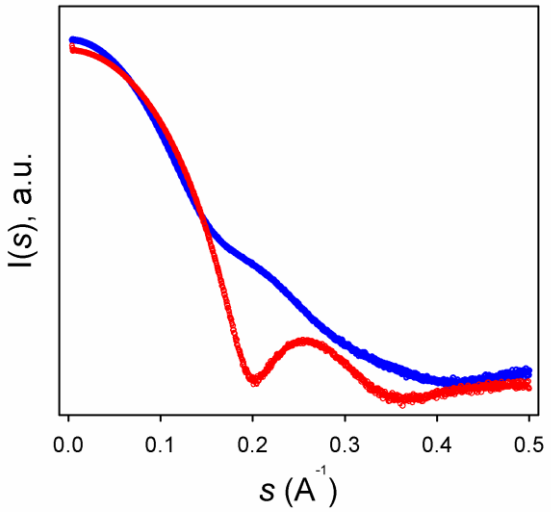


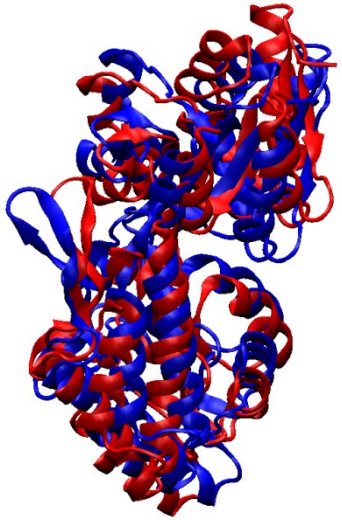
What does it mean Resolution in SAXS?



Adenylate Kinase

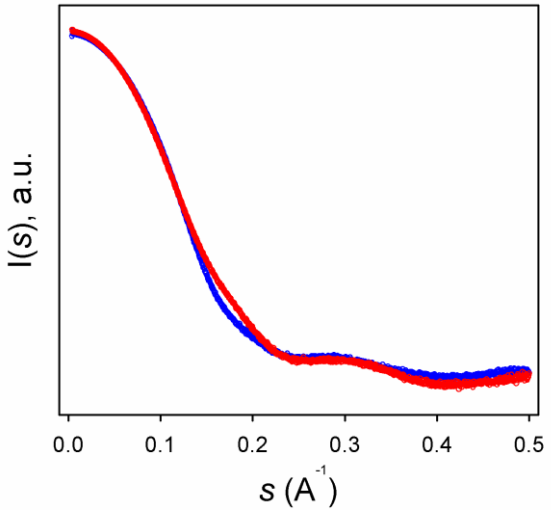
RMSD = 7.0 Å





Maltose Binding Protein

RMSD = 3.8 Å

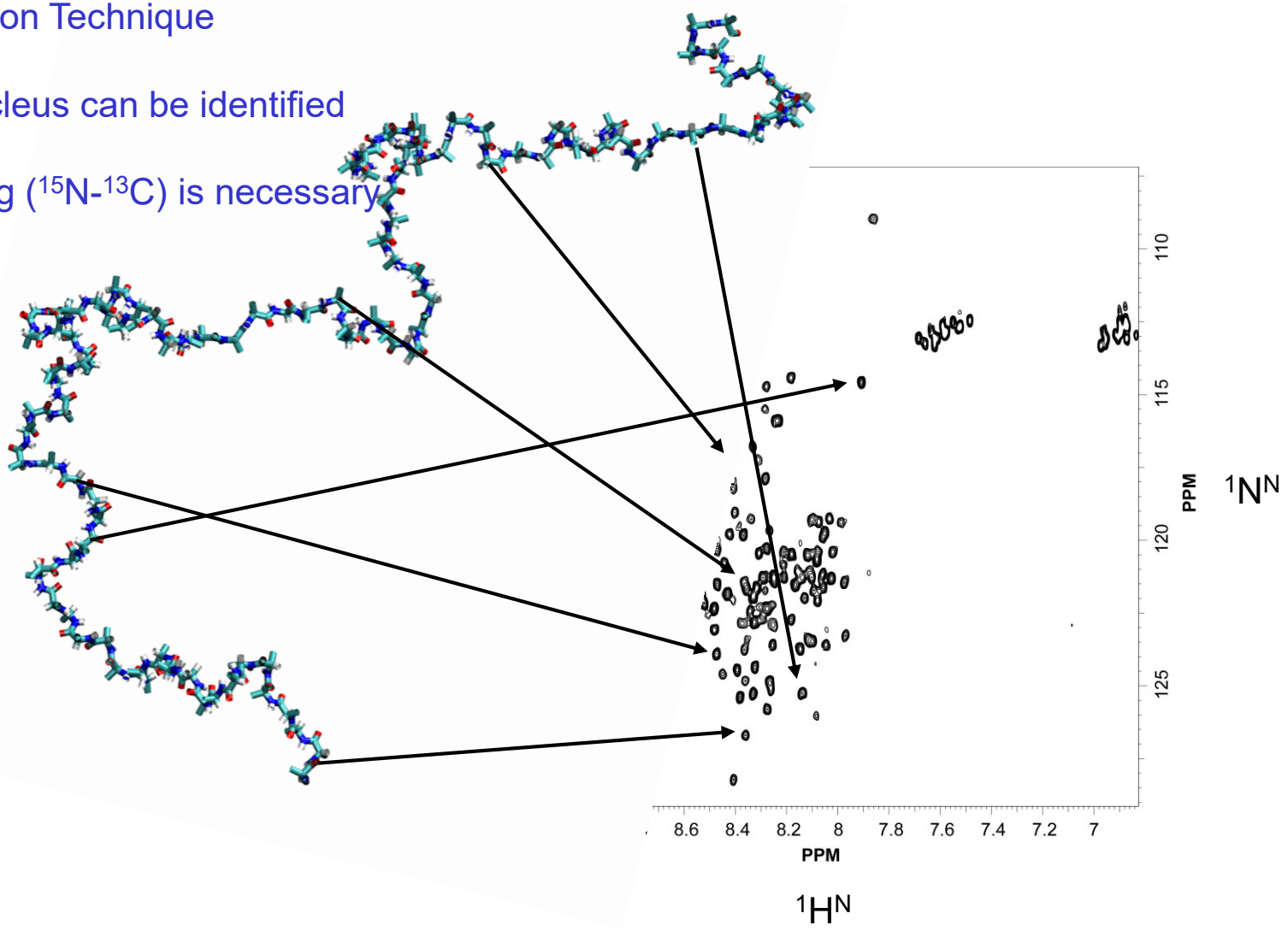


Nuclear Magnetic Resonance

Atomic Resolution Technique

Each active nucleus can be identified

Isotopic labelling (^{15}N-^{13}C) is necessary



NMR a versatile Source of Structural Information

Local Conformations
Chemical Shifts, J-Couplings

Interatomic Distances
^{1}H-^{1}H NOEs, PREs, Pseudocontact Shifts

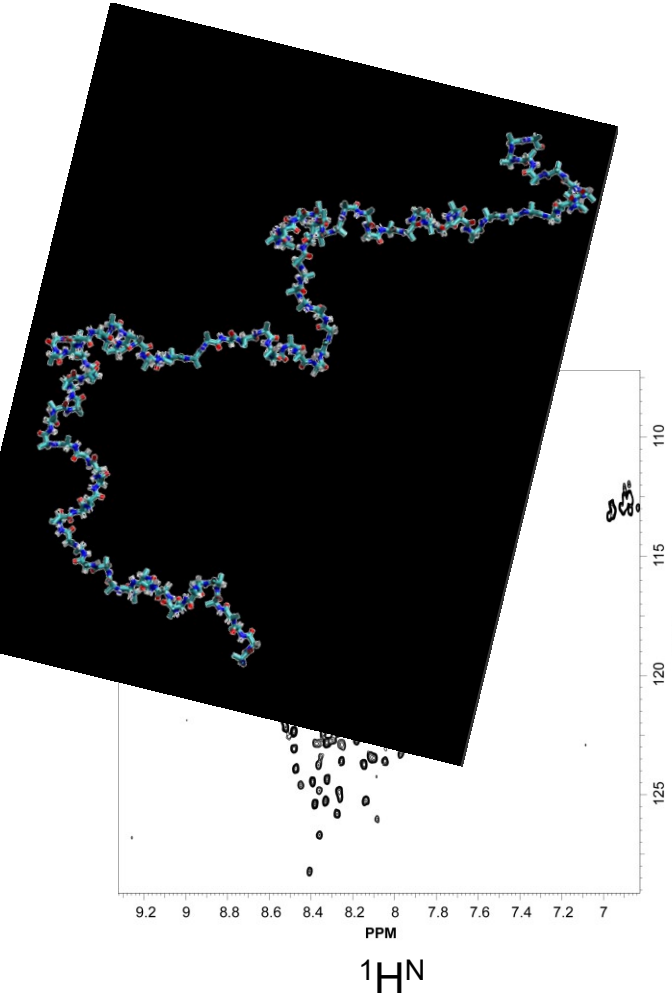
Orientational Restrictions
RDCs, Relaxation (R_2/R_1), Pse

Overall Size
Diffusion Coefficie ional Diffusion Tensor), RDCs (Alignment Tensor)

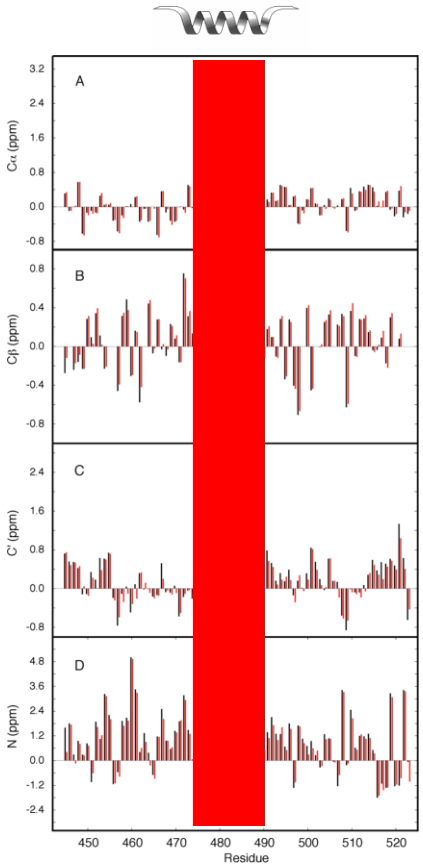
Time-Sc
Rela ε), Relaxation Dispersion, $R_1\rho$, CEST

The Majority of these Observables are also Ensemble Averaged

Chemical Shifts

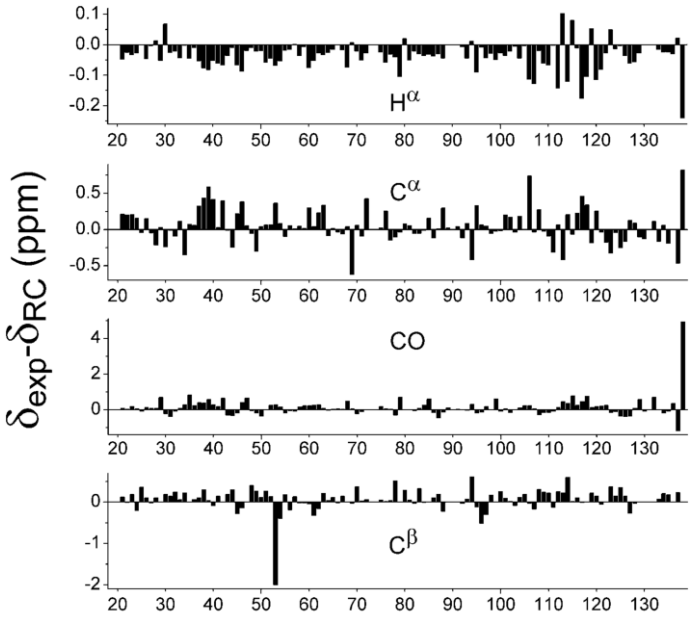


$^1H^N$



Sendai virus N-tail

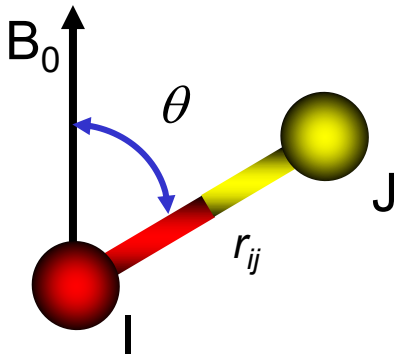
Identification of transiently formed seconday structural elements

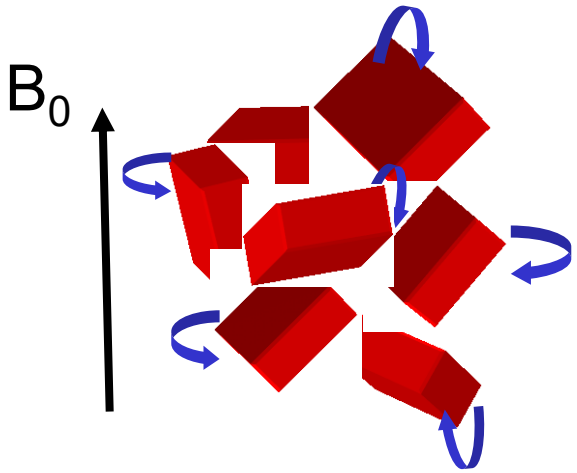


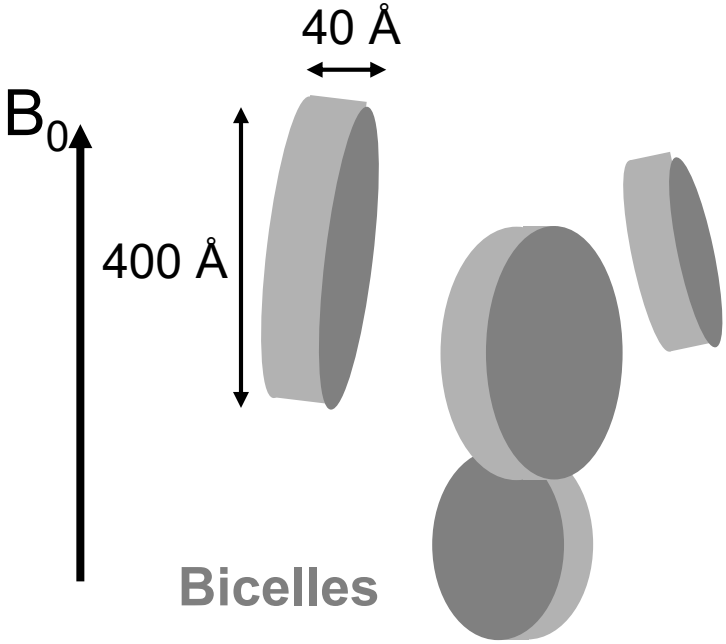
Kiss1 residue number

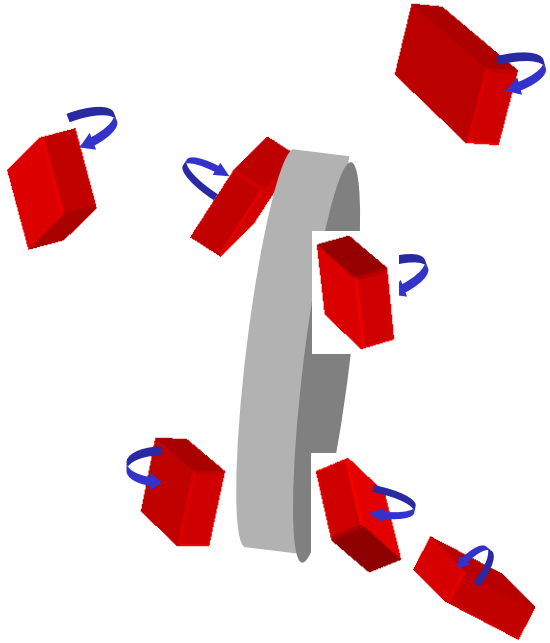
Residual Dipolar Couplings (RDCs) in Partially Aligned Systems

$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{8\pi^3}\left\langle \frac{P_2(\cos\theta(t))}{r_{ij}^3} \right\rangle$$



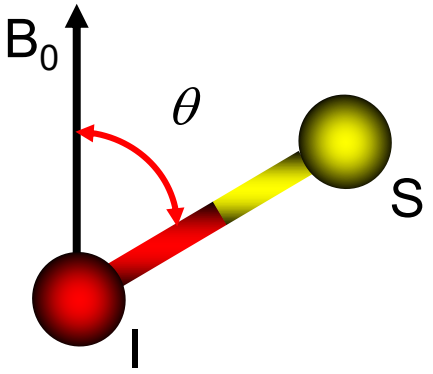






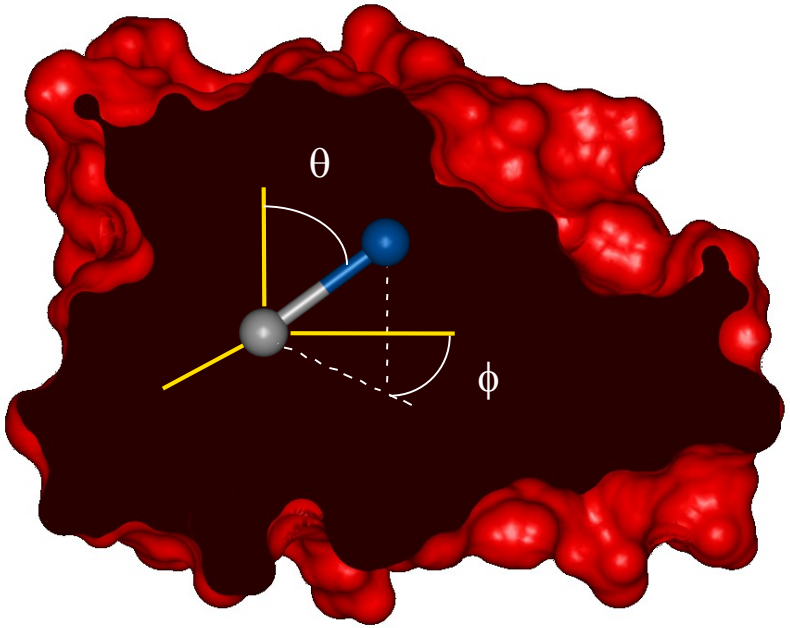
Tjandra, N.; Bax, A. *Science*, **1997**, *278*, 1111-1114.

Residual Dipolar Couplings by NMR



$$D_{ij} = -\frac{\gamma_i \gamma_j \mu_0 h}{8\pi^3}\left\langle \frac{P_2(\cos\theta(t))}{r_{ij}^3} \right\rangle$$

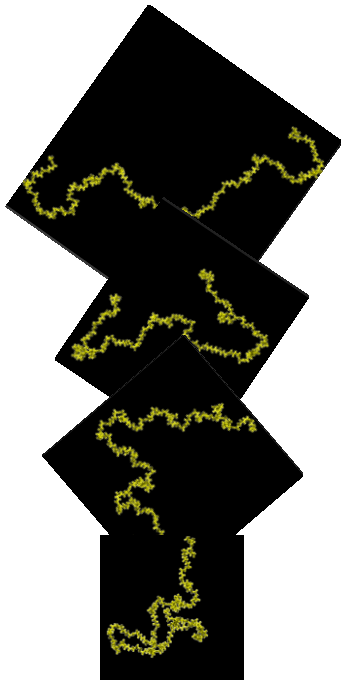
$$D_{ij} = -S\frac{\gamma_i \gamma_j \mu_0 h}{16\pi^3 r_{ij}^3}\left[A_a\left(3\cos^2\theta - 1\right) + \frac{3}{2}A_r \sin^2\theta\cos 2\varphi\right]$$

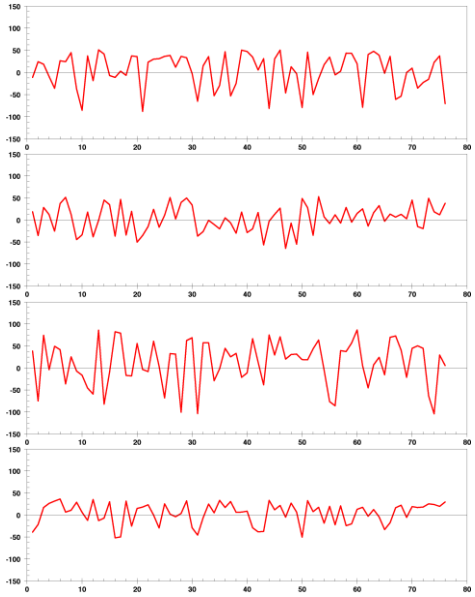


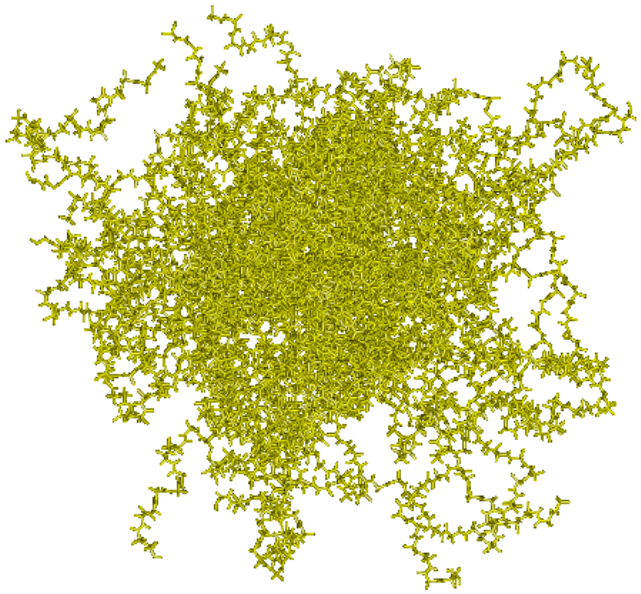
Residual dipolar couplings depend on the orientation of internuclear vectors relative to the alignment frame

RDCs are a valuable source of information to study structure and dynamics of biomolecules

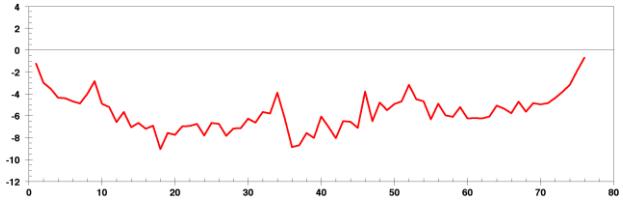
RDCs in Flexible Proteins





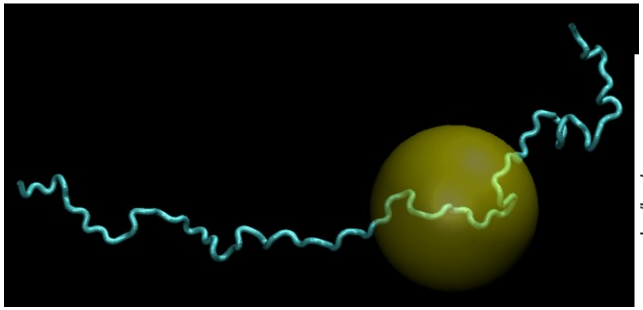


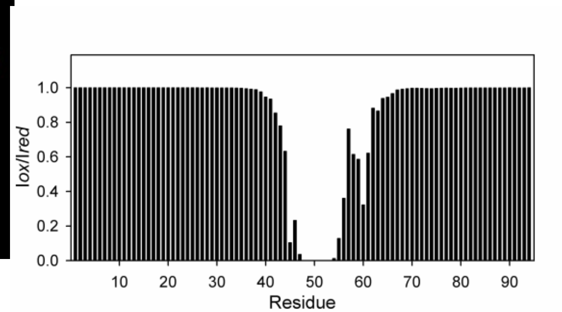
$\langle D_{ij} \rangle$



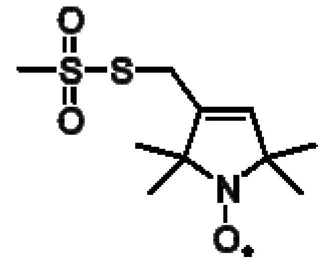
RDCs are the averaged value for all conformations in the ensemble

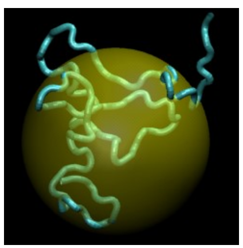
Paramagnetic Relaxation Enhancement (PRE)

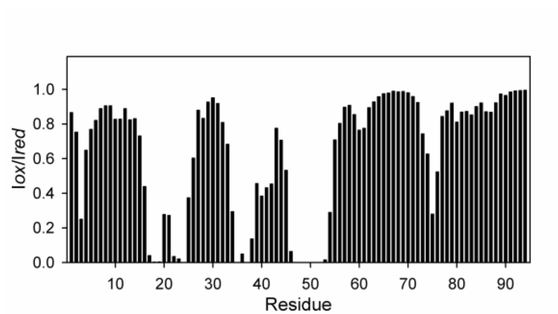




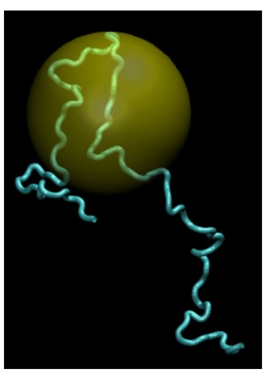
Free electron Radical attached to a Cys mutant

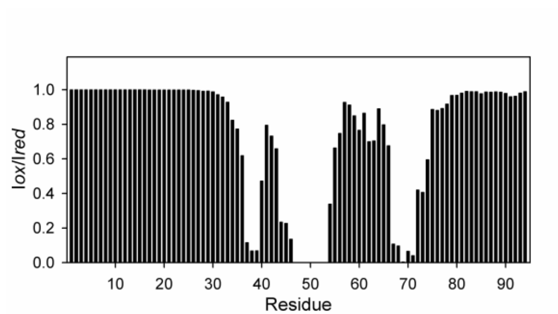






Residues in the proximity experience an enhancement of their R_2 relaxation



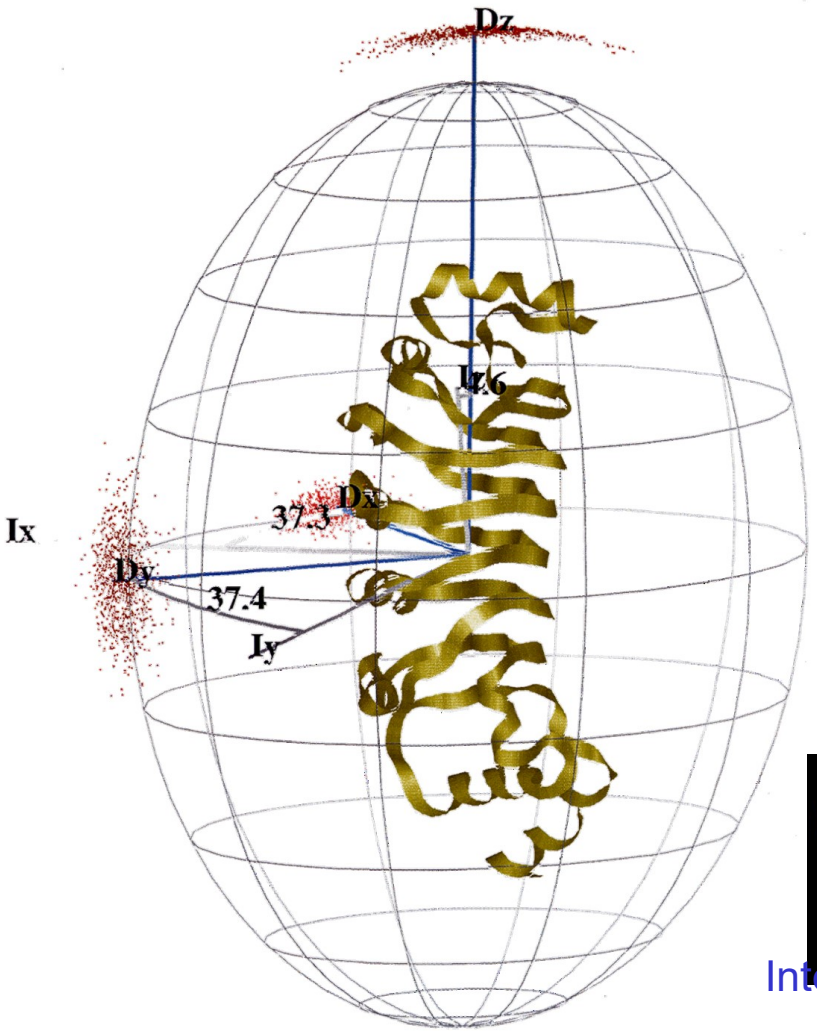


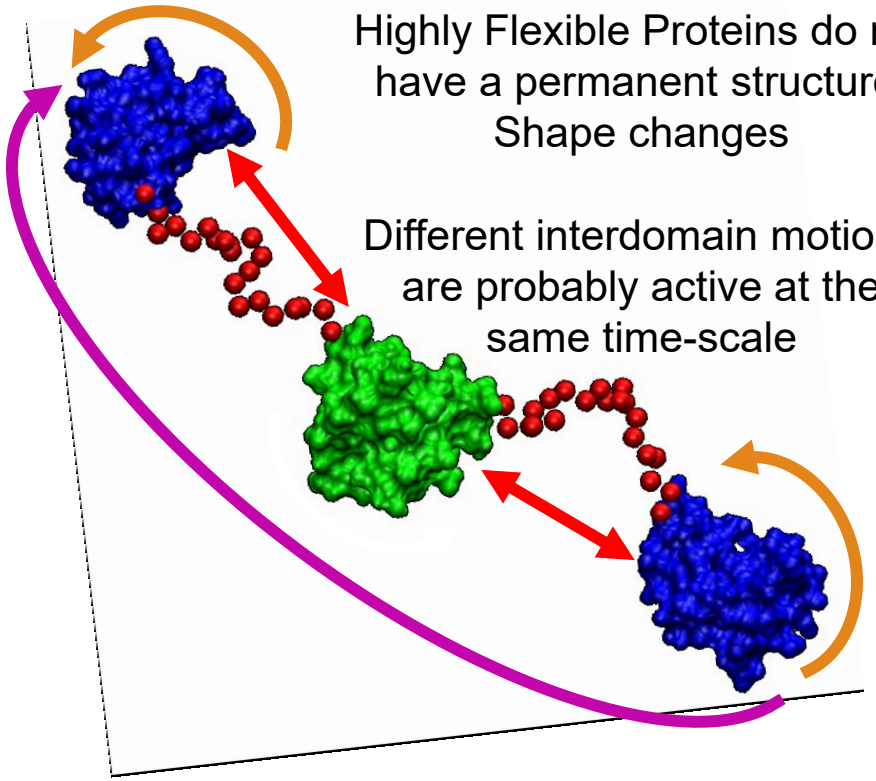
Effect up to r < 25 Å
r^6 dependence

Although it is a relaxation phenomena, in ensembles properties are normally averaged

Protein Dynamics from Relaxation Rates

Hydrodynamic Properties
Rotational Diffusion Tensor (ns)

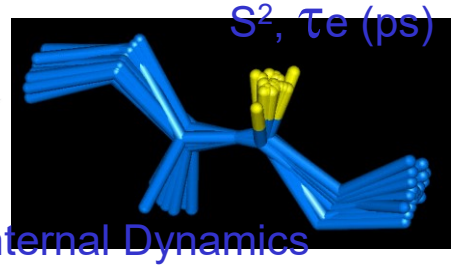




Highly Flexible Proteins do not have a permanent structure. Shape changes

Different interdomain motions are probably active at the same time-scale

S^2, τe (ps)



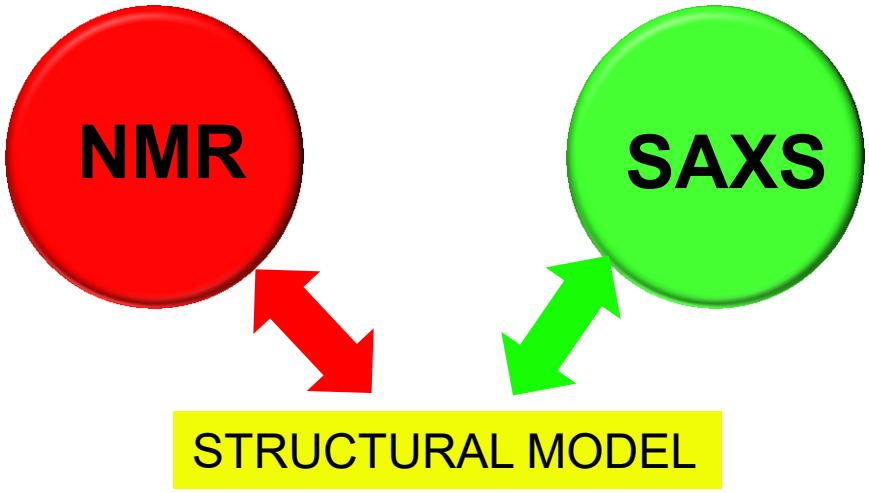
Internal Dynamics

Complementarity between SAXS and NMR in Structural Biology

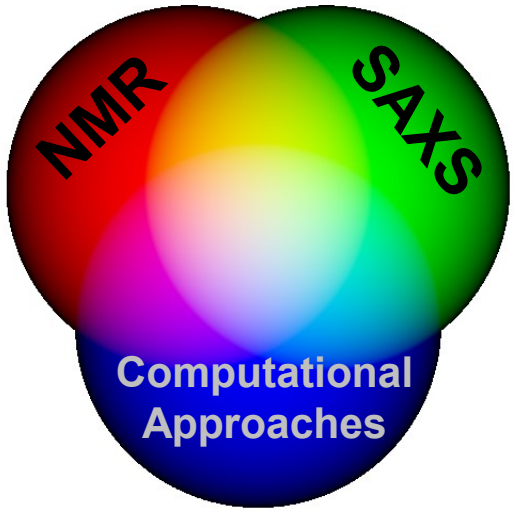
NMR		SAXS
Close Contacts: ^{1}H-^{1}H nOes LR relationship: RDCs, R_2/R_1	**Globular Proteins**	Overall Shape Validation
Interfaces: ^{1}H-^{1}H nOes, CS, PRE Orientation: RDCs, R_2/R_1	**Complexes and Rigid Multi-Domain Proteins**	Translational Information Overall Shape Validation
Conform. Sampling: CS, RDCs LR Contacts: PREs	**Unstructured Proteins**	Dimensions of the Ensemble
Time-Scale Information: Spin Relaxation	**Flexible Multi-Domain Proteins**	Volume Sampled by the Domains

Combining SAXS and NMR: Two Aproaches

Model Validation



Data-Driven Modelling

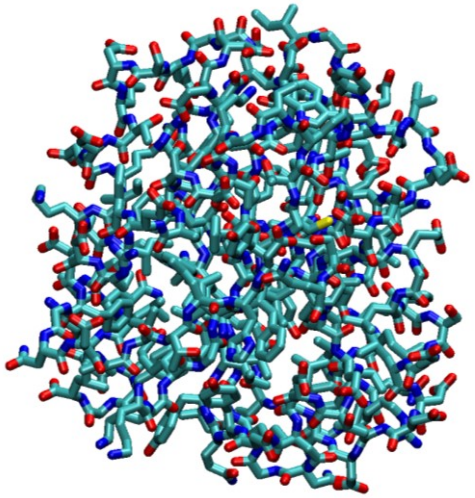


Capacity to address complex biological systems

Structural models with better resolution

More complete models embedding structure and dynamics

Globular Protein

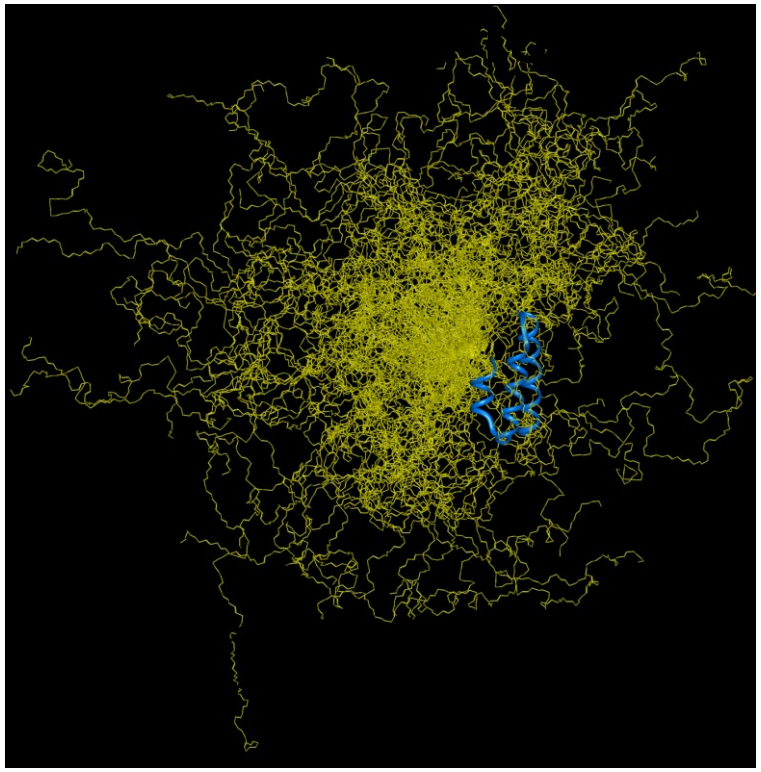


Multiple Constraints are measured (Short and Long Range)

Optimization of a single set of coordinates to simultaneously describe all data

Single Structure

Limited number of degrees of freedom (previous knowledge)



Optimization of an **ensemble of coordinates** to simultaneously describe all data available

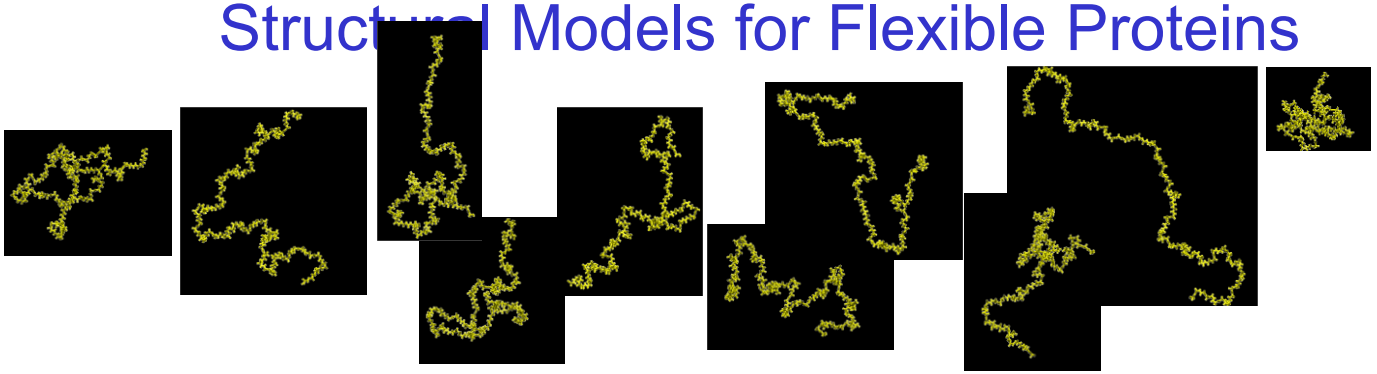
Ensemble Model

Ill-defined problem…

Data - Information - Model
Validation and cross-validation

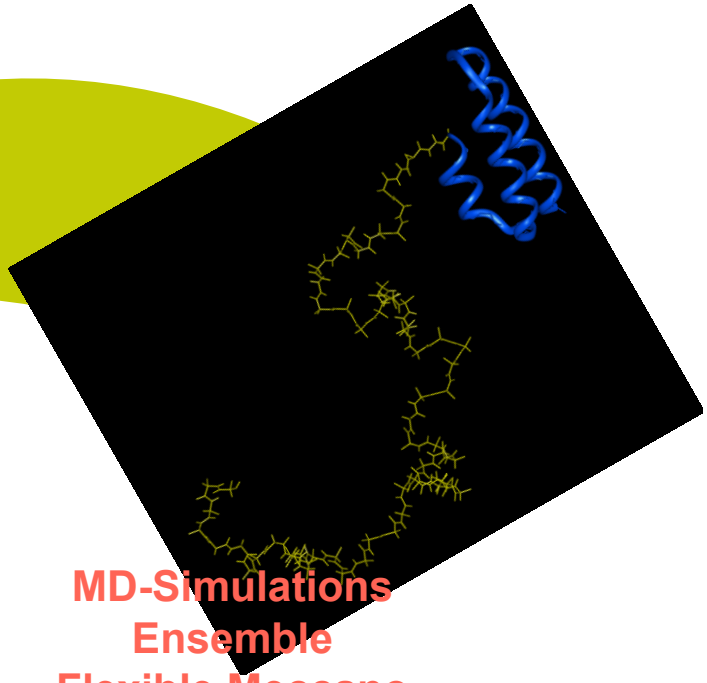
Needs for the Structural Characterization of Flexibility

Structural Models for Flexible Proteins



Calculation of NMR and SAXS Properties from Individual Conformations

Calculation of Properties from Conformations



SAXS:

CRYSOL, AXES, FOXS, AquaSAXS, pepsiSAXS, WAXSIS…

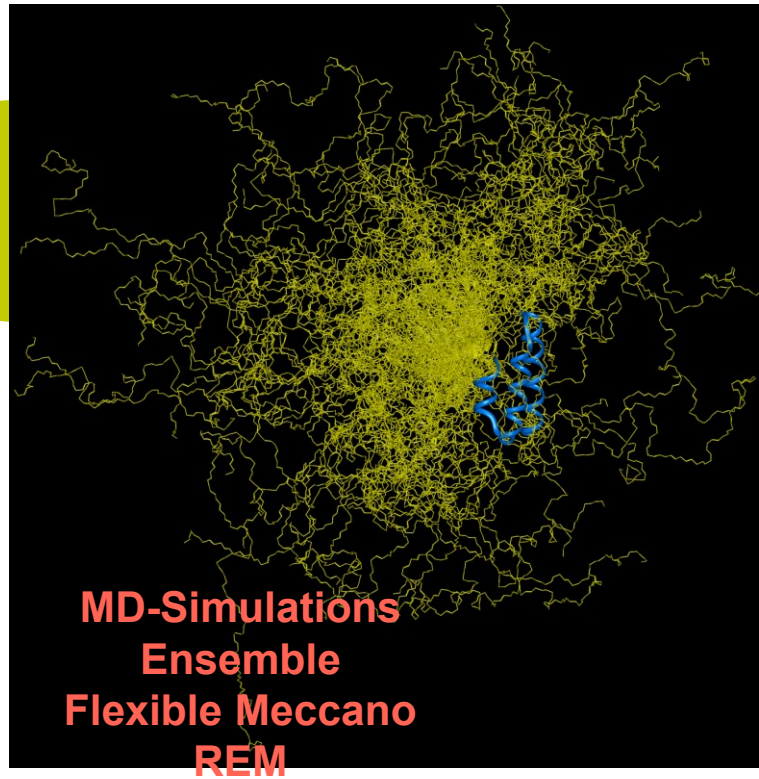
NMR:

CS: Sparta, Sparta+, ShiftX, CamShift
RDCs: PALES, Flexible-Meccano
PREs: Flexible-Meccano
Hydrodynamics: HydroPro

MD-Simulations
Ensemble
Flexible Meccano
REM
IDPConformerGenerator
CG-Simulations
Ranch…

PROPERTY

Calculation of Properties from Conformations



MD-Simulations
Ensemble
Flexible Meccano
REM
IDPConformerGenerator
CG-Simulations
Ranch…

SAXS:

CRYSOL, AXES, FOXS, AquaSAXS, pepsiSAXS, WAXSIS…

NMR:

CS: Sparta, Sparta+, ShiftX, CamShift
RDCs: PALES, Flexible-Meccano
PREs: Flexible-Meccano
Hydrodynamics: HydroPro

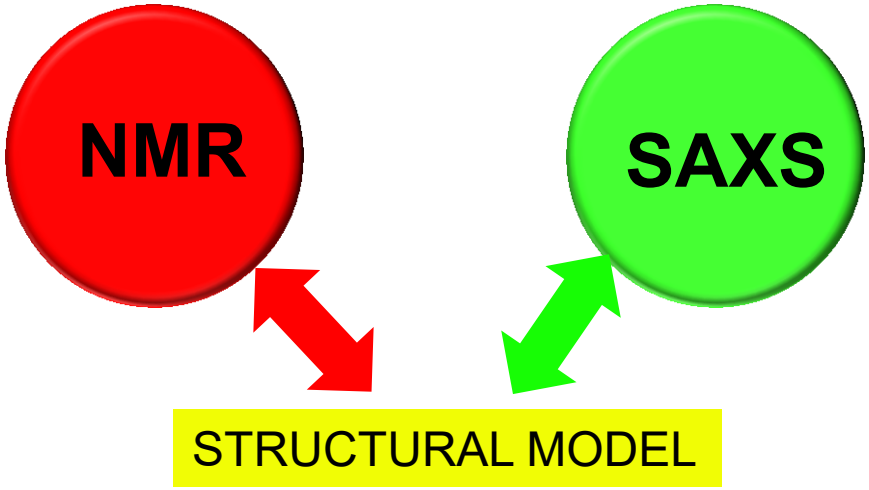
Warning!!!

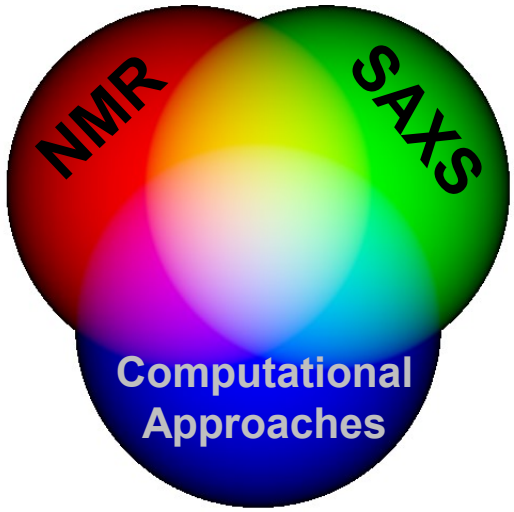
Not all properties can be averaged

Not all properties require the same number of conformations

AVERAGED PROPERTY

Two Philosophies

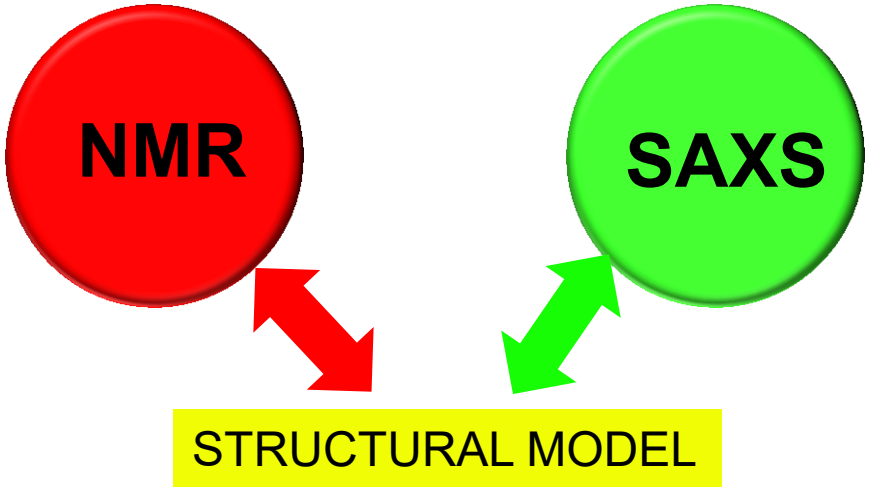




Model Validation

Data-Driven Modelling

Two Philosophies

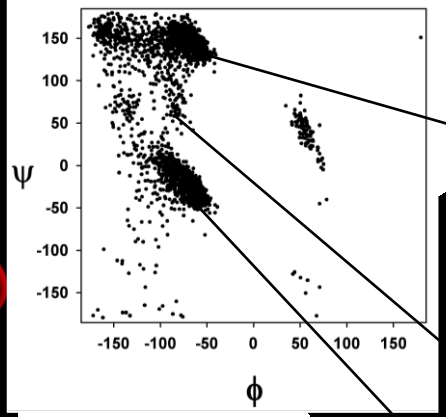


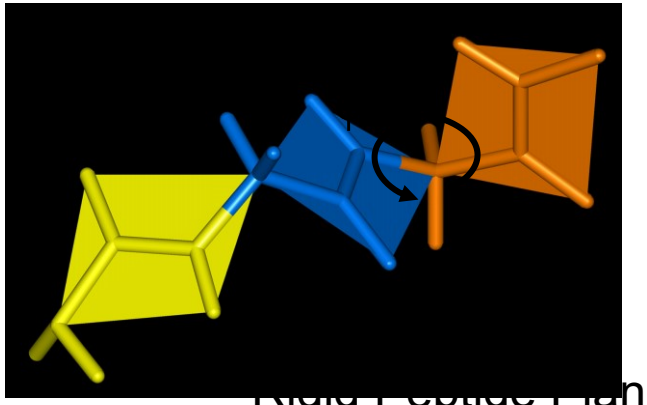
Model Validation

Coil Models to Generate Disordered Ensembles

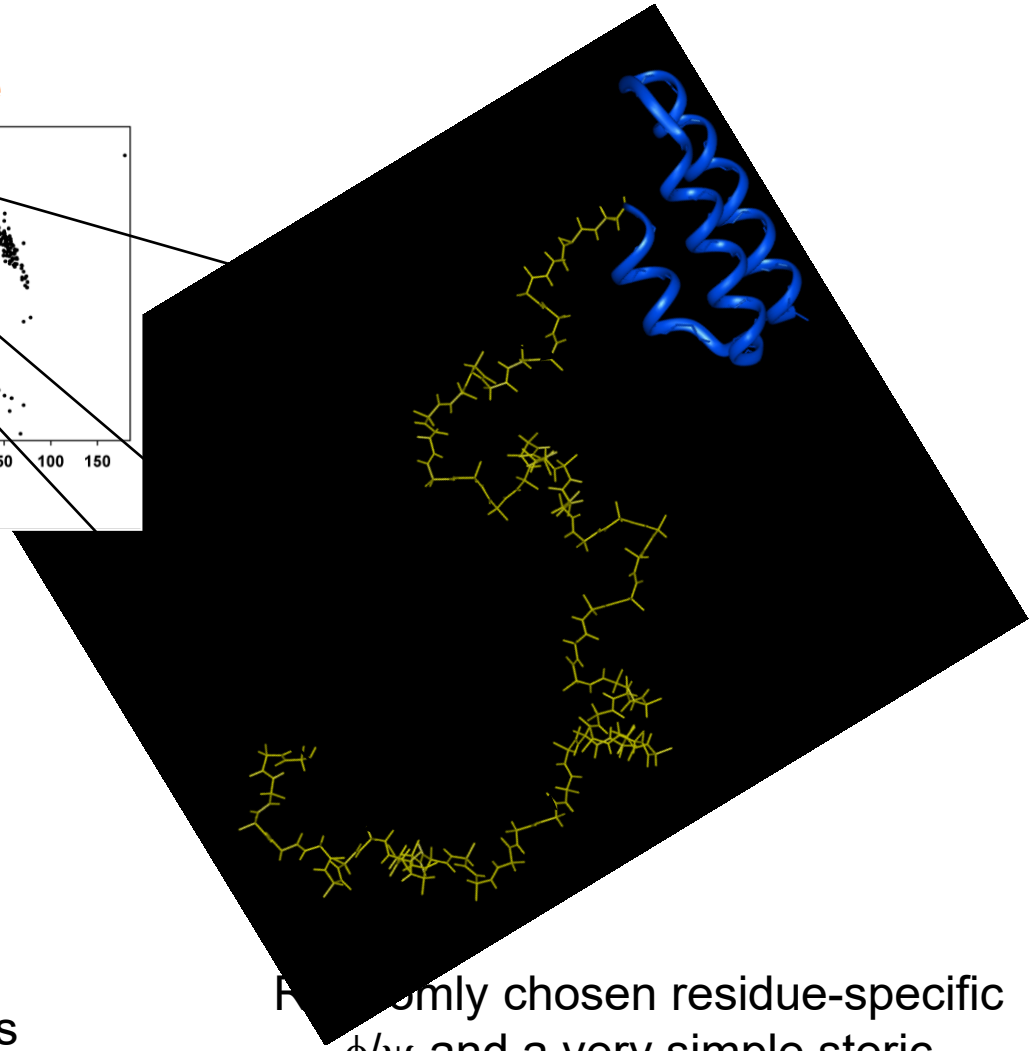
Coil Library

Alanine





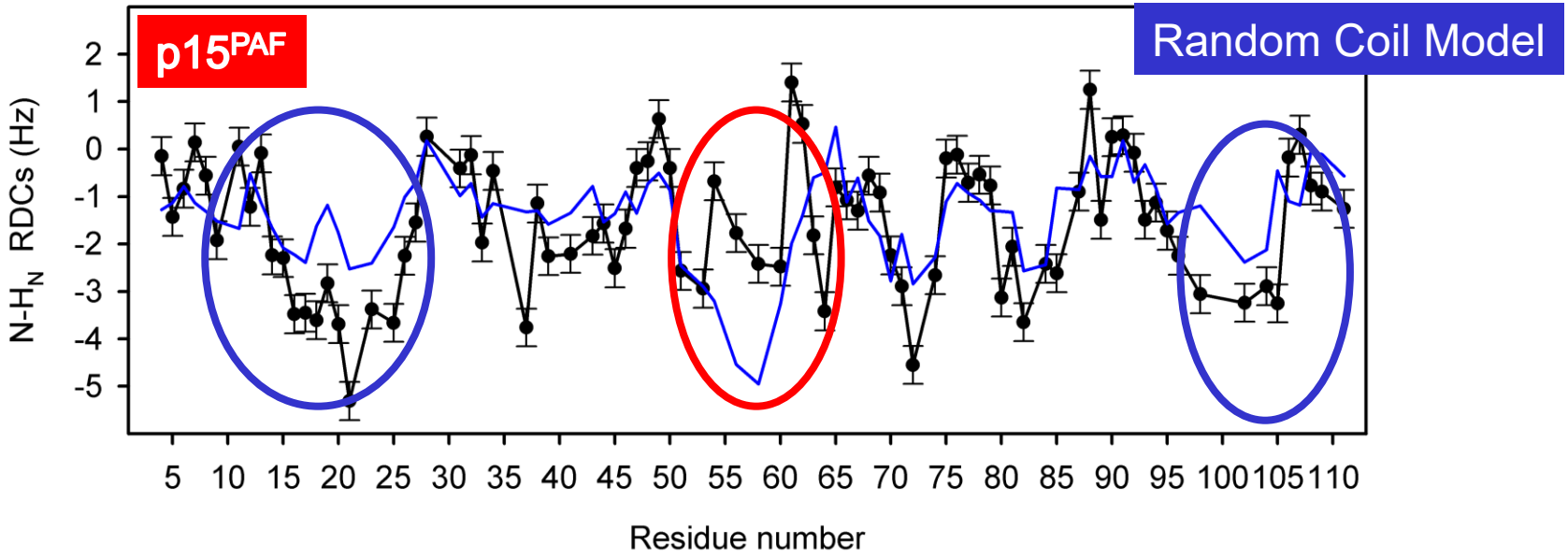
Rigid Peptide Planes



Randomly chosen residue-specific ϕ/ψ and a very simple steric potential are used to define a single conformation

Bernadó et al. *PNAS* 2005, 102, 17002
Ozenne et al. *Bioinformatics* 2012, 28, 463

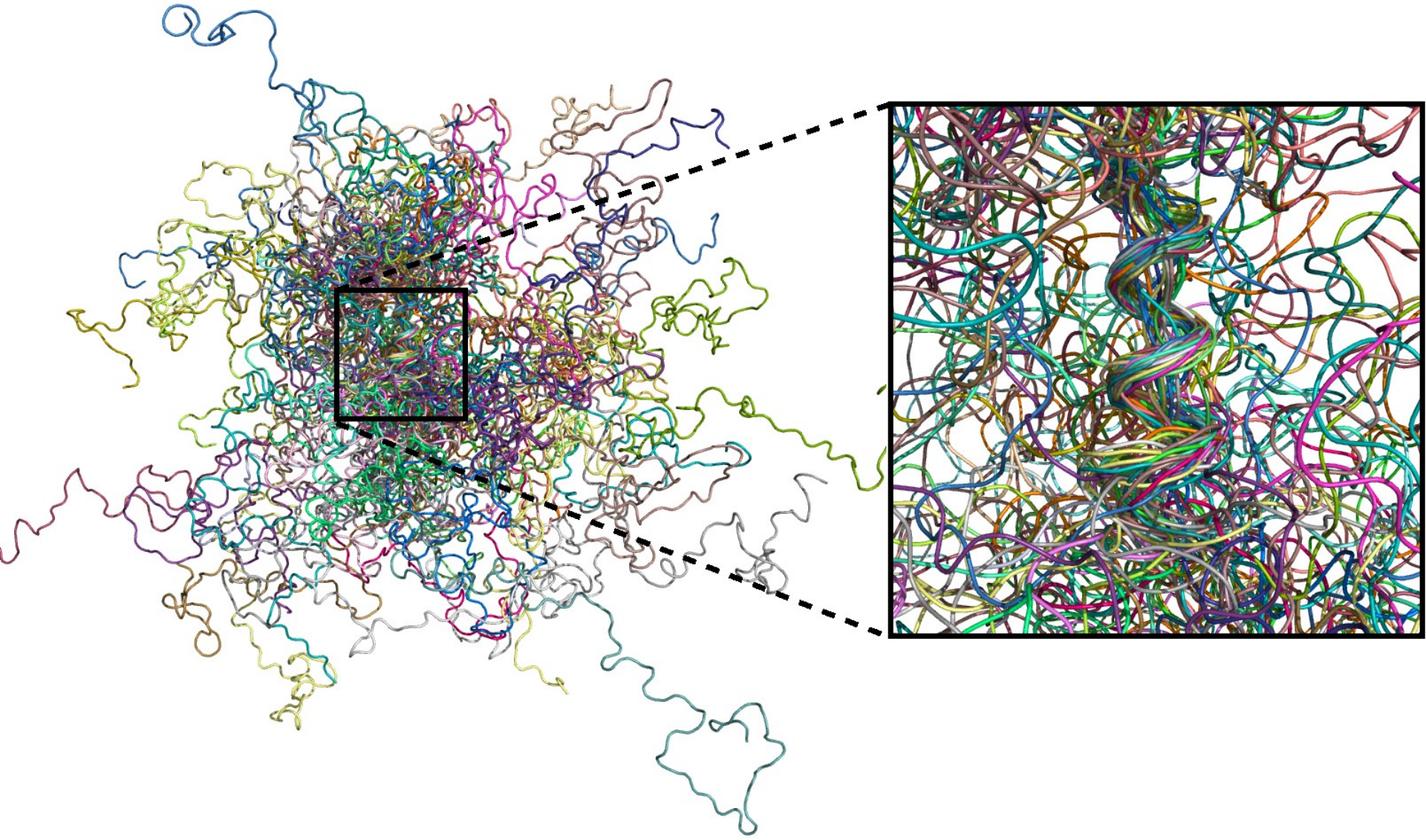
Interpreting RDCs with *FM*: Local Structural Information



Tiago Cordeiro - Col. Francisco Blanco (Biogune, Bilbao)

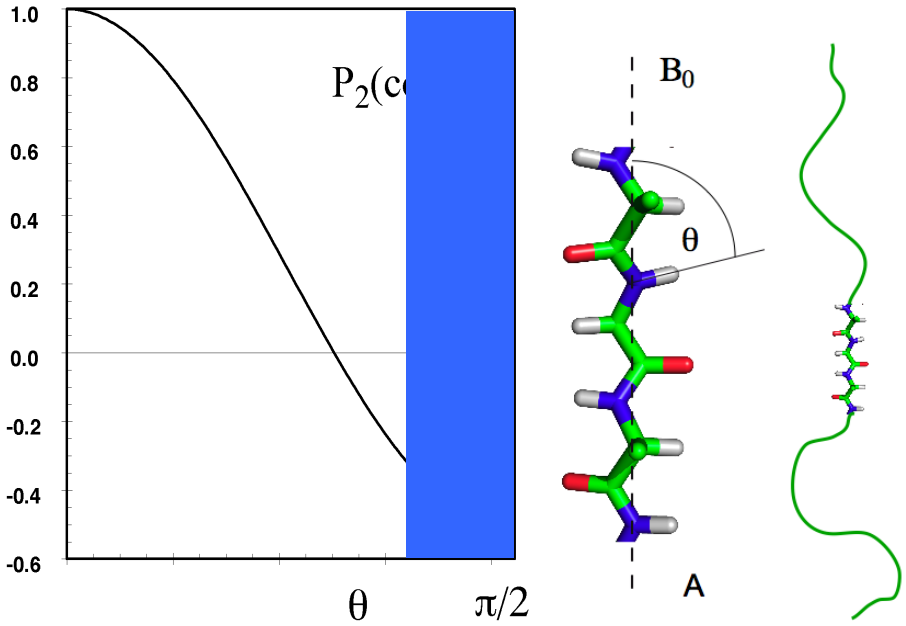
Di Biasio et al Biophys J. 2014

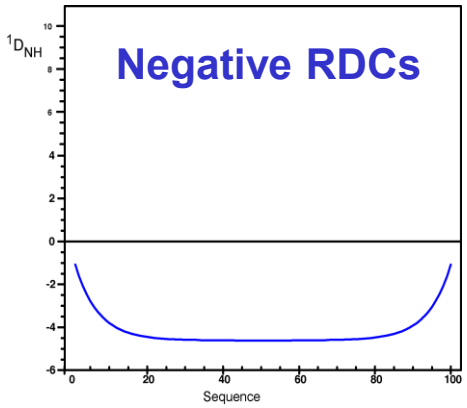
Partial Structure in Intrinsically Disordered Proteins



RDCs in Disordered States

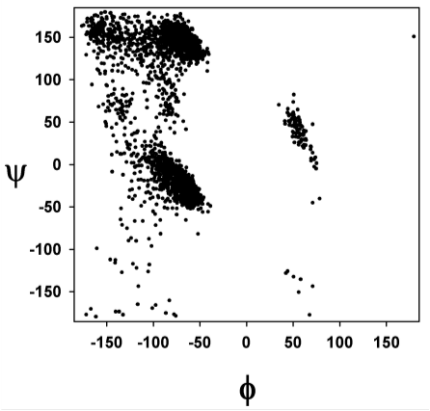
Disordered and Extended Regions

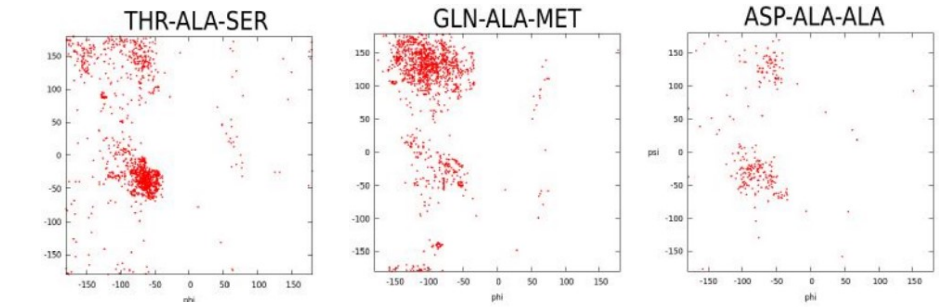




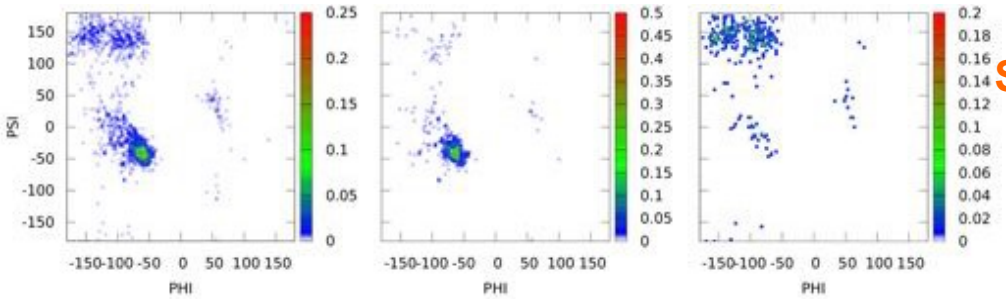
Realistic Models of IDPs – Tri-peptide Coil Model

Alanine



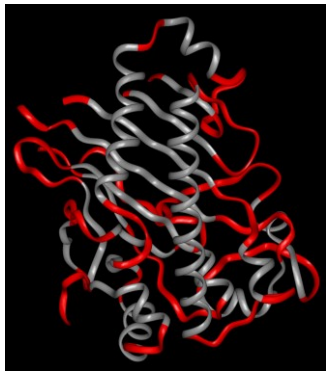


SEQUENCE- Dependent Conformation

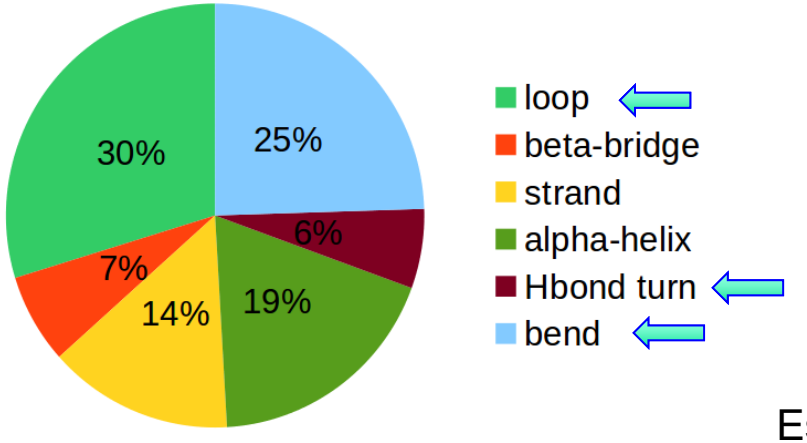


SER-ARG-ALA SER(α)-ARG-ALA(α) SER(β)-ARG-ALA(β)

STRUCTURE- Dependent Conformation



81.749 Proteins from SCOPe < 95% identity



COIL Database (LTS)

Tri-Peptides = 3.645.381

España et al. *Structure*, 2019, 27, 381

Application to NMR Residual Dipolar Couplings

Single-Residue Sampling (SRS)

M V K P G T F D P E M K...

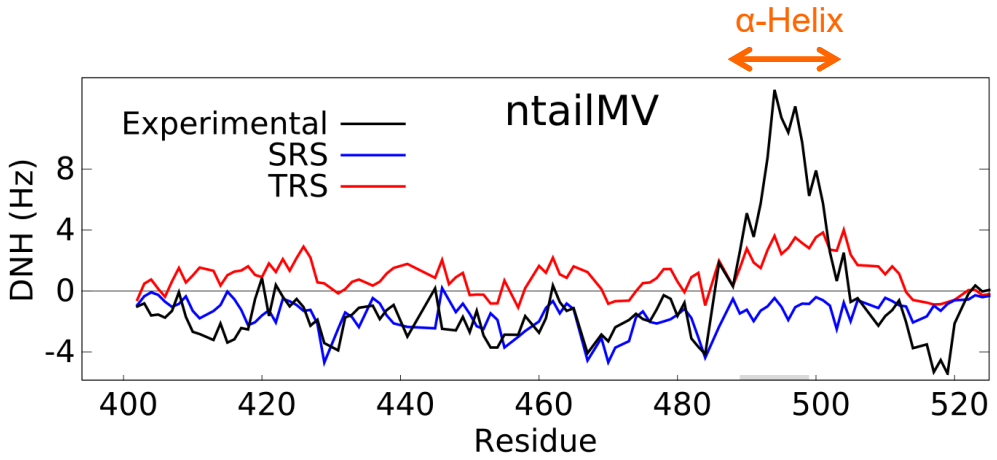
Tripeptide-Residue Sampling (TRS)

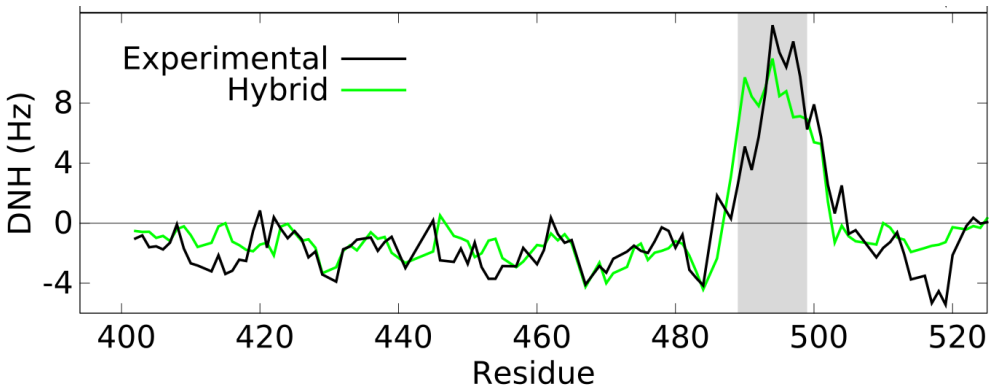
M V K P G T F D P E M K...

M V K P G T F D P E M K...

M V K P G T F D P E M K...

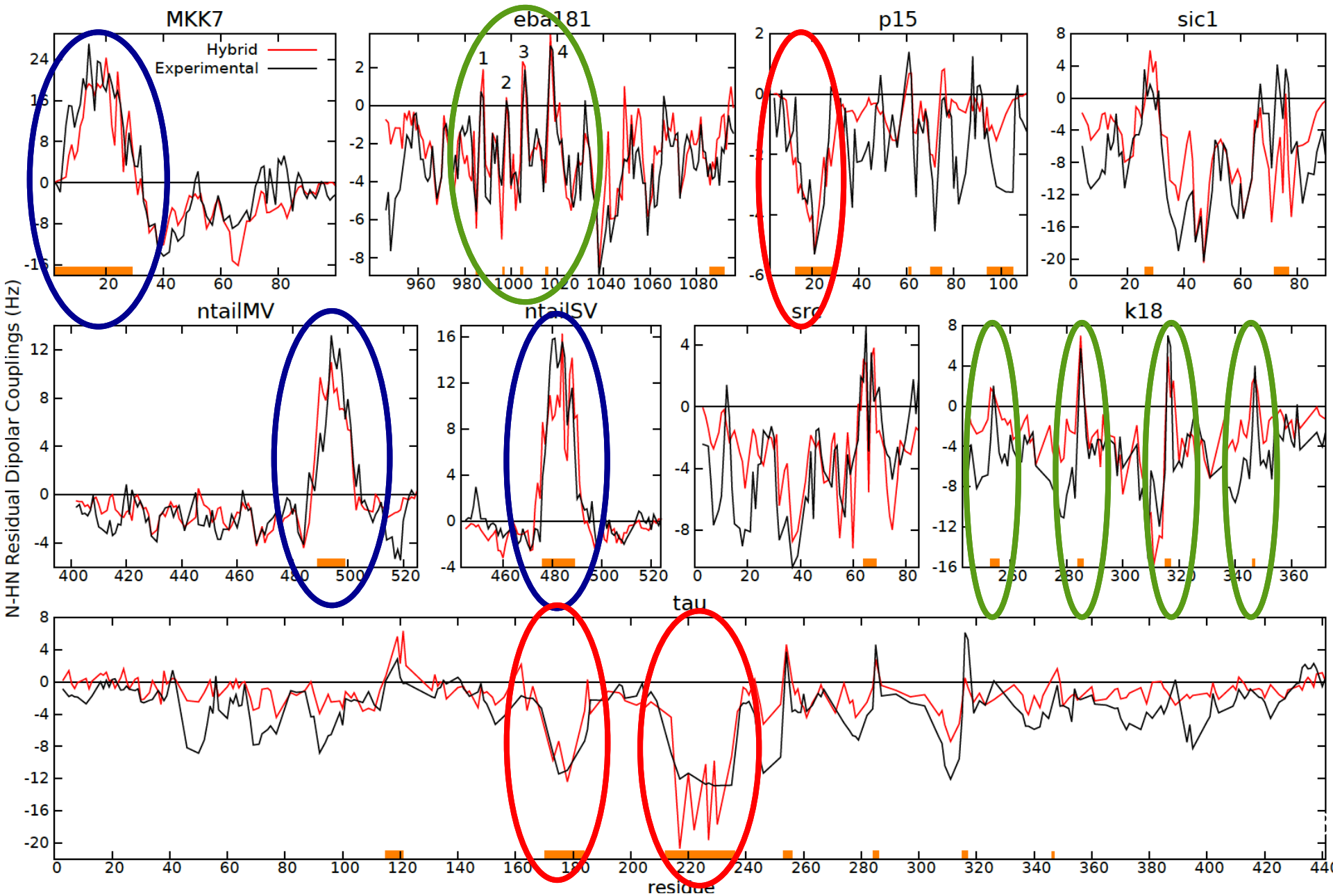
M V K P G T F D P E M K...



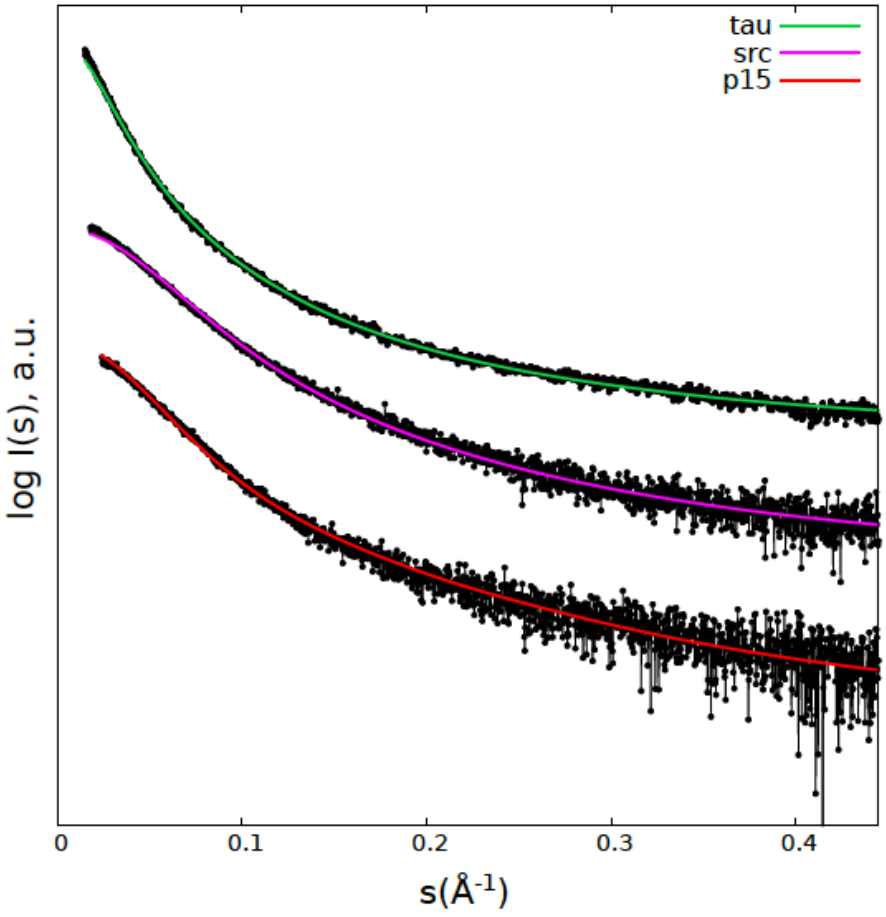


The appropriate combination of the SRS and TRS strategies provides an excellent description of the RDCs measured for ntail MV

Application to NMR Residual Dipolar Couplings



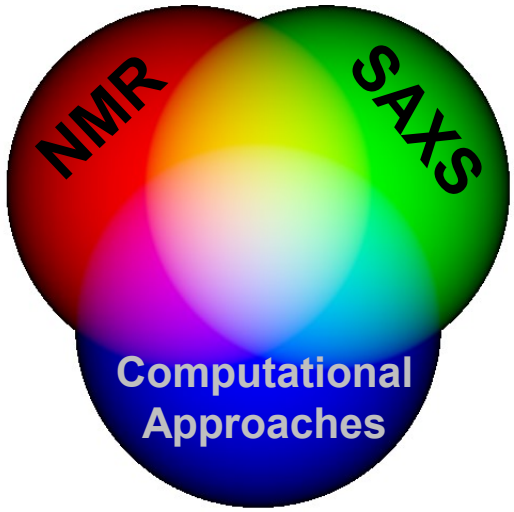
Application to SAXS curves



The ensembles generated are also in agreement with the SAXS curves

The building strategy based on the tri-peptide database produces realistic ensembles of IDPs that are compatible with NMR and SAXS data

Two Philosophies



Data-Driven Modelling

Ensemble Methods in SAXS

- The Ensemble Optimization Method (EOM)

Bernadó, Mylonas, Petoukhov, Blackledge, and Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 2007, **129**:5656-64.

- Minimal Ensemble Search (MES)

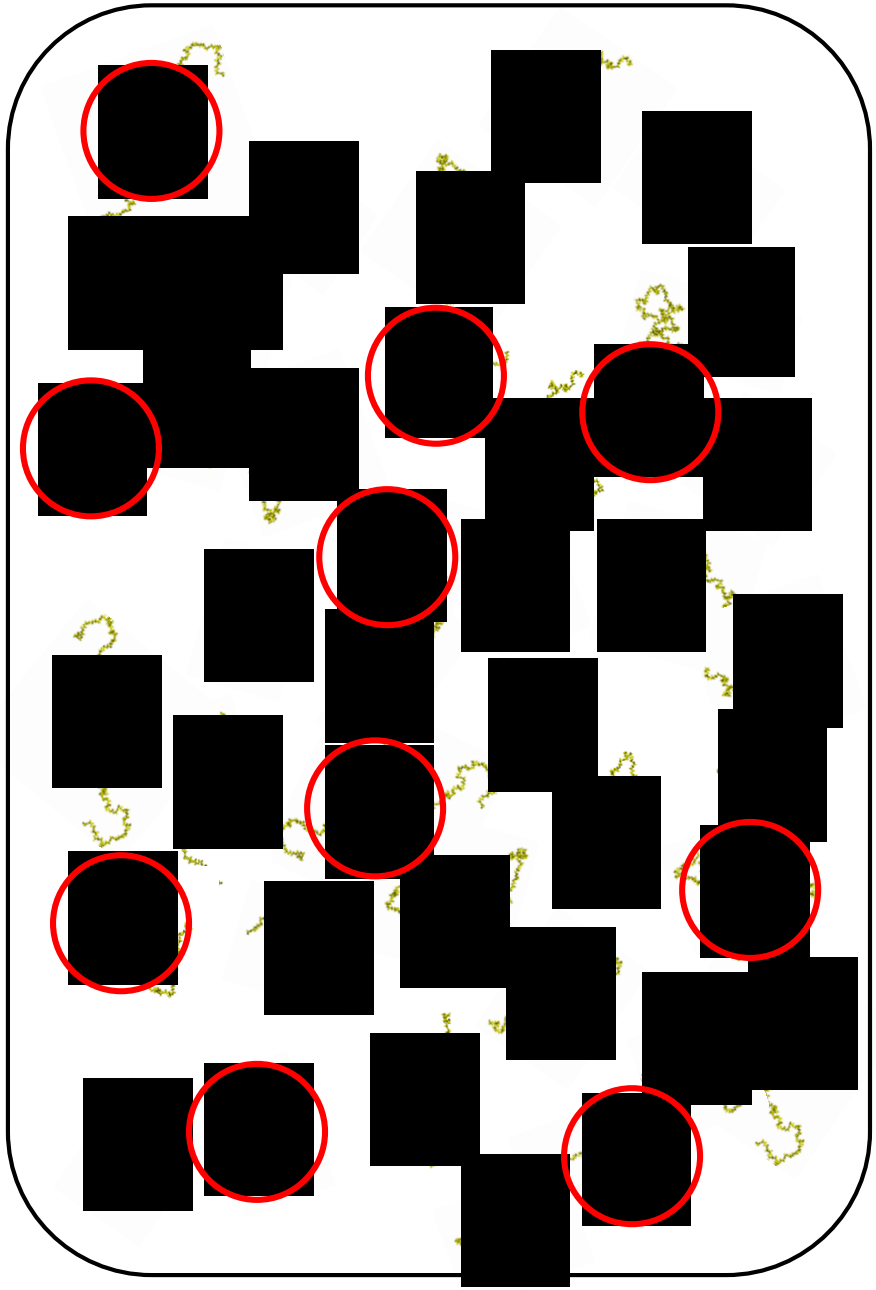
Pelikan, Hura, and Hammel. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* 2009, **28**:174–189.

-Basis-Set Supported SAXS (BSS-SAXS)

Yang, Blachowicz, Makowski, Roux. Multidomain assembled sates of Hck tyrosine kinase in solution. *PNAS* 2010, **107**:15757-15762.

-Ensemble Refinement of SAXS (EROS)

Rozycki, Kim, Hummer. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure*, 2011, **19**:109-116.



1. Calculation of a conformational ensemble
2. Computation of the theoretical SAXS profiles
3. Optimization of a subensemble
4. Structural interpretation of the subensemble

Ensemble Methods in SAXS

- The Ensemble Optimization Method (EOM)

Bernadó, Mylonas, Petoukhov, Blackledge, and Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 2007, **129**:5656-64.

- Minimal Ensemble Search (MES)

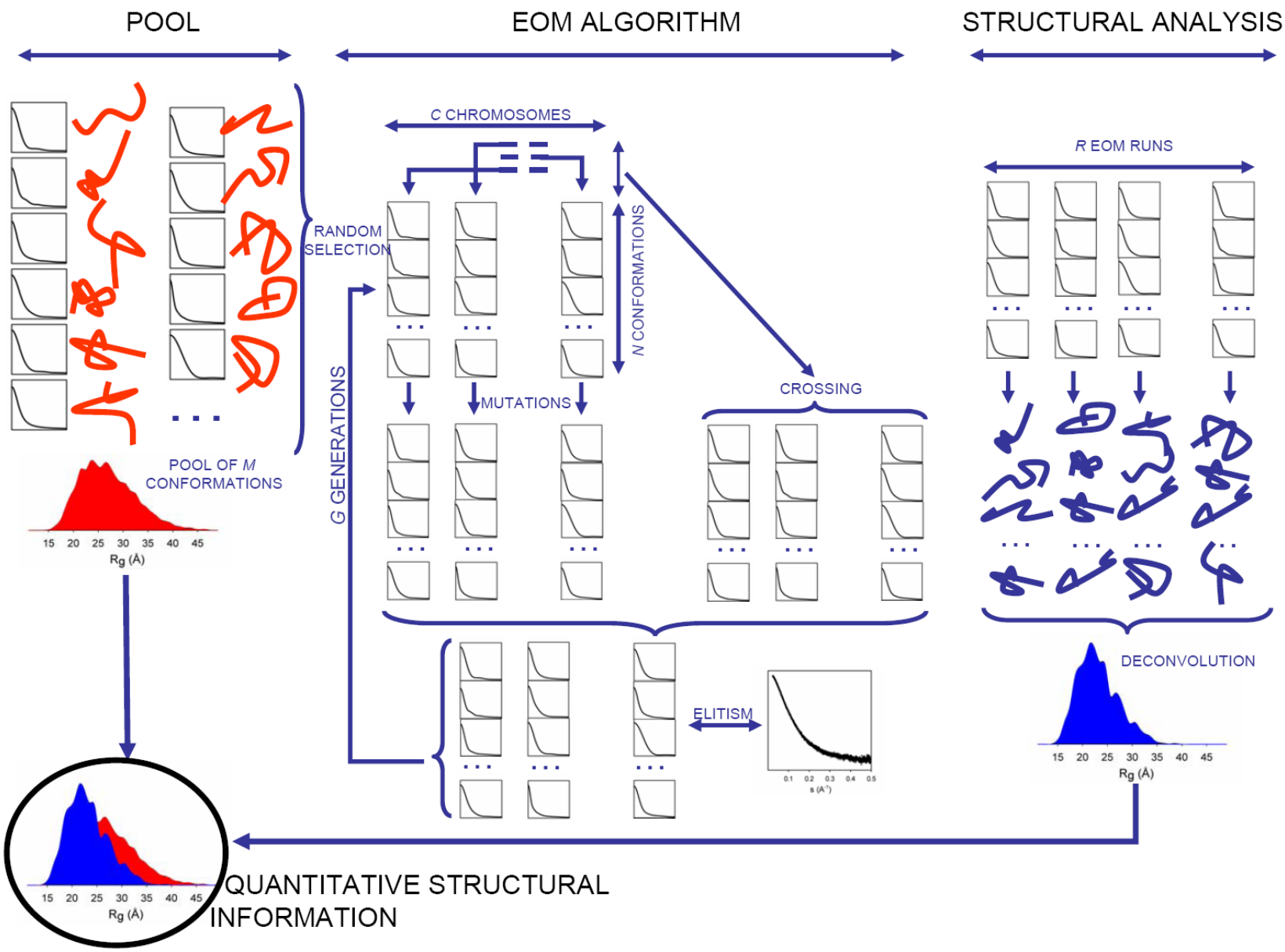
Pelikan, Hura, and Hammel. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen Physiol Biophys* 2009, **28**:174–189.

-Basis-Set Supported SAXS (BSS-SAXS)

Yang, Blachowicz, Makowski, Roux. Multidomain assembled sates of Hck tyrosine kinase in solution. *PNAS* 2010, **107**:15757-15762.

-Ensemble Refinement of SAXS (EROS)

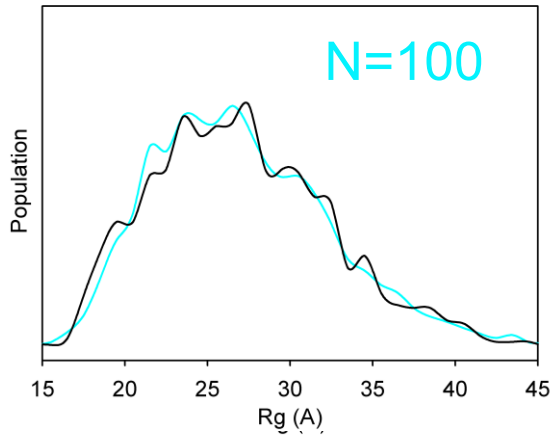
Rozycki, Kim, Hummer. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure*, 2011, **19**:109-116.



Bernadó, Svergun et al. *J.Am.Chem.Soc.* 2007, 129, 5656-5664.

Structural Meaning of the Selected Conformations

How many structures have to be used? It depends on the degree of flexibility



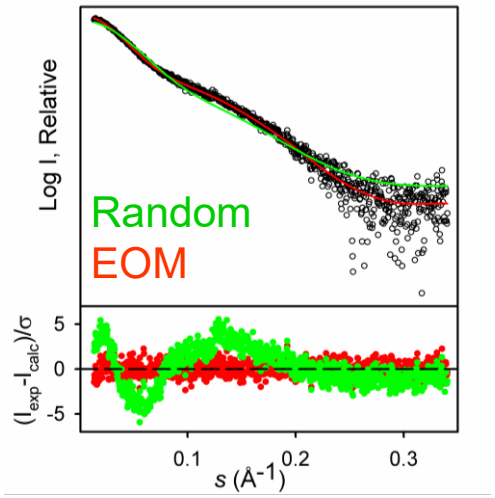
It is tempting to look at the structures at atomic/residue level… Don't do this because (Remember that) SAXS is a low resolution technique and the information content is limited… Only the size/shape of the molecule can be recovered

Structures collected are simply a TOOL to describe the size/shape distributions...

If certain structure is collected at each run… It does not mean that it is prevalent in solution

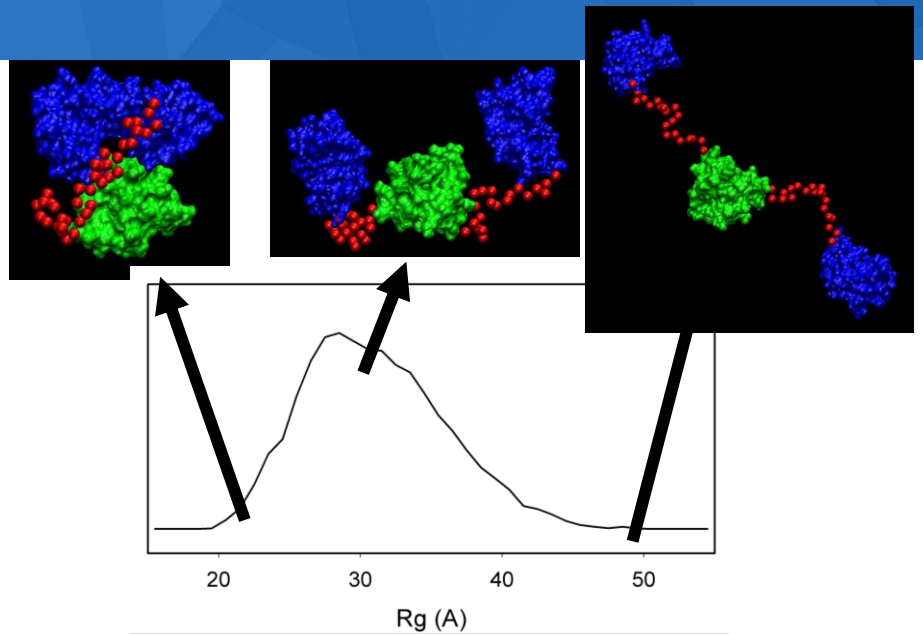
Structural interpretation can ONLY be done if the fit is from excellent to perfect!!!

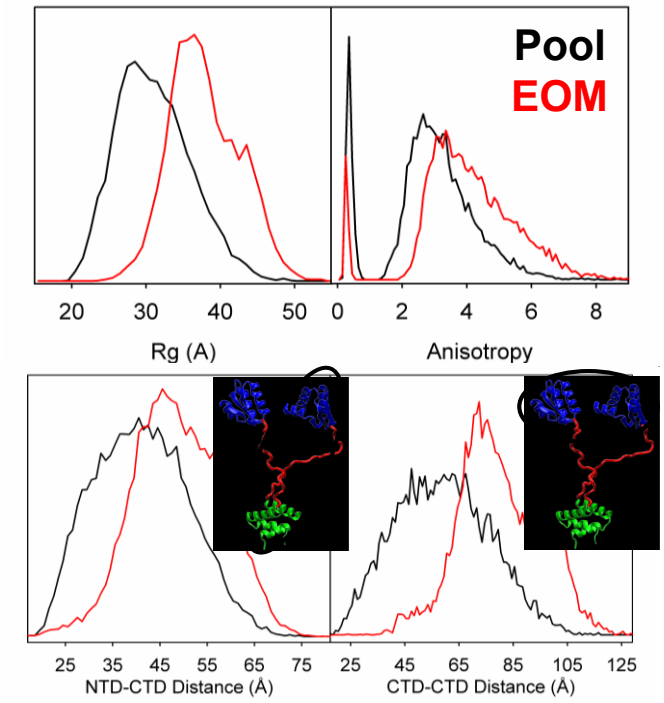
Ribosomal L12



Three samples at 7, 15 and 20 mg/ml were measured and merged

Data measured at X33 Beamline in DESY-Hamburg





Although highly dynamic, L12 samples a reduced conformational space, and mainly adopts highly elongated (anisotropic) structures.

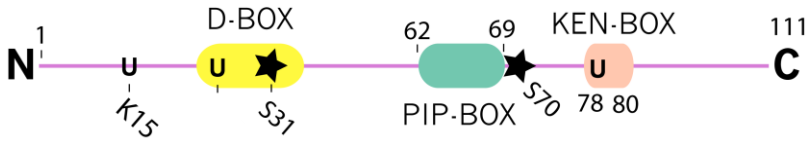
This induces large distances between both CTD domains that leave the NTD dimer in the middle

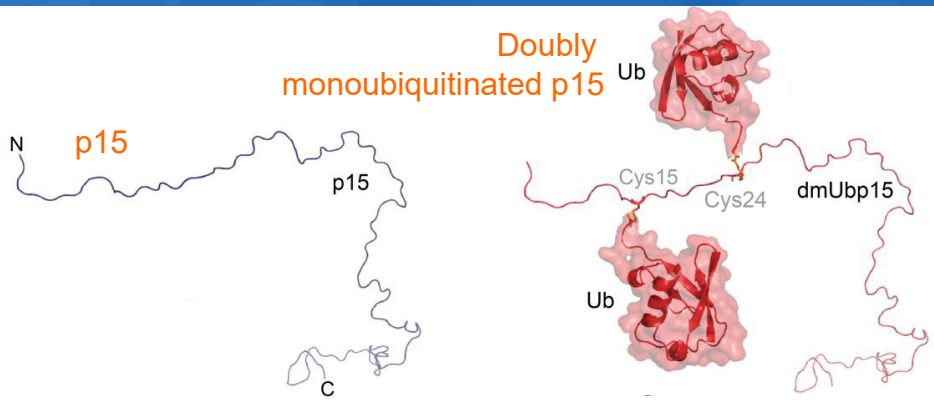
The linker is partially structured

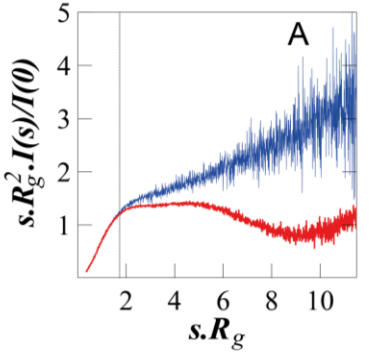
Bernadó et al. Biophys. J. 2010, 98, 2374

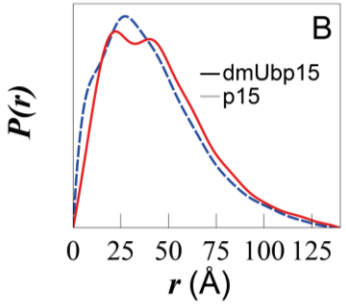
Col. D. Svergun, M. Akke, M. Tchorzewski

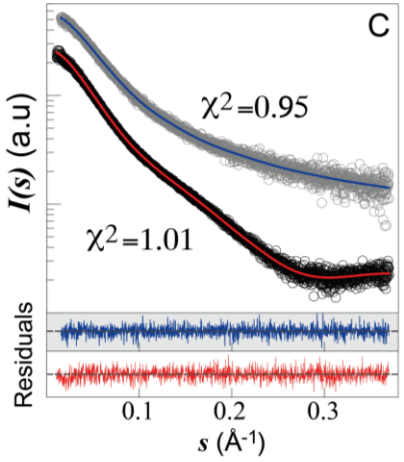
Doubly Monoubiquitinated p15

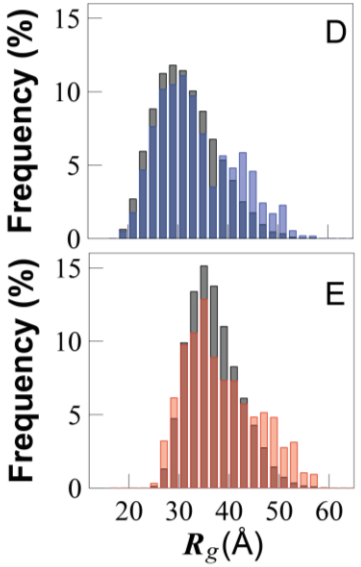


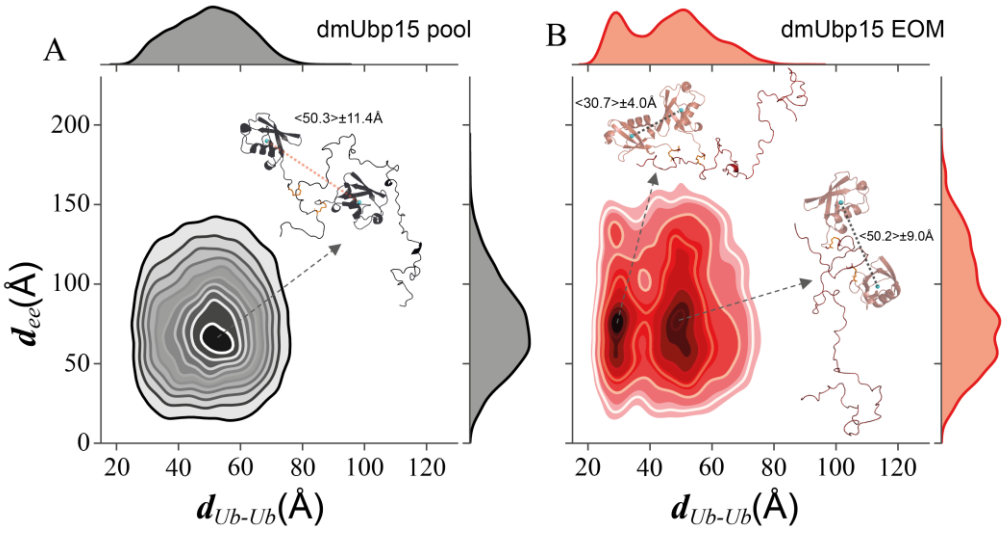












The EOM analysis indicates that both ubiquitins form transient dimers when grafted to p15

Coll. Francisco Blanco (CICBiogune-Bilbao) and Tiago Cordeiro (ITQB-Lisbon)

González-Magaña et al. ACS Chem Biol, 2019

FACritizism about EOM

"*With EOM you can fit an elephant if you want*…"

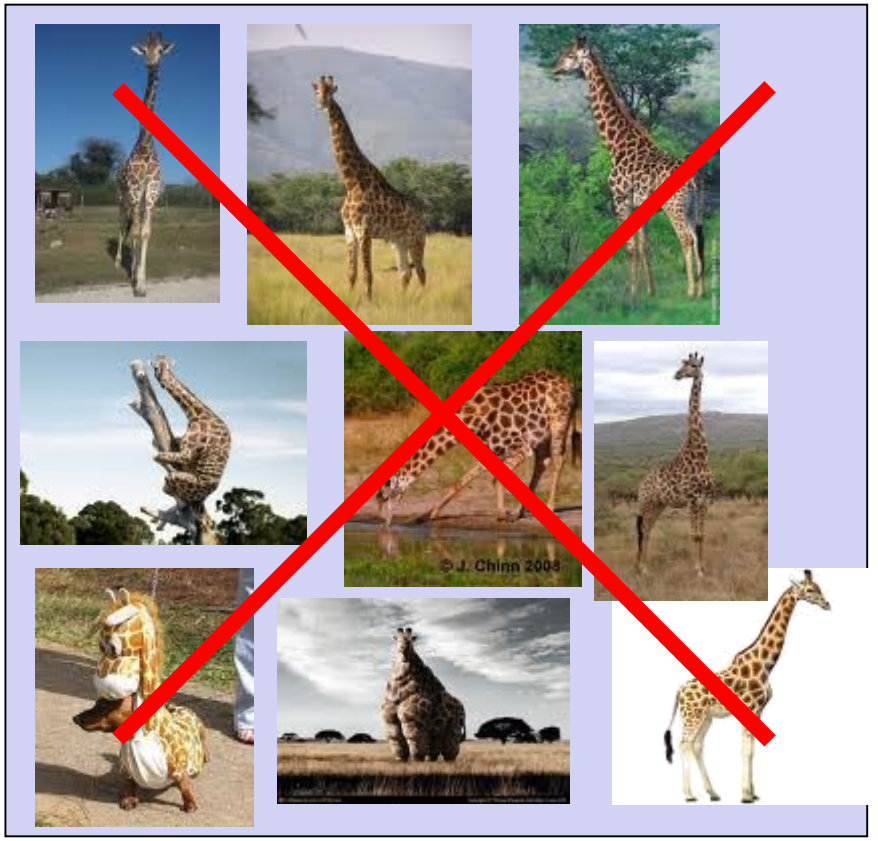
Maria Garcia-Parajo

FACritizism about EOM

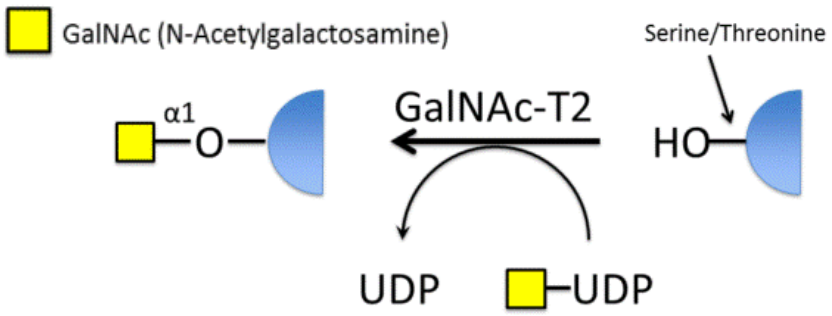
"With EOM you can fit an elephant if you want…"

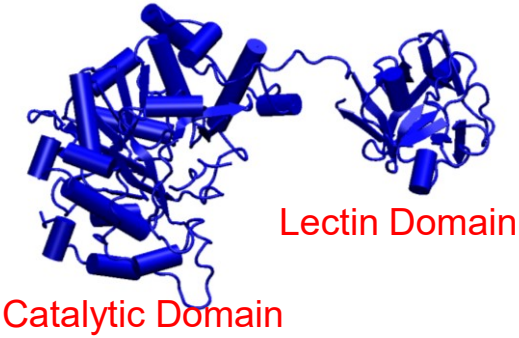
Maria Garcia-Parajo

THIS IS NOT TRUE

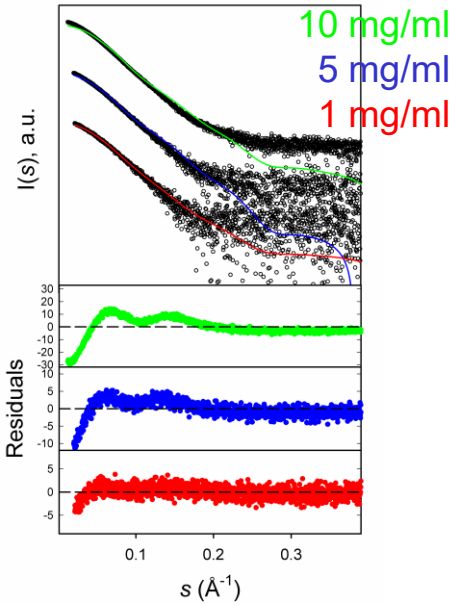


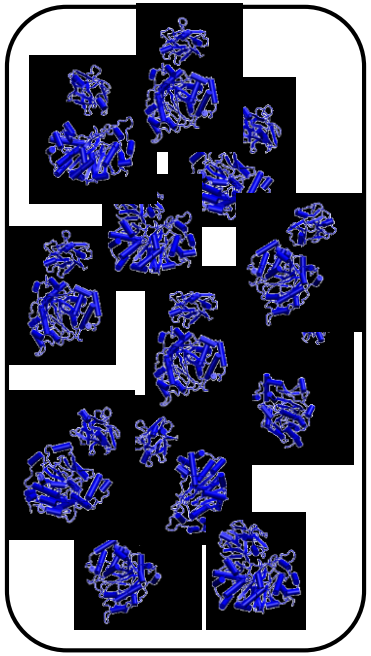
GalNAc-T2





EOM (single Chain)

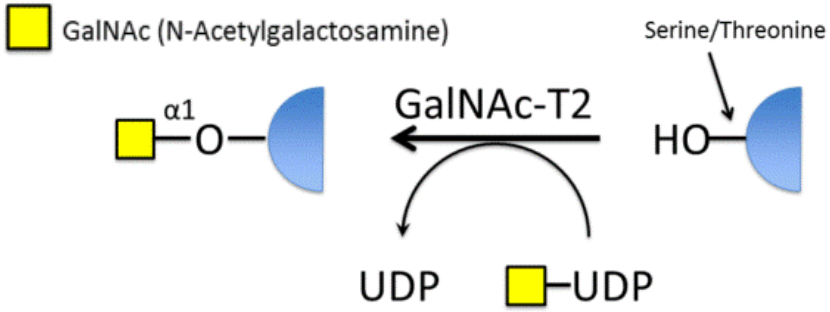


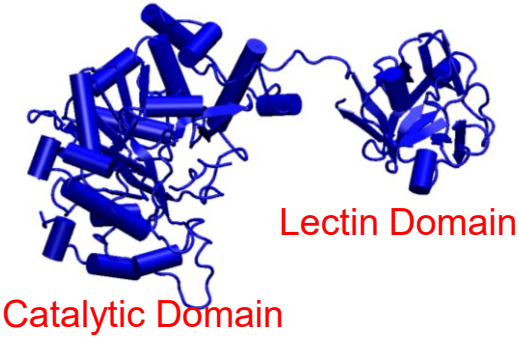


Col. Ramón Hurtado (Zaragoza-Spain)

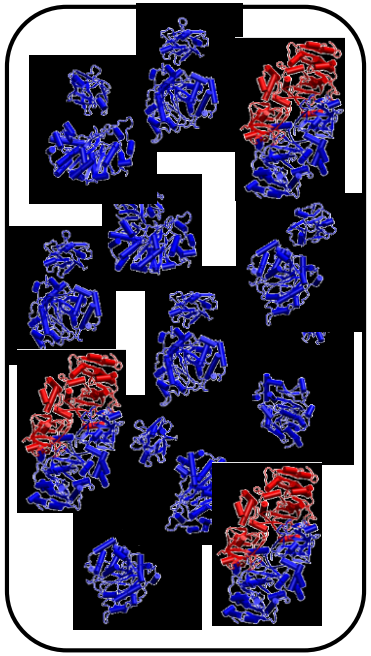
Lira-Navarrete et al. Nat. Comm 2015

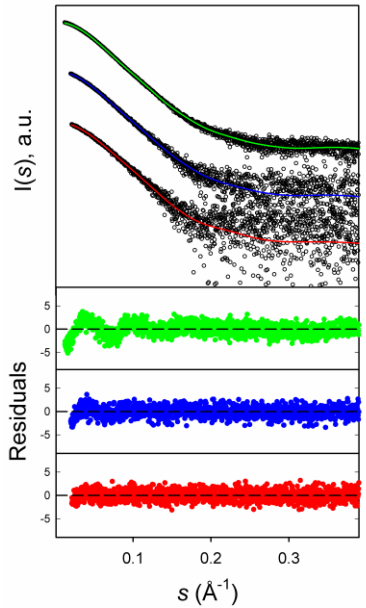
GalNAc-T2

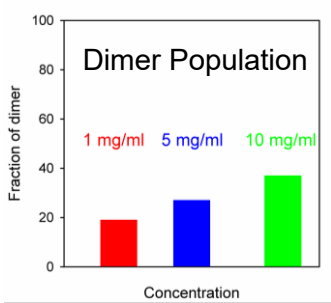


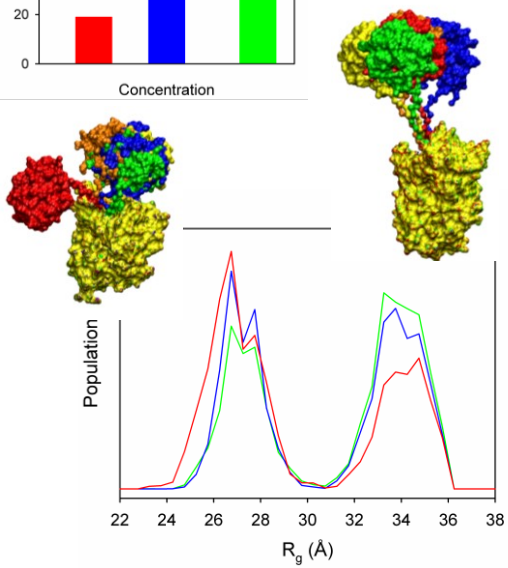


EOM (with dimers)





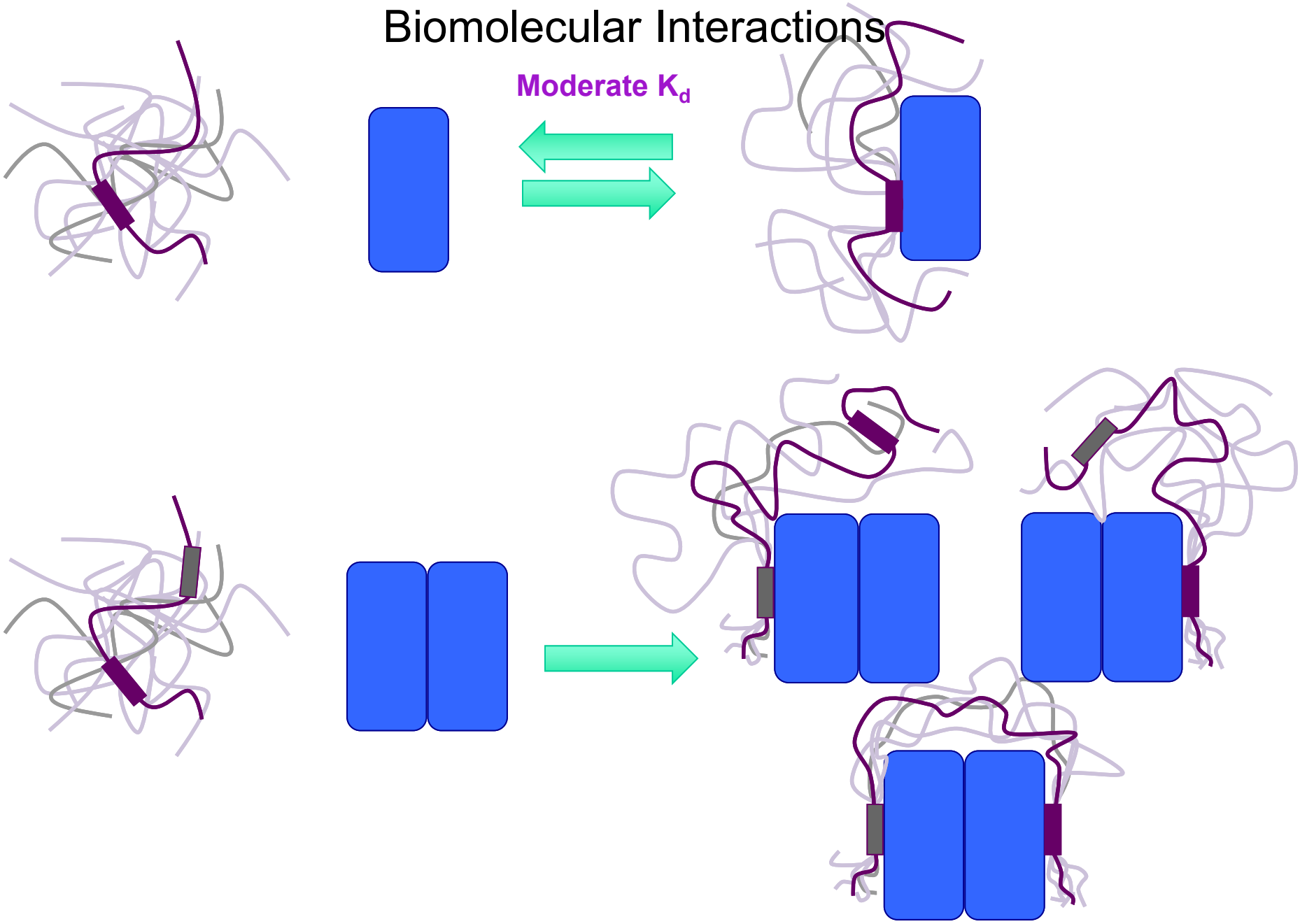




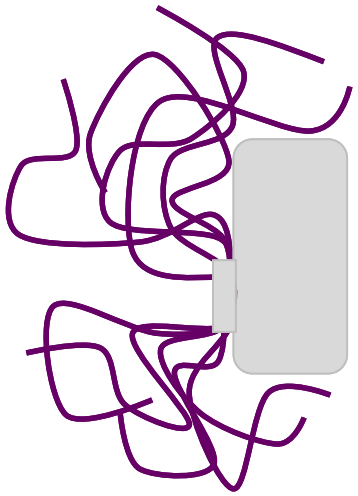
Col. Ramón Hurtado (Zaragoza-Spain)

Lira-Navarrete et al. Nat. Comm 2015

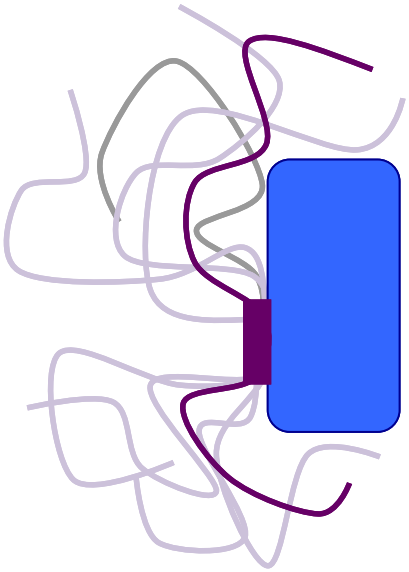
Biomolecular Interactions

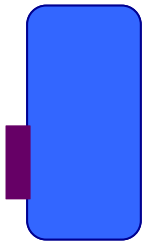


A Structural Perspective



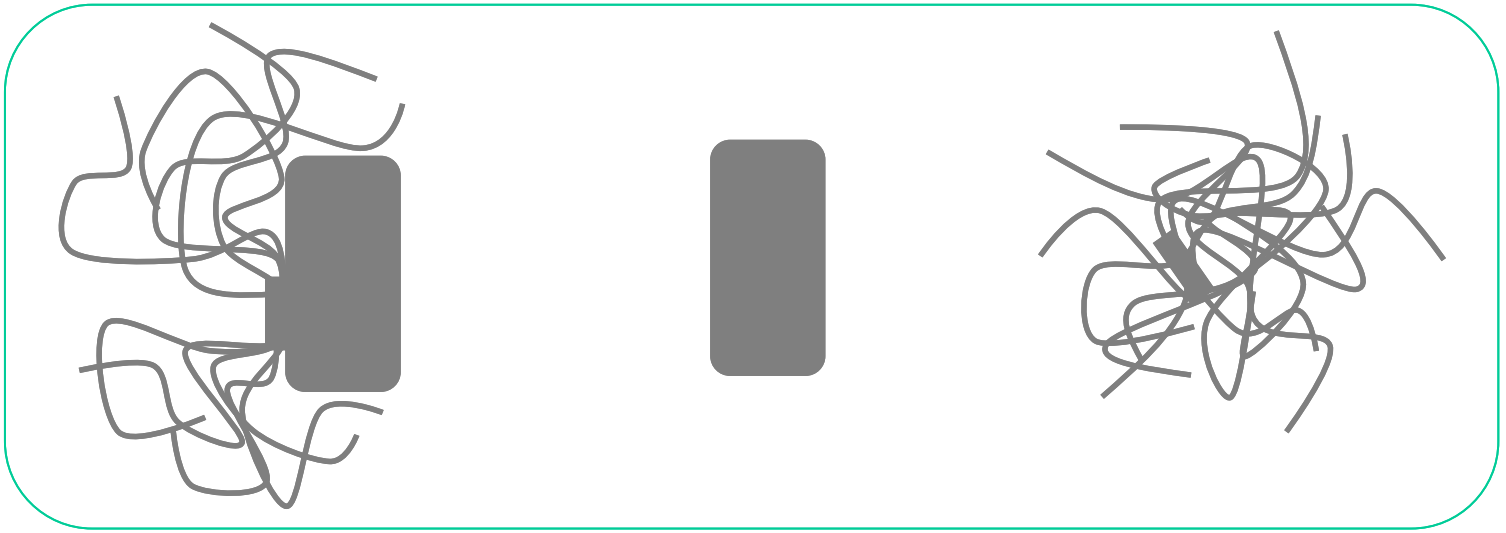
NMR



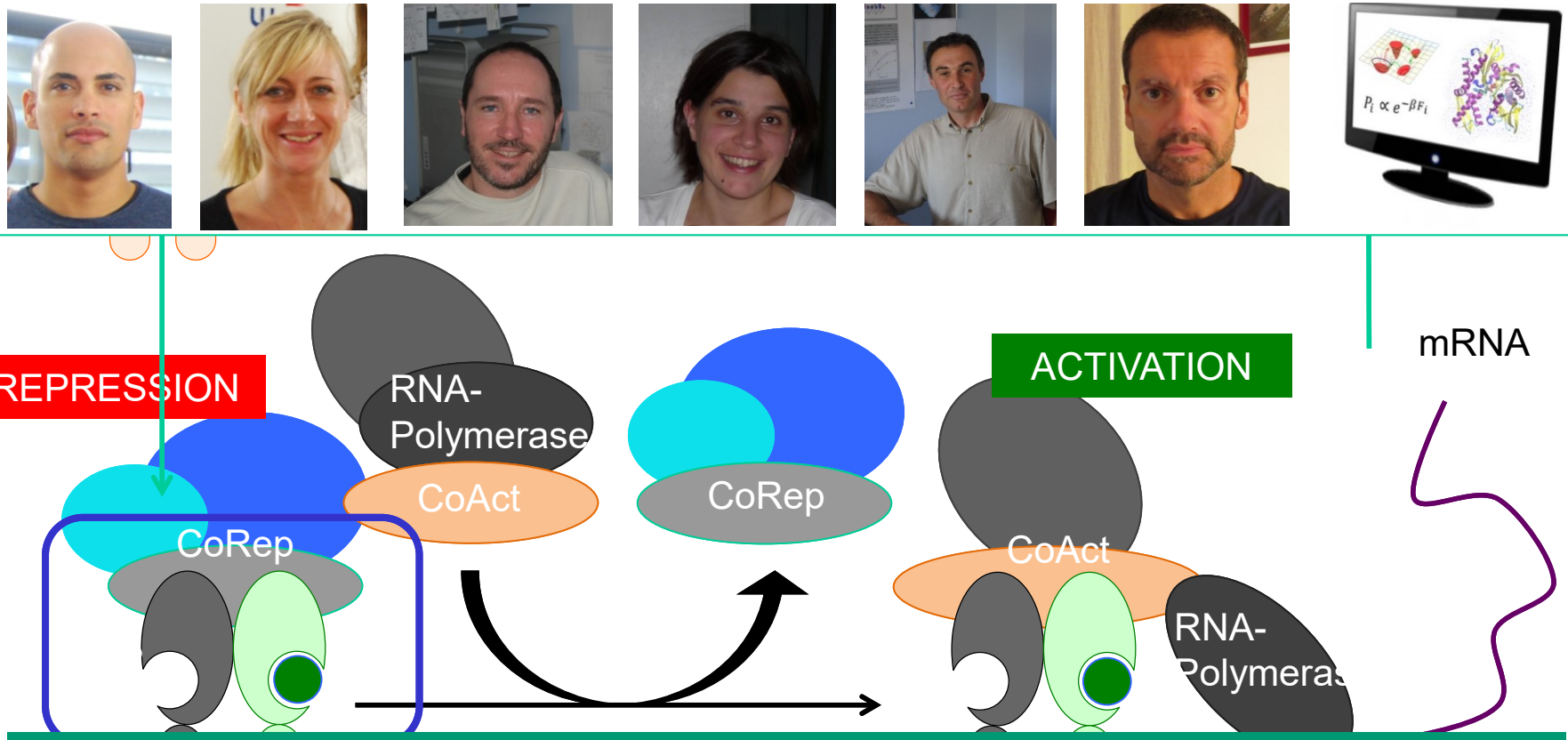


Crystallography

SAXS

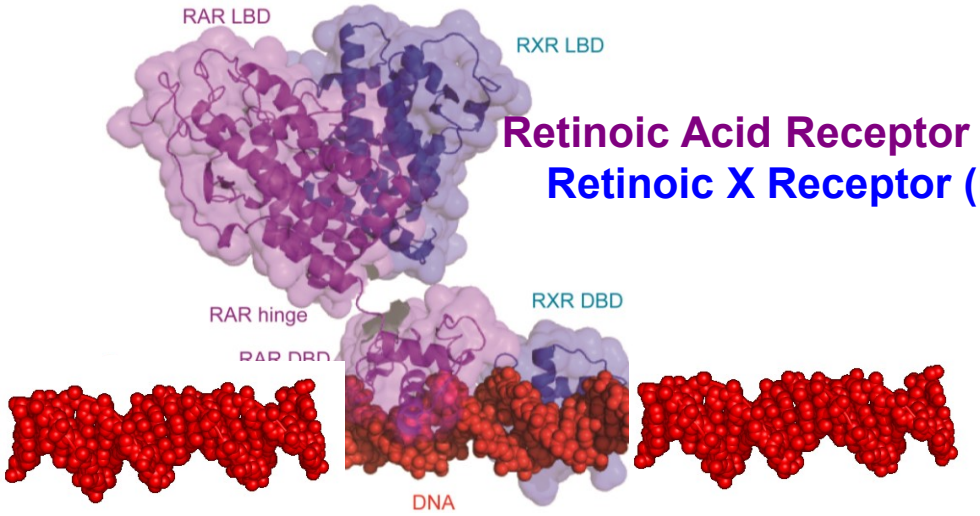


Regulation of the Retinoic Acid Nuclear Receptor

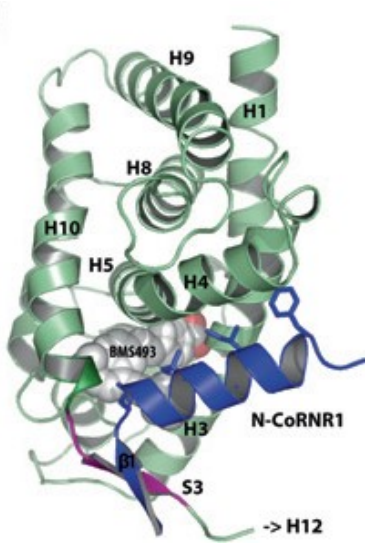


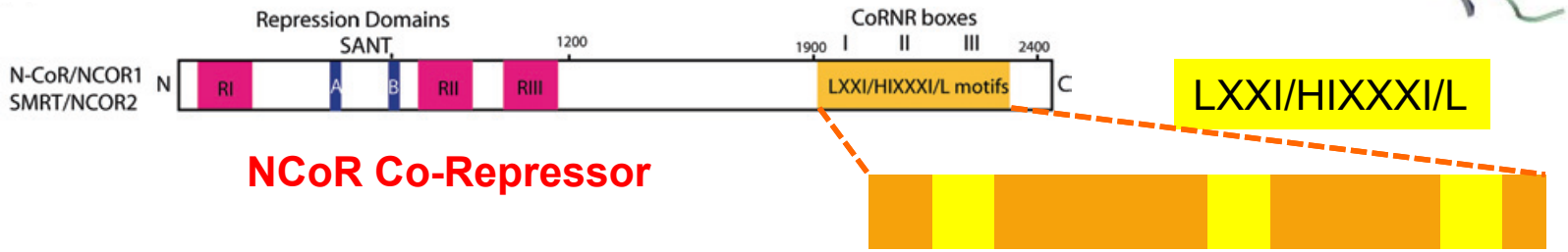
Cordeiro, T.N. ; Sibille, N. ; Germain, P. ; Barthe, P. ; Boulahtouf, A. ; Allemand, F. ; Bailly, R. ; Vivat, V. ; Ebel, C. ; Barducci, A. ; Bourguet, W. ; le Maire, A.* ; Bernadó, P.* *Interplay of protein disorder in retinoic acid receptor heterodimer and its corepressor regulates gene expression,* **Structure**, 2019, 27, 1270.

Retinoic Acid Receptor (RAR) / Retinoic X Receptor (RXR)

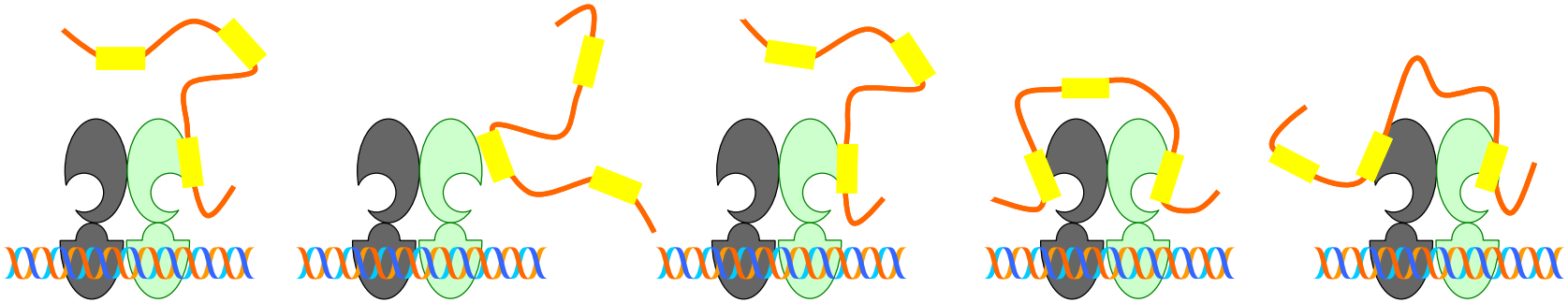


Retinoic Acid Receptor (RAR)
Retinoic X Receptor (RXR)

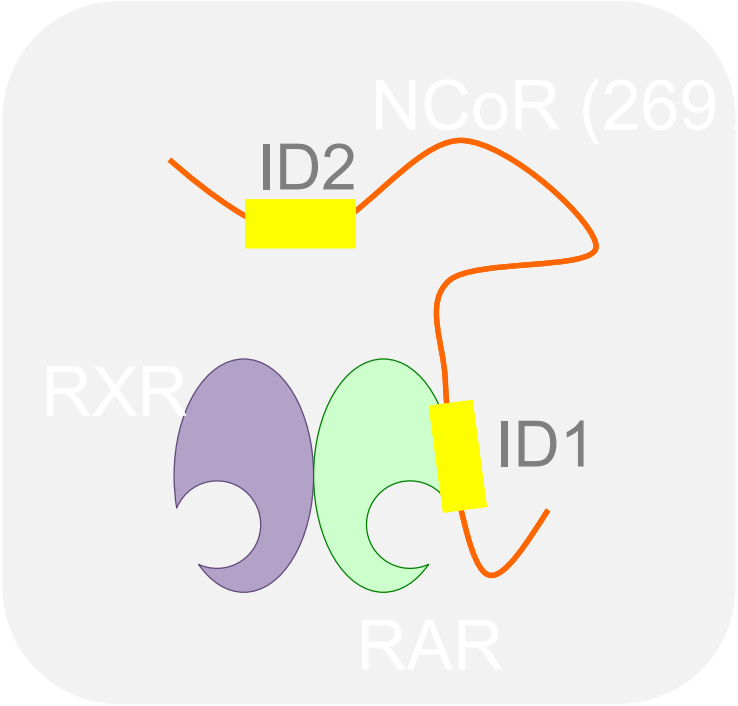




NCoR Co-Repressor



Questions

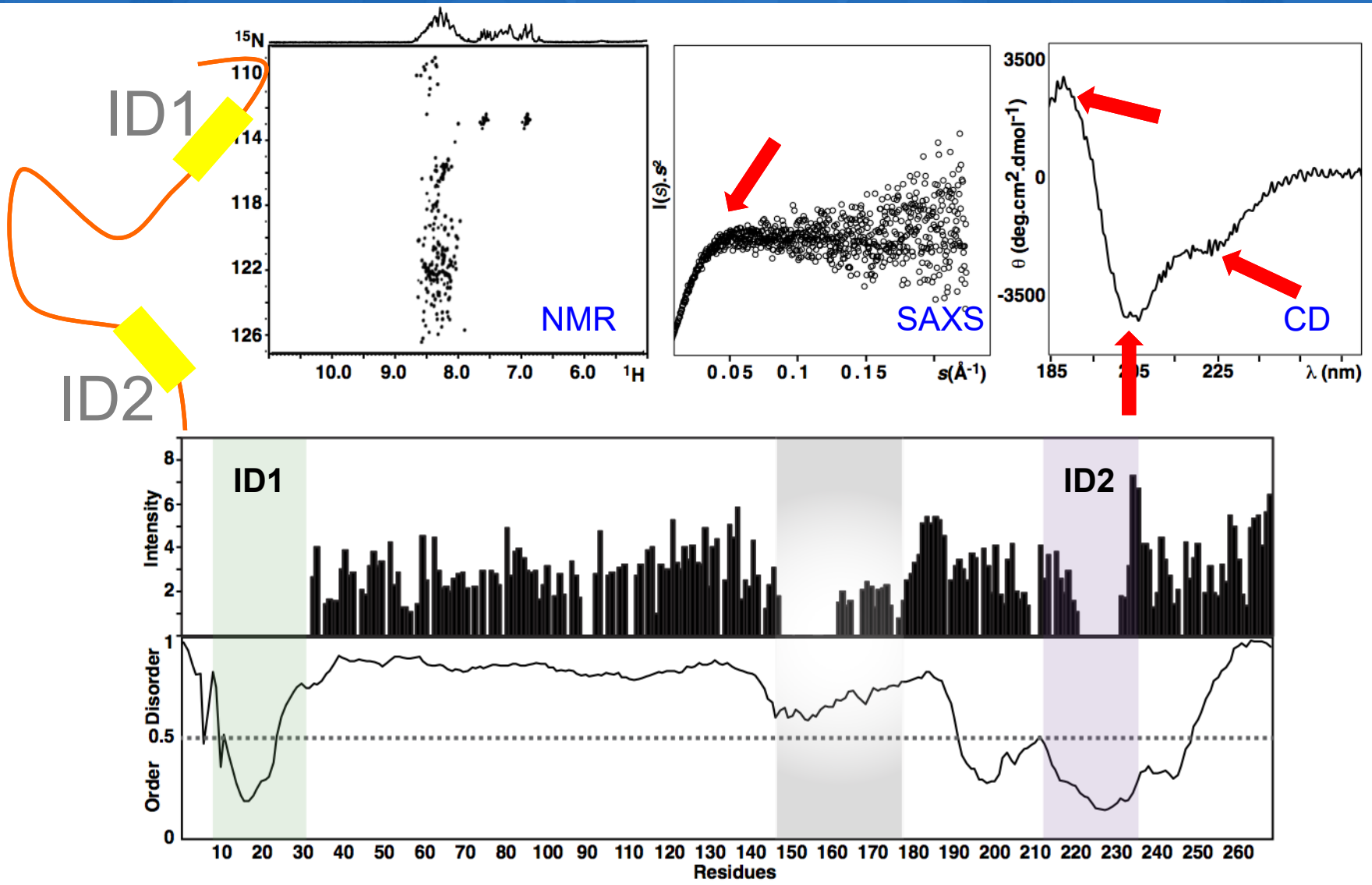


Relevant conformational state in solution?

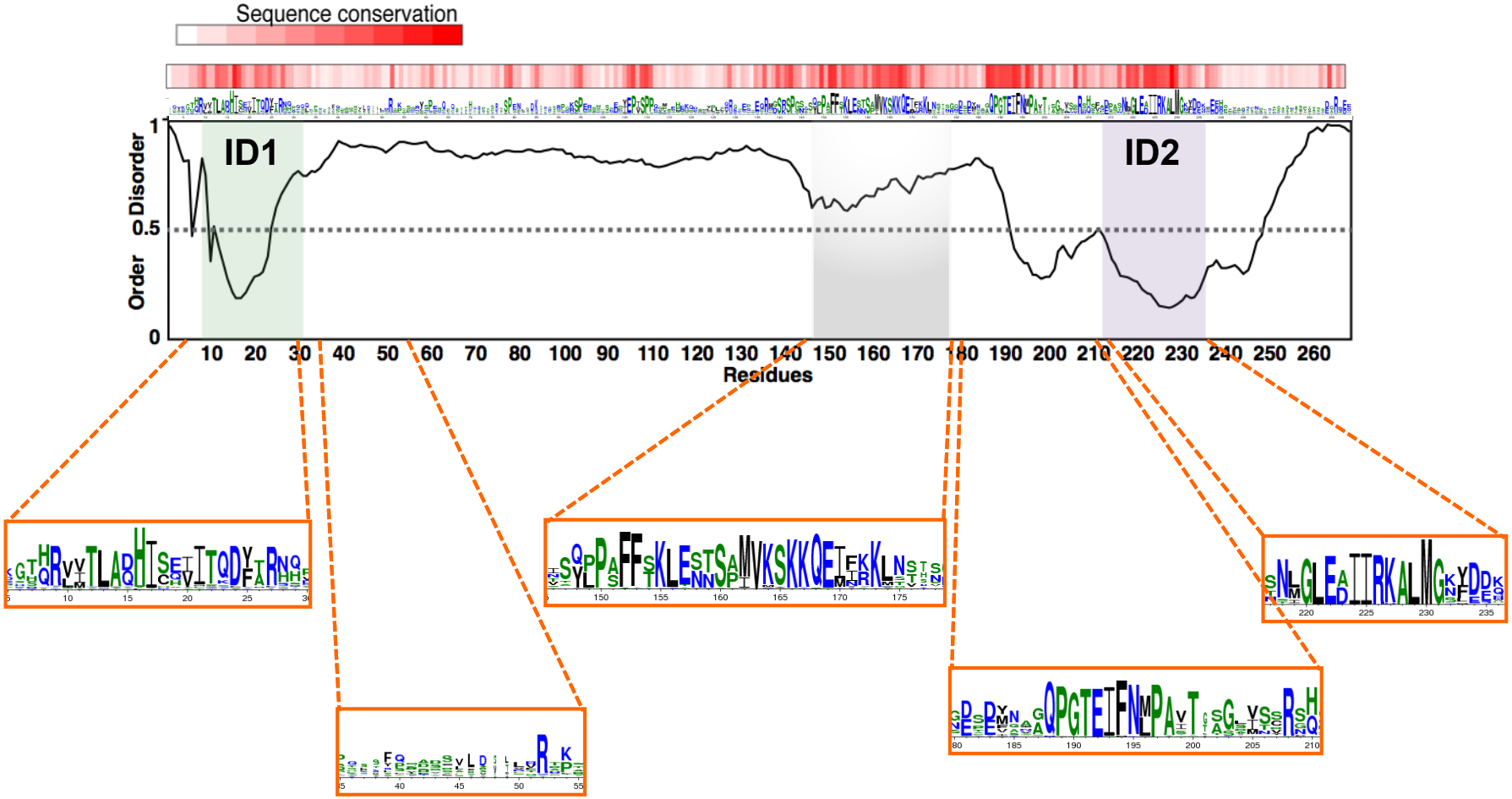
How this scenario is modified by ligands?

How the transition from repressive to active states is achieved?

NCoR is an IDP… but not a Random Coil



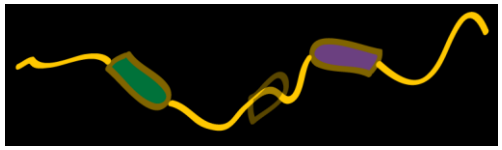
Sequence Conservation in NCoR

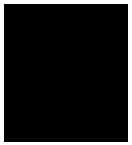


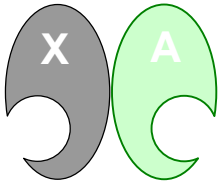
Long-Range Contacts & Co-Evolution

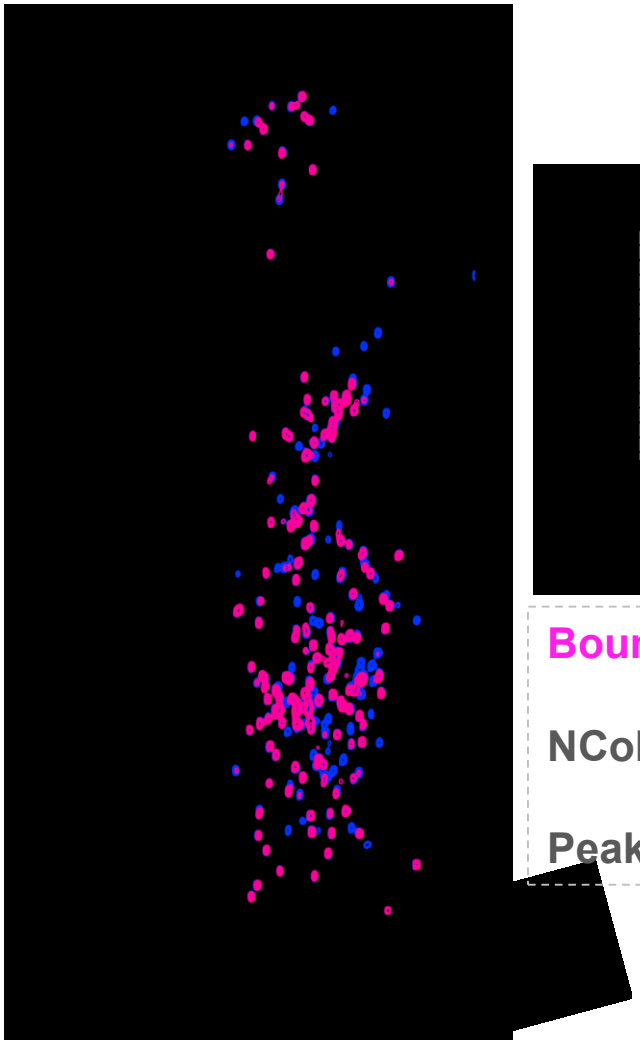
N-CoR is largely a disordered protein with partially structured regions and long-range contacts within the C-terminal region of NCoR

NCoR in Complex with RAR/RXR



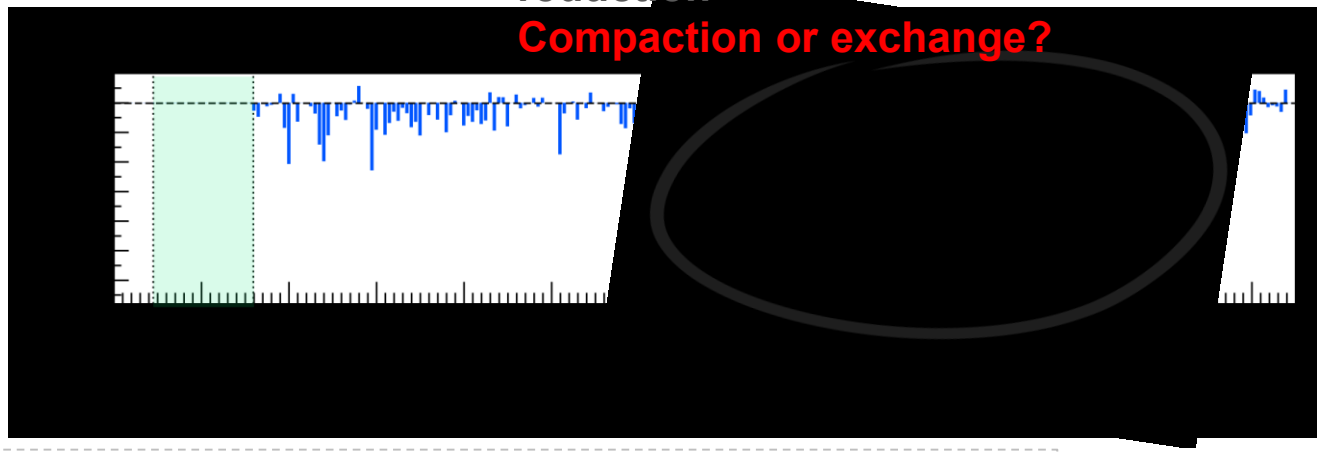






Region connecting IR and ID2 displays a important intensity reduction

Compaction or exchange?

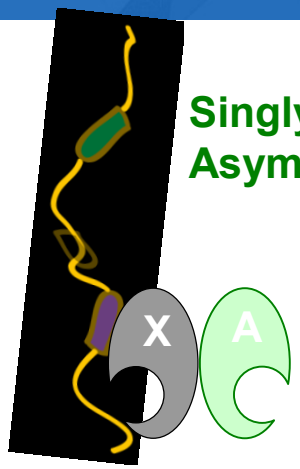


Bound and **Free** spectra are superimposable.

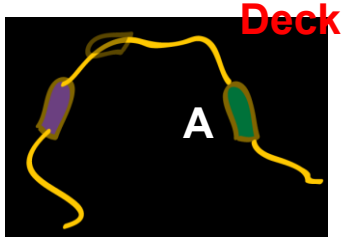
NCoR is essentially disordered in bound state.

Peak Intensities decrease in a heterogeneous way

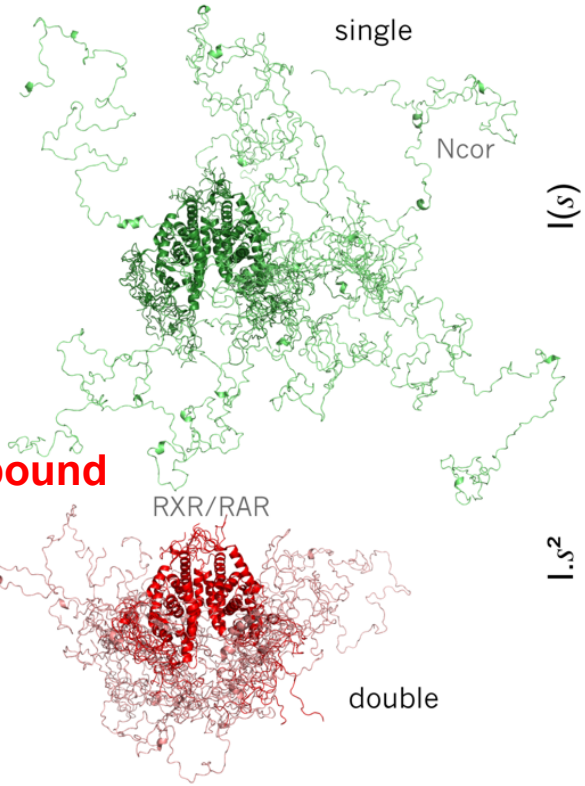
Modeling the interaction of RXR/RAR with NCoR

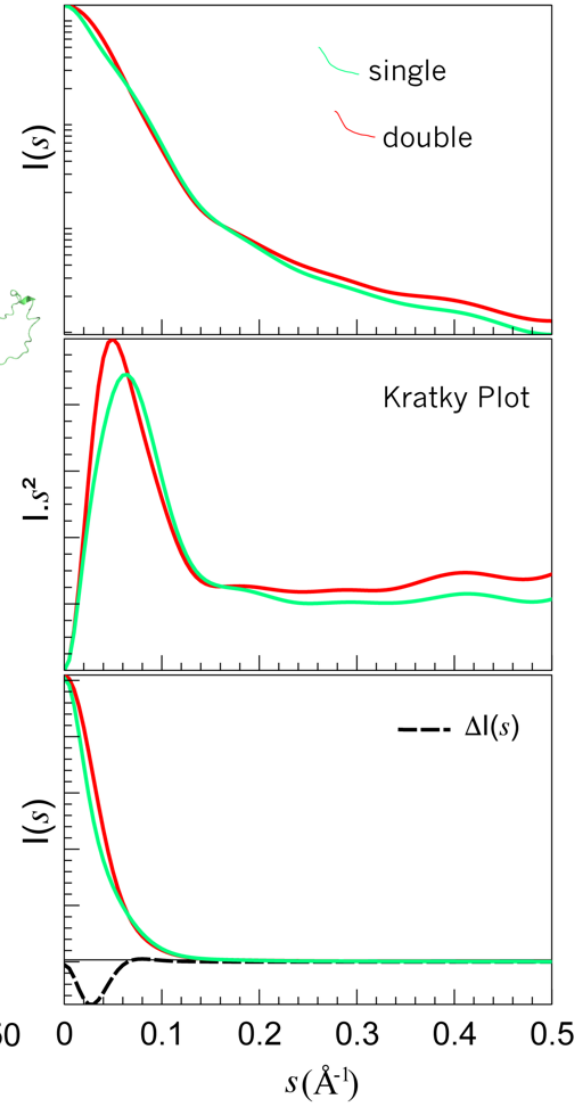


Singly bound
Asymmetric



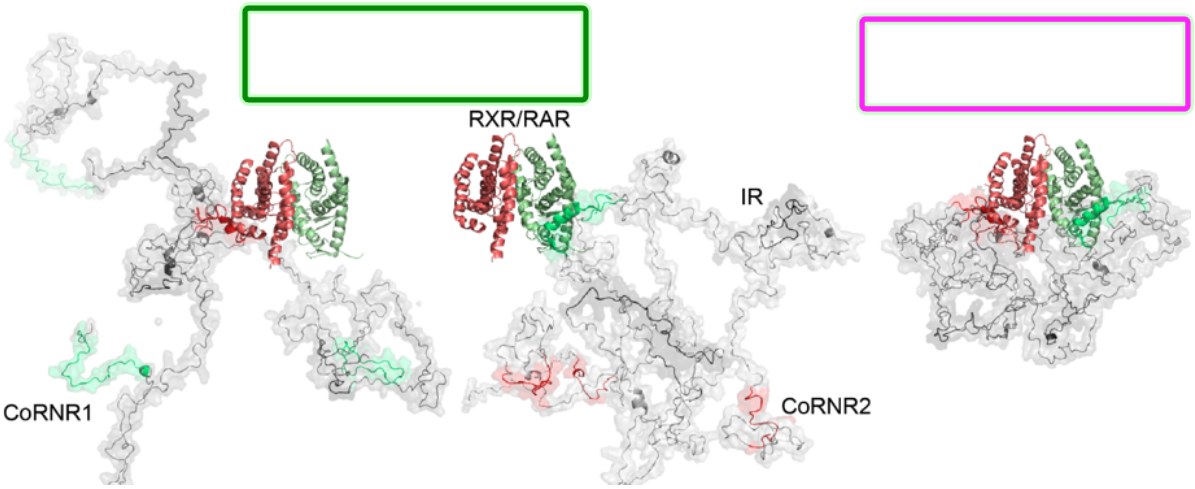
Doubly bound
Deck

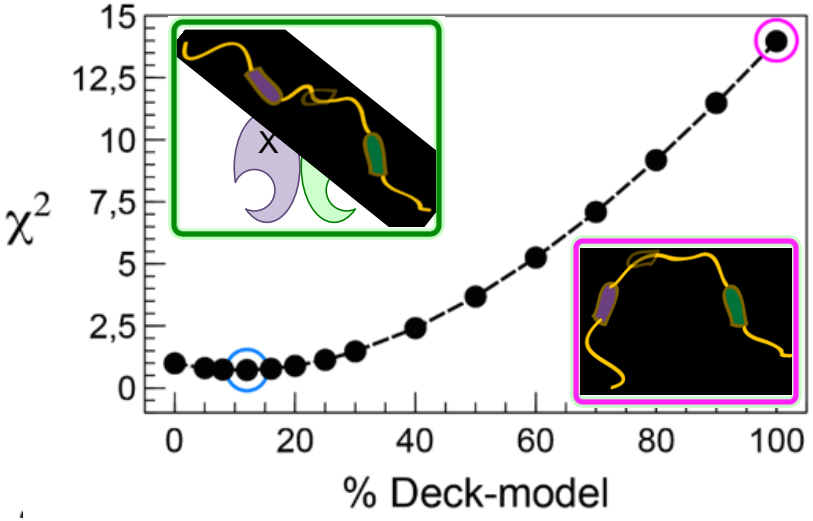


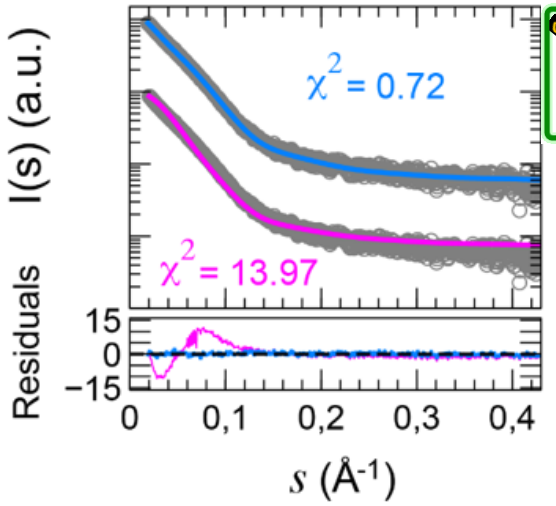


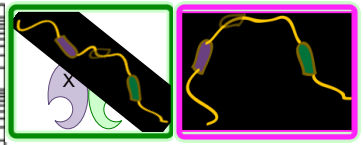
SAXS discriminates between singly- and doubly-bound scenarios

Interaction of RXR/RAR with NCoR









An equilibrium between singly- and doubly-bound scenarios is observed

Conclusions

- In highly flexible proteins, NMR provides the conformational sampling at residue level. SAXS provides the overall size and shape.

- Synergistic application of NMR and SAXS (with the help of computational tools) provides more accurate structural/dynamic models for flexible proteins…

- SAXS provides information about large-amplitude motions in biomolecules and reaches novel and biologically relevant information.

- SANS combined with amino acid specific labelling provides insights into the structure of low-complexity regions in proteins.

- Progress in the structural interpretation of SAXS data (in terms of conformational dynamics) will come from the development of theoretical methods to generate and perturb 3D structures…

Structure and Function of Highly Flexible Proteins

Pau Bernadó (DR2)
Nathalie Sibille (CR1)
Frédéric Allemand (IR1)
Aurelie Fournet (AI)

Aurelien Thureau (Soleil)

Xamuel Lund (PhD)
Amin Sagar (Pdoc)
Carlos Elena-Real (Pdoc)
Assia Mouhand (Pdoc)

Alejandro Estaña (PhD)
Annika Urbanek (Pdoc)
Lucile Senicourt (Pdoc)
Matija Popovic (Pdoc)
Anna Morató (IR)