Using extensive or sparse experimental constraints for (integrative) structural modeling

Ewen Lescop

Renafobis workshop

June 5-12 2015

Me

1999-2000 DEA cristallographie et RMN biologiques (Paris XI, Grenoble, Strasbourg)
2000-2003 PhD ICSN Gif-sur-Yvette (Dir Eric Guittet)
Structure, dynamics and interaction of an honeybee odorant binding protein by NMR

2004-2006 Postdoc in Peking University (Pr Changwen Jin)
Structure of a DNA/transcription factor (RX/NMR)
Structure and dynamics of a phosphotyrosine phosphatase (NMR)
Structure and dynamics of the HPPK enzyme (NMR)

2006-2007 Postdoc at IBS (Grenoble) (Bernhard Brutscher)
Methodological developments in protein NMR: fast NMR data collection (BEST, ..) and analysis (COBRA), new pulsesequences (HADAMAC, …)

Since 2007 CNRS Research associate at ICSN (Gif-sur-Yvette)
Methodological developments in NMR
Dynamics and interaction involving redox multidomain protein by integrative structural biology

cnrs IR THC RMN
TRÈS GRANDES INFRASTRUCTURES DE RECHERCHE
Résonance Magnétique Nucléaire, Très Hauts Champs
FR3050 CNRS

English Français
Soumettre une proposition>
Recherche... >>
Suivre un projet>

Présentation Projets Actualité Évènements Résultats Contact Activité et Satisfaction

Réunion d'Utilisateurs

15 Oct **7ème Réunion des Utilisateurs organisée par ICSN Gif/Yvette**
Gif-sur-Yvette

Evènements

22 Mai **Formation Atelier Pratique en RMN**
Grenoble

Remerciements : « Financial support from the TGIR-RMN-THC Fr3050 CNRS for conducting the research is gratefully acknowledged.»

Le TGE décentralisé RMN Très Hauts Champs, est un réseau constitué d'équipes de recherche reconnues au niveau international en RMN, exploitant des spectromètres RMN Hauts Champs.

Le Réseau est une structure ouverte à une communauté nationale et internationale d'utilisateurs. Il a pour but de répondre au mieux aux attentes scientifiques des communautés d'utilisateurs et aux experts de la spectroscopie RMN.
Les laboratoires d'accueil proposent l'accès à leurs installations à hauts champs magnétiques, accompagné d'une expertise scientifique des possibilités offertes par les méthodes les plus récentes pour les développements de nouvelles directions de recherche.

Actualité

27 avril 2015
Commande du premier 1.2 GHz français
La première commande d'un spectromètre 1.2 GHz vient d'être passée auprès de la compagnie Buker. Il s'agit de la 7ème commande d'un appareil de...
Lire la suite...

Overview

1. The concept of free energy landscape

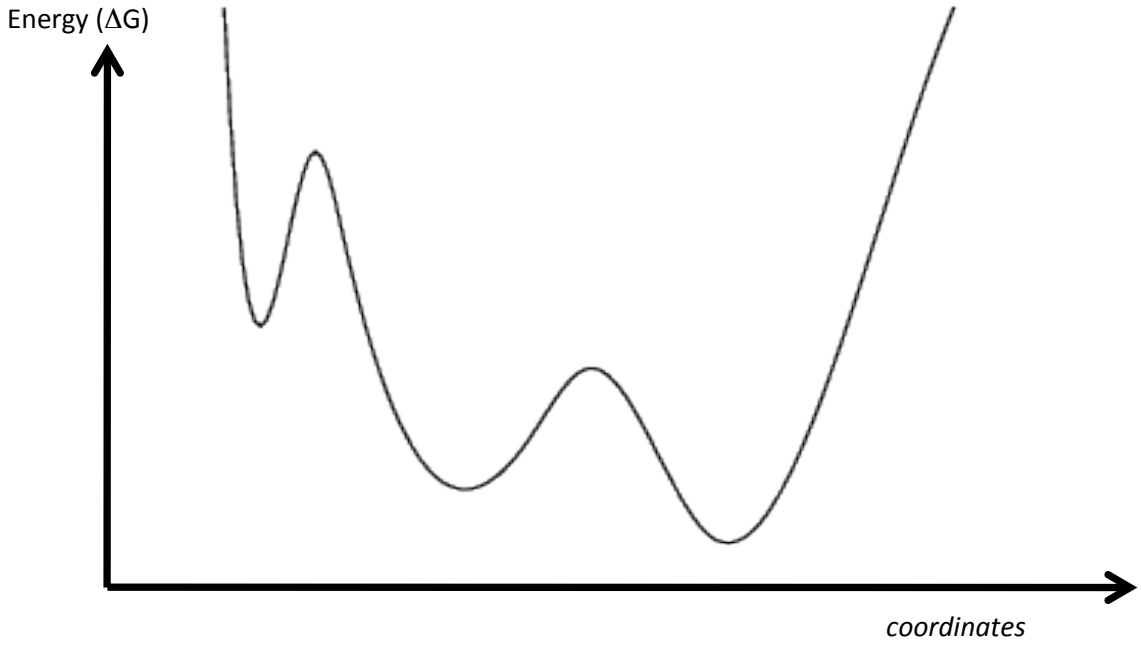
2. Basic reminders on molecular modeling

3. Applications to structure determination by Xray and NMR:
   « Almost » overdetermined systems

4. Applications to sparse experimental data:
   Largely under-determined systems
   1. flexible systems (IDP)
   2. macromolecular assemblies

Free Energy landscape

☞ Every single protein conformation is associated to a given (free) energy (Gibbs energy)

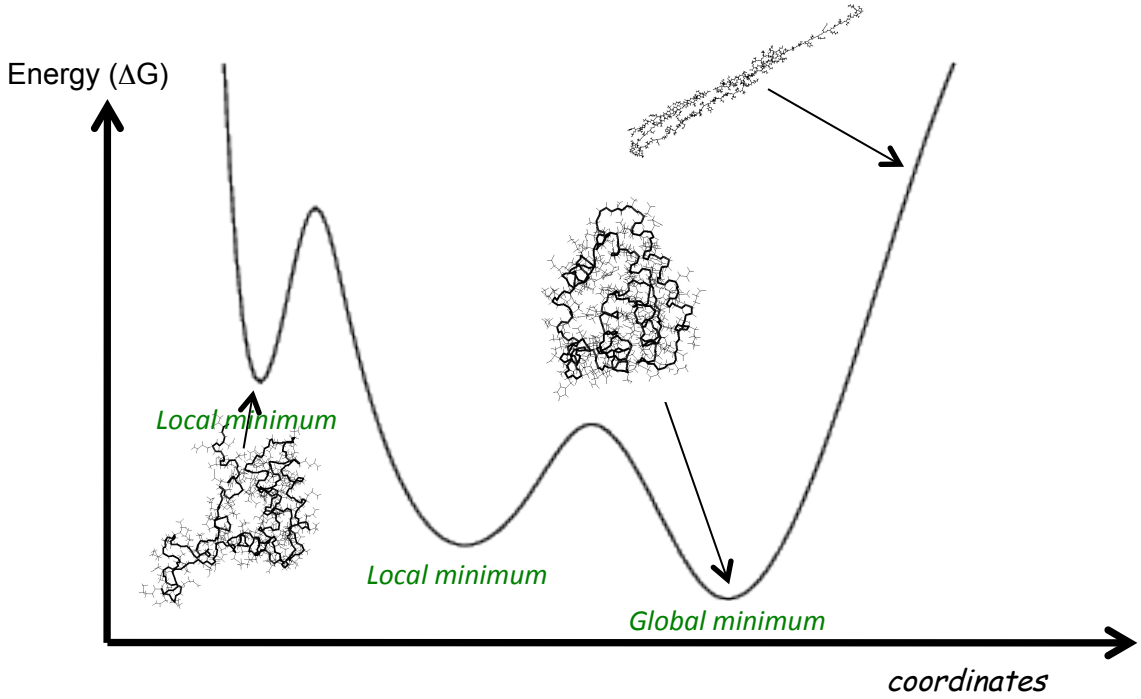


☞ X-axis : conformational axis

- 1D when only one degree of freedom (e.g. ^{1}H-^{1}H in H_2)
- Multidimensional (nD) space associated to the nD degrees of freedom
- Reduction to the most relevant direction in the nD dimension (*collective variable*)

Energy landscape

☞ Every single protein conformation is associated to a given (free) energy (Gibbs energy)



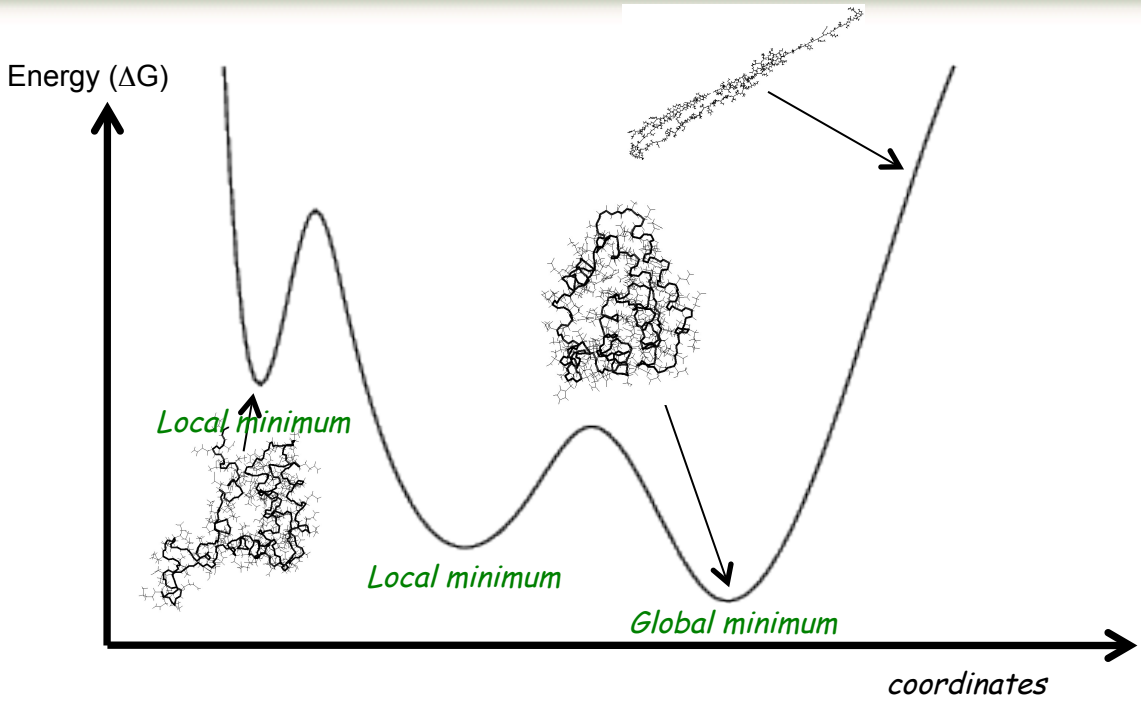
☞ Proteins follow Boltzmann's law:

"at thermal equilibrium, the population distribution

depends on the energy of each state (conformation)":

$$p(state) \propto e^{-\frac{\Delta G}{kT}}$$

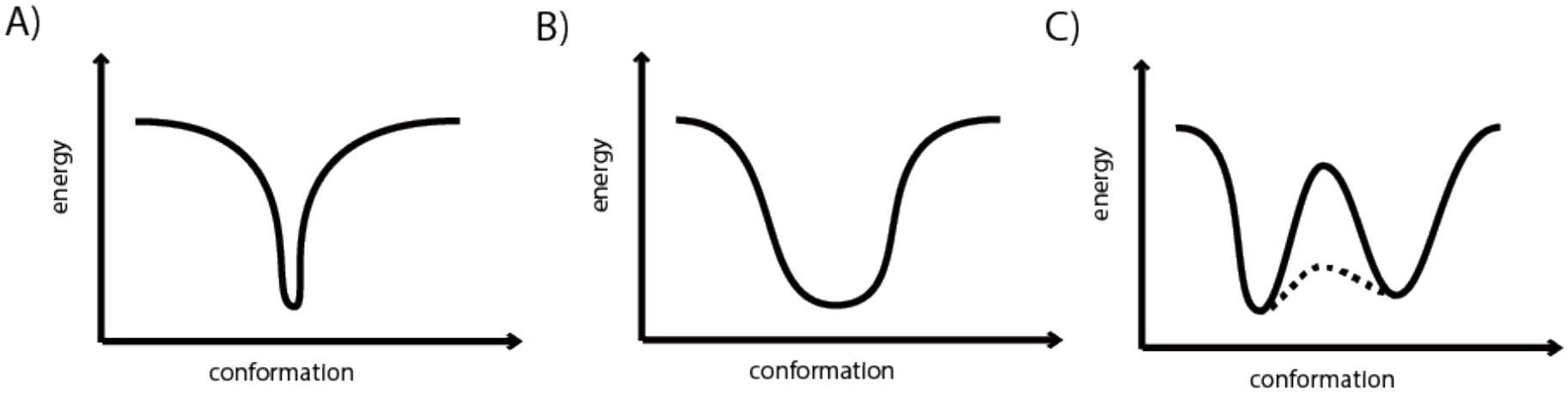
Energy landscape



- The most stable (populated) conformation lies at the global minimum
- Local minima are “excited” states that are lowly populated
- The barriers between minima report on the kinetics of conformational transitions

The free energy landscape is a convenient way to represent structure **and** dynamics of proteins

Examples of energy landscapes



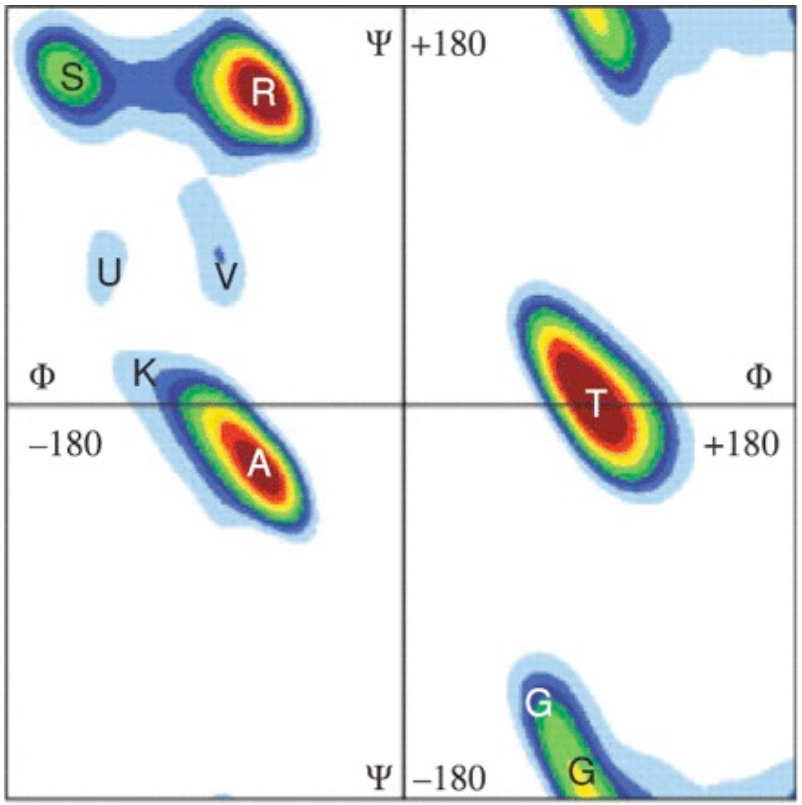
Unique rigid native conformation

Unique flexible native conformation

Two alternative flexible native conformations

Slow or rapid interconversion

Ramachandran plot: a 2D energy landscape



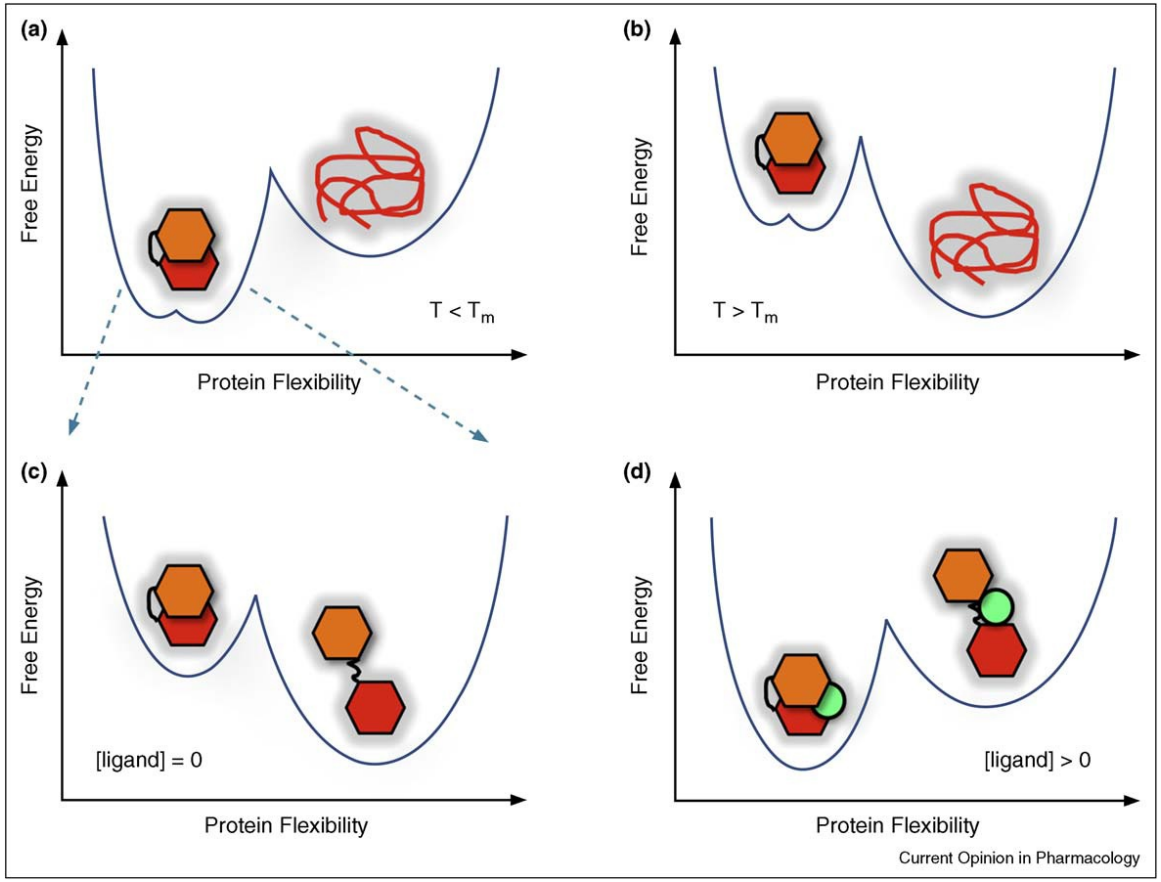
A: α-helices
S: β-sheets
R: polyproline type II

K: ends of helices or in 3_{10} helices.
T: turn
G: glycine.

Generated from the analysis of (Ψ, Φ) angles for each residue from crystal structures.

Can be interpreted as the 2D free energy landscape revealing the population distribution of (Ψ, Φ) in the so-called "random-coil" (or denatured).

Remodeling of Energy landscape



Temperature-induced denaturation:

The native folded state is favoured at low temperature

The denatured unfolded state is favoured at high temperature

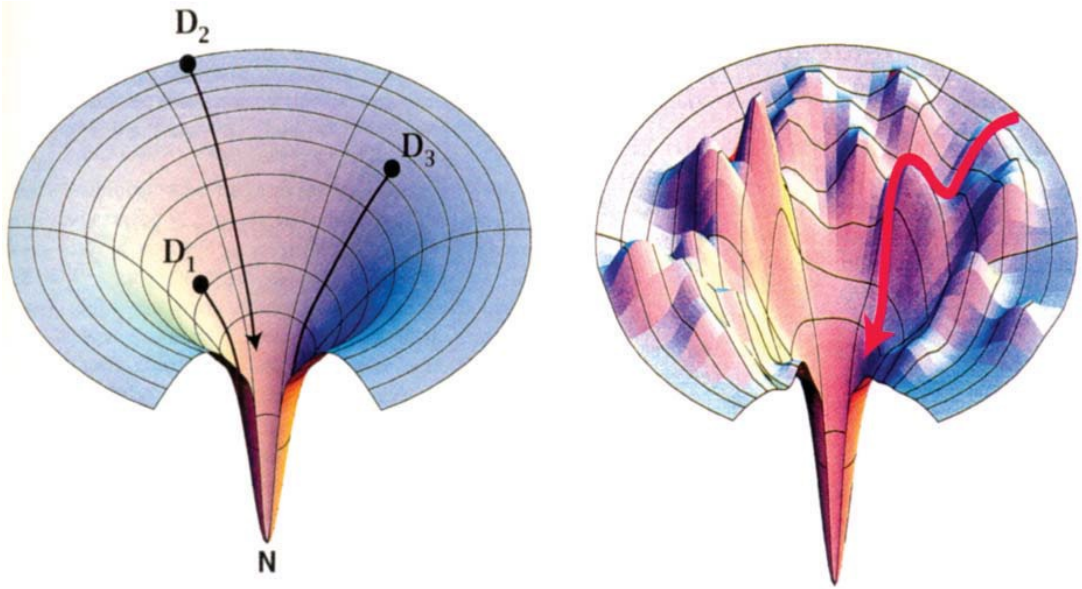
Ligand-binding effects:

Ligand-binding stabilizes conformational states that were lowly populated in the apo-state

Related to “lock-and-key”, “induced-fit” or conformational selection” theories.
Allostery

Dennis R Livesay Current Opinion in Pharmacology 2010, 10:706–708

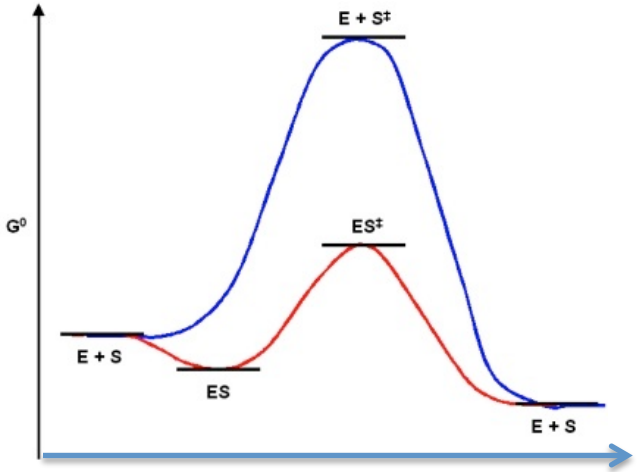
Energy landscape and folding



Modern view:
Protein folding results from the exploration of a funnel-type landscape

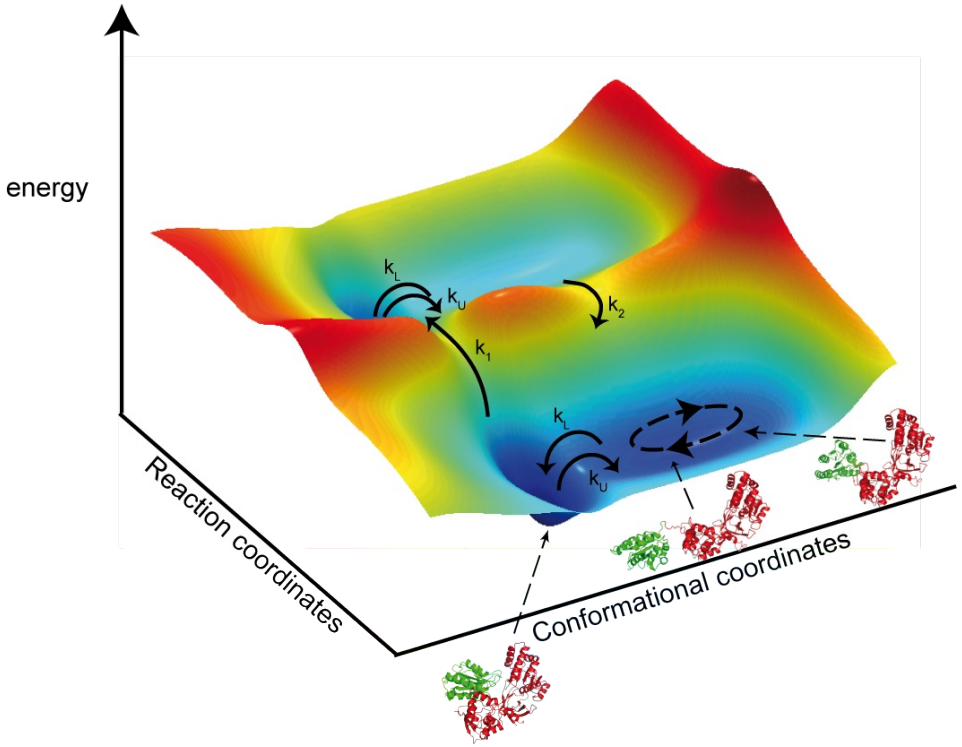
K.A.Dill & H.S.Chan (1997) Nature Struct.Biol. **4**, 10-19;

Energy landscape and catalysis



Reaction coordinates

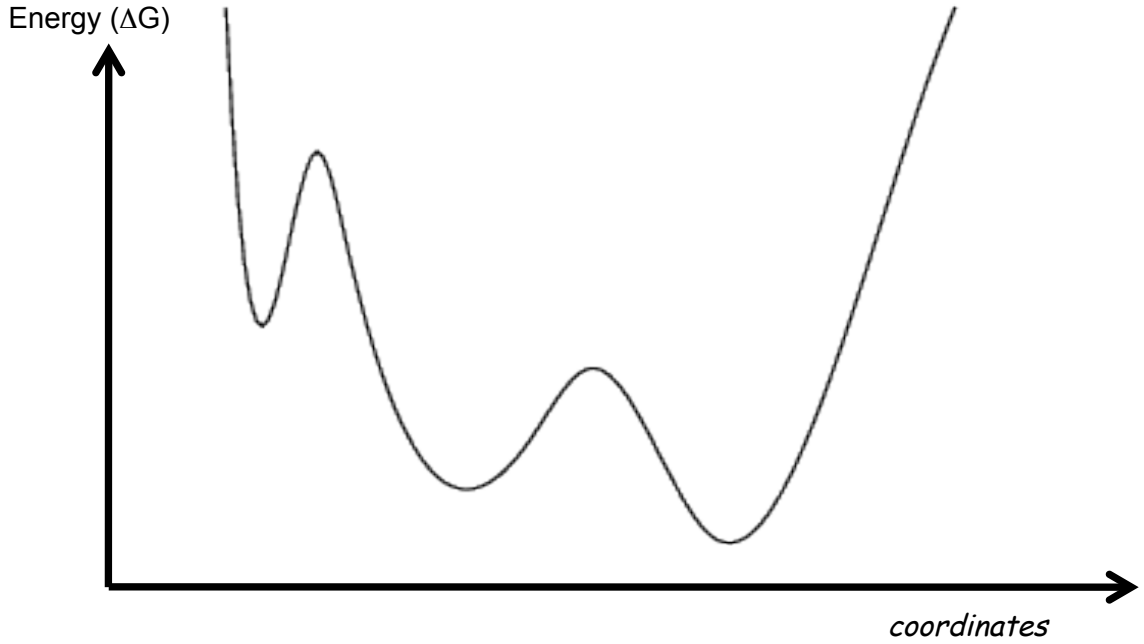
Conventional view of catalysis:
The reaction progresses through energy barriers from the substrate S to product P through the transition-state.



Multidimensional energy landscape

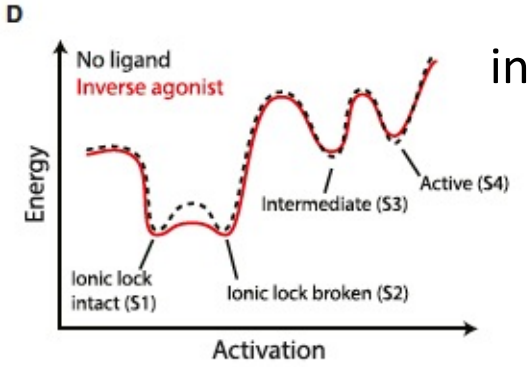
Frances O, Fatemi F, Pompon D, Guittet E, Sizun C, Pérez J, Lescop E, Truan G. Biophys J. 2015;108(6):1527-36

And so?

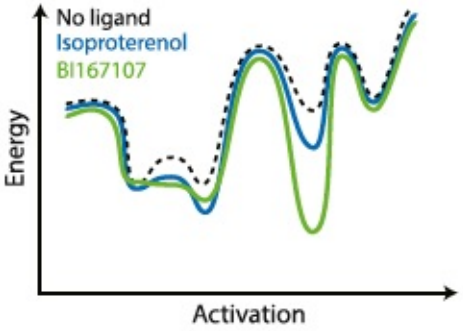


What is the active conformations?
What is the meaning of the active conformation?
How the landscape is modified upon interaction with partners, substrates, etc…?
Correlating the energy landscape with the biological function

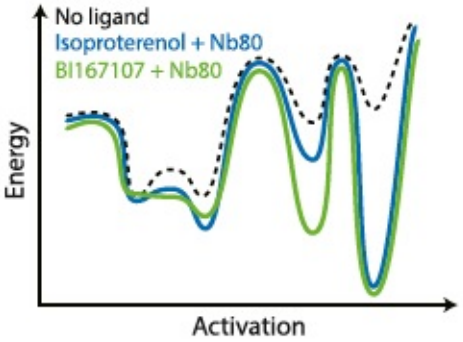
Activation of β2-adrenergic receptor (GPCR)

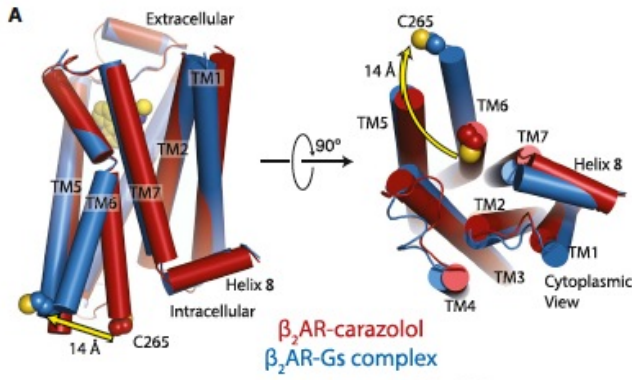


inverse agonists



agonists

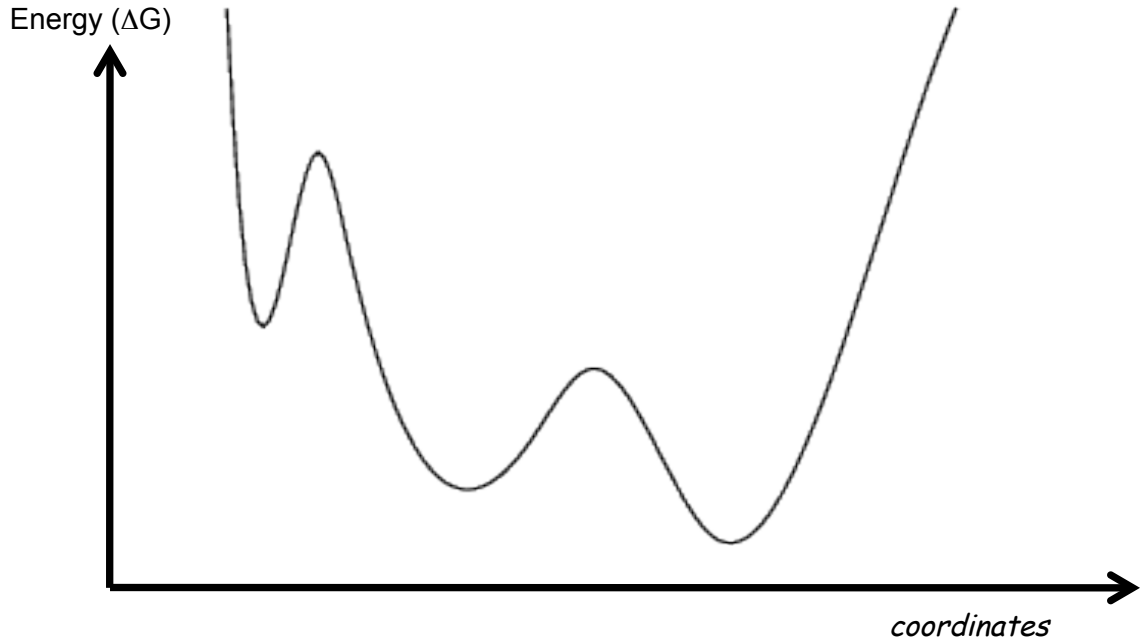




The relative depth (populations) of the various conformers are modified upon agonist/antagonist binding.
The energy barriers (kinetic) are also adapted.
The active conformer is populated upon agonist binding.

Manglik A, Kim TH, Masureel M, Altenbach C, Yang Z, Hilger D, Lerch MT, Kobilka TS, Thian FS, Hubbell WL, Prosser RS, Kobilka BK. Cell. 2015 May 21;161(5):1101-11.

The dream of theorists



The energy landscape is encoded in the protein sequence and is adjusted to the local environment and physico-chemical parameters

So: for a given protein, energy calculation -> energy landscape -> interpretation
-> “Protein structure prediction without experimental data”

But does not work for protein

Why it does not work for proteins

Quantum Mechanics allowed the (almost) exact energy calculation

But not tractable for (very) large molecules

Approximations based on empirical data are not accurate enough

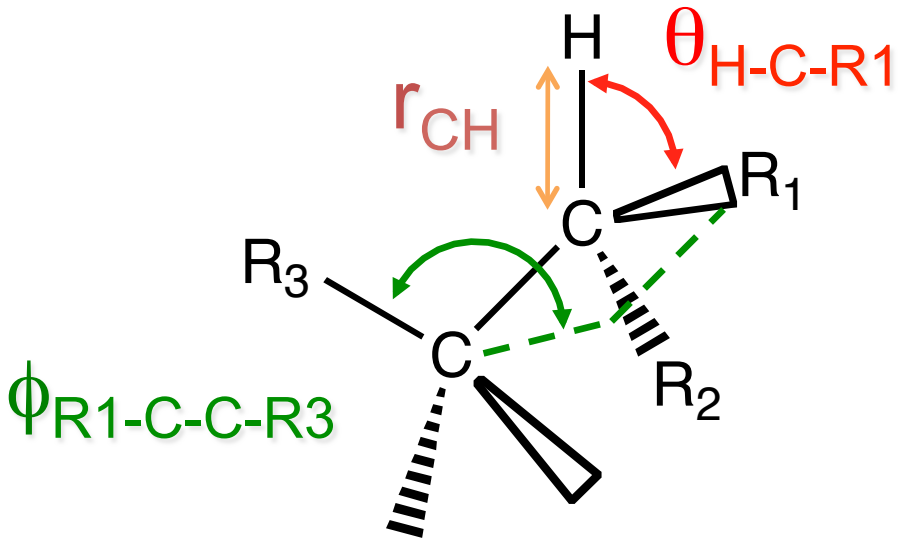
The difference in energy between folded (native) and unfolded (denatured) structure is low (related to function).

The environment (solvent, membrane, partners, etc…) plays a crucial role.

Fortunately for us, theorists are not good enough at the moment

Protein structure respects chemistry

☞ As all molecules, proteins should respect chemistry (bond length, angles, …)



$$E_{bonded} = E_{bond} + E_{angle} + E_{tor} + E_{improper}$$

Bonded terms

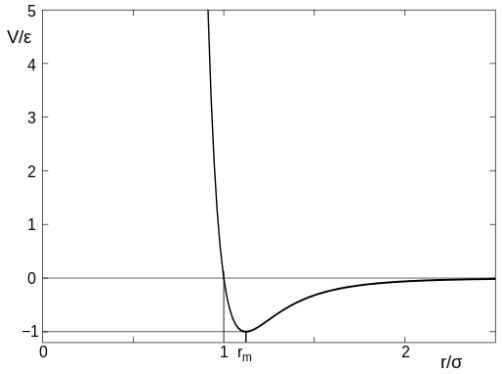
$$E_{\text{bond}} = \sum_{i,j \in S_{\text{B}}} S_{ij}\left(r_{ij} - \bar{r}_{ij}\right)^2$$

$$E_{\text{bang}} = \sum_{i,j,k \in S_{\text{BA}}} K_{ijk}(\cos\theta_{ijk} - \cos\bar{\theta}_{ijk})^2$$

$$E_{\text{tor}} = \sum_{ijk\ell \in S_{\text{DA}}} \sum_{n} \left(\frac{V n_{ijk\ell}}{2} \left[1 \pm \cos(n\tau_{ijk\ell})\right] \right)$$

Force field

- Protein 3D folding results from short and long-range physico-chemical interactions: electrostatic contacts / Van der Waals / hydrogen bonds, etc….
- Electrostatic interactions $$E_{\mathrm{coul}} = \sum_{i,j \in S_{\mathrm{NB}}} \left(\frac{q_i q_j}{\epsilon(r_{ij})\, r_{ij}} \right)$$
- Lennard-Jones potential $$E_{\mathrm{LJ}} = \sum_{i,j \in S_{\mathrm{NB}}} \left(\frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} \right)$$



Non-Bonded terms

$$E_{\mathrm{nonbonded}} = E_{\mathrm{electrostatic}} + E_{\mathrm{van\ der\ Waals}}$$

Force field

- Protein folding results from physico-chemical interactions

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$$

$$E_{non\text{-}bonded} = E_{electrostatic} + E_{van\ der\ Waals}$$

- Total energy used in most force-fields (Amber, …)

$$E = E_{bonded} + E_{non-bonded}$$

- This empirical energy calculation is rapid to calculate at the expense of an approximative estimation

Molecular dynamics

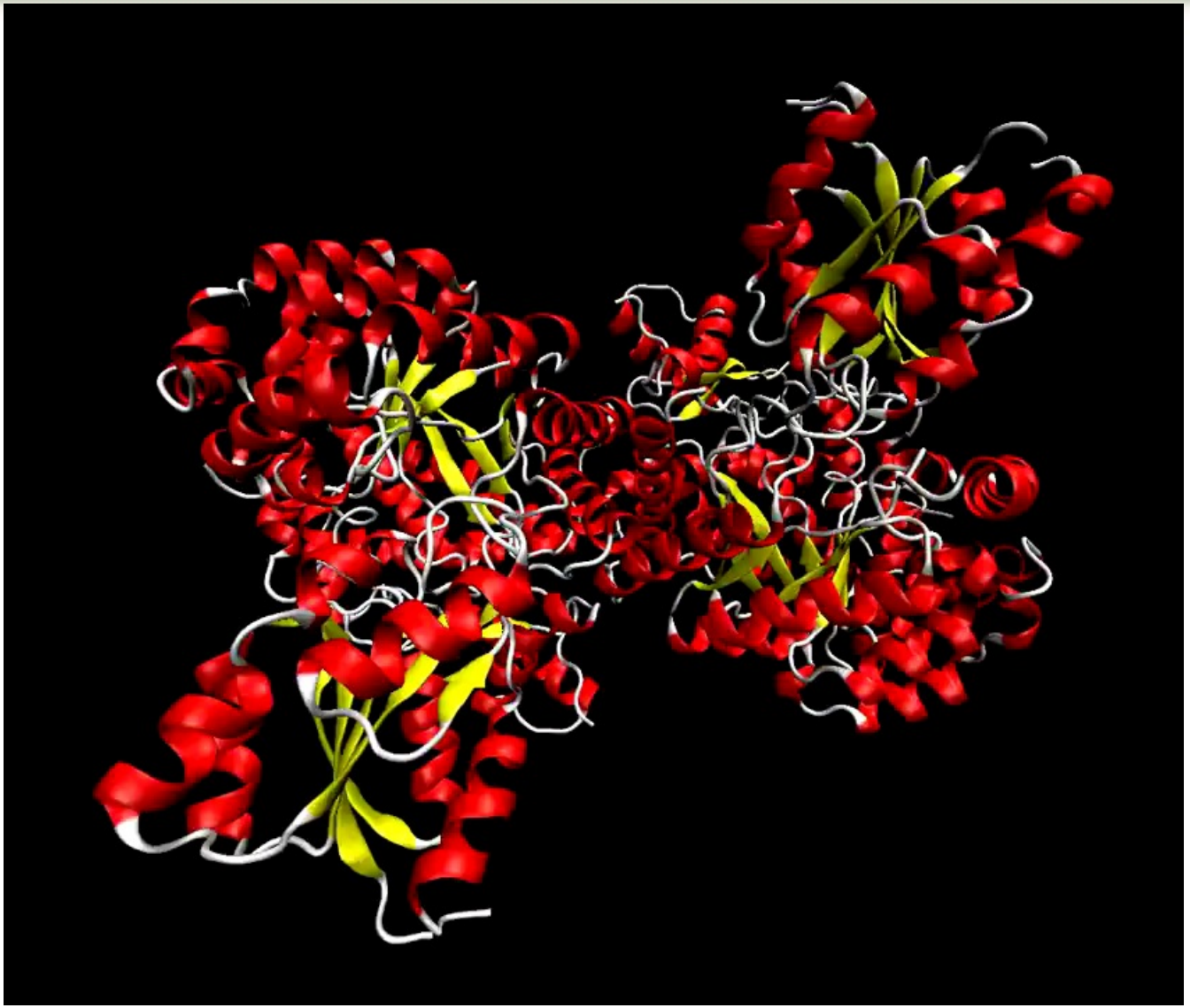
A time t_0, structure -> position of all atoms in space r=(x,y,z) -> energy

Equation of motion $$\frac{d^2\vec{r}_i}{dt^2} = -\frac{c}{m_i}\frac{\partial}{\partial\vec{r}_i}E_{hybrid}$$

Predict the new position at time *t* (step by step)

Molecular dynamics trajectory

Molecular Dynamic simulation of Glutamate Mutase (pdb code 1I9C) in explicit solvent (not shown) using NAMD molecular dynamics software and the CHARMM22 molecular mechanics forcefield.



https://www.youtube.com/watch?v=ziveAfGp7lY

Molecular dynamics trajectory

www.youtube.com/watch?v=bRDpoo9R5N0

Molecular dynamics of CmAAT1 Enzyme, CmAAT1-Hex-aCoA complex

Hybrid energy

$$E_{hybrid} = E_{phys} + w_{data} E_{data}$$

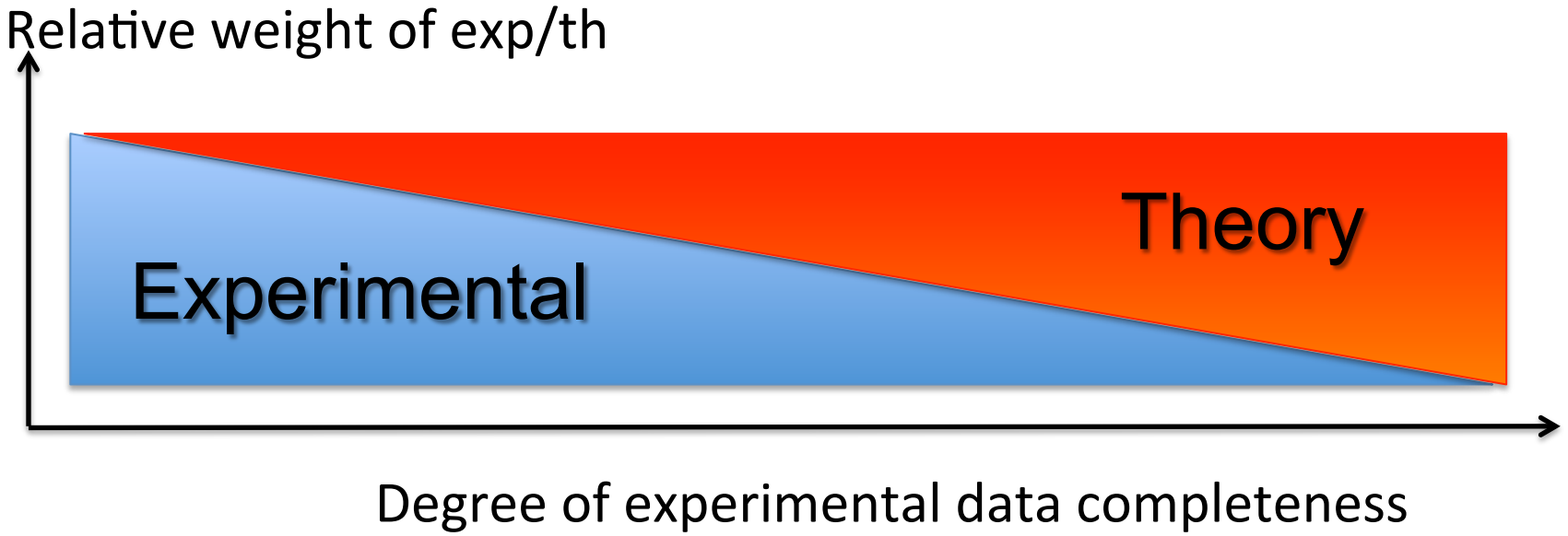
Geometry

Real physical energy term

Experiment

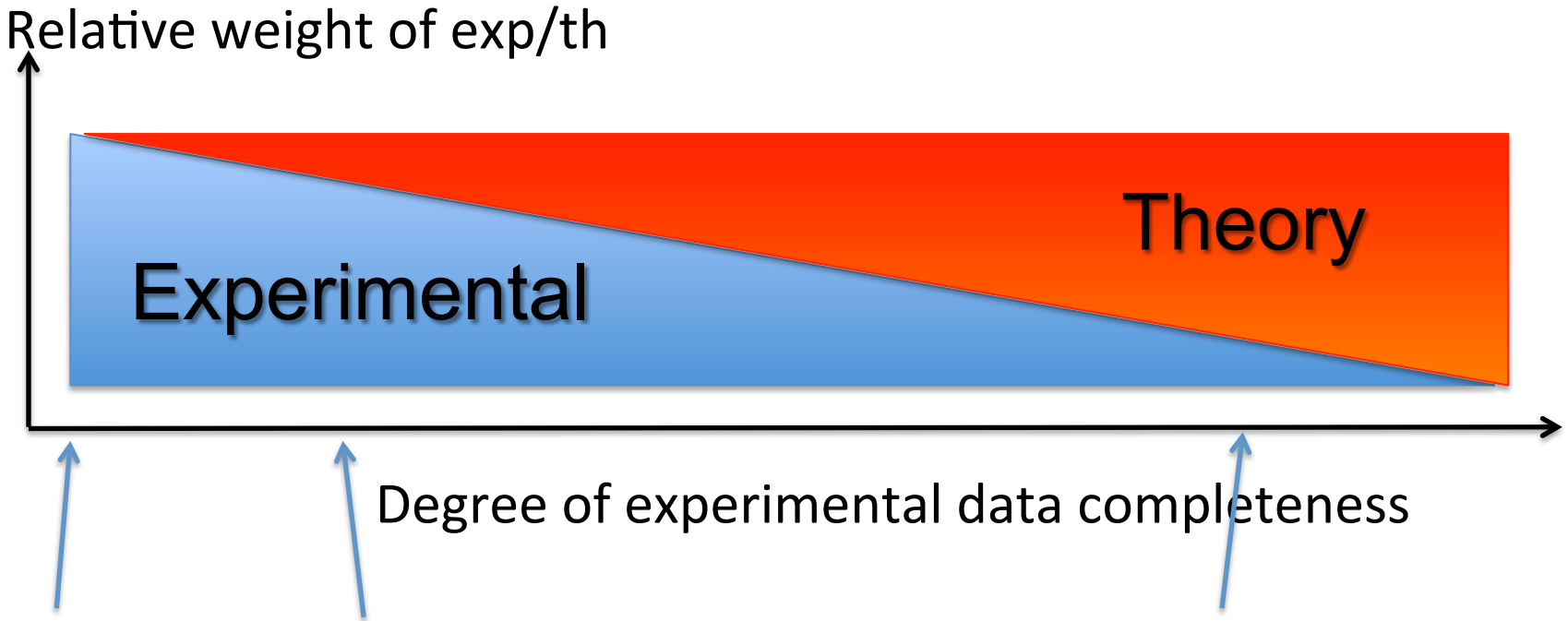
Adhoc energy term to take experimental into account

Hybrid energy



The most complete the experimental data are,
the less theoretical data you need

Hybrid energy



Very high resolution (<1Å) Xray diffraction data

High resolution (>1Å) Xray diffraction data / NMR / high-res Cryo EM

1. Highly flexible systems (IDP)
2. Sparse data (low-res Cryo EM / NMR / MS / cross link / EPR / …)

Overdetermined experimental data

Underdetermined experimental data

Experimental data and pseudo-energy

$$E_{data}=\Sigma k(A_{calc}-A_{exp})^2$$

A_{exp}: an experimental observation such as

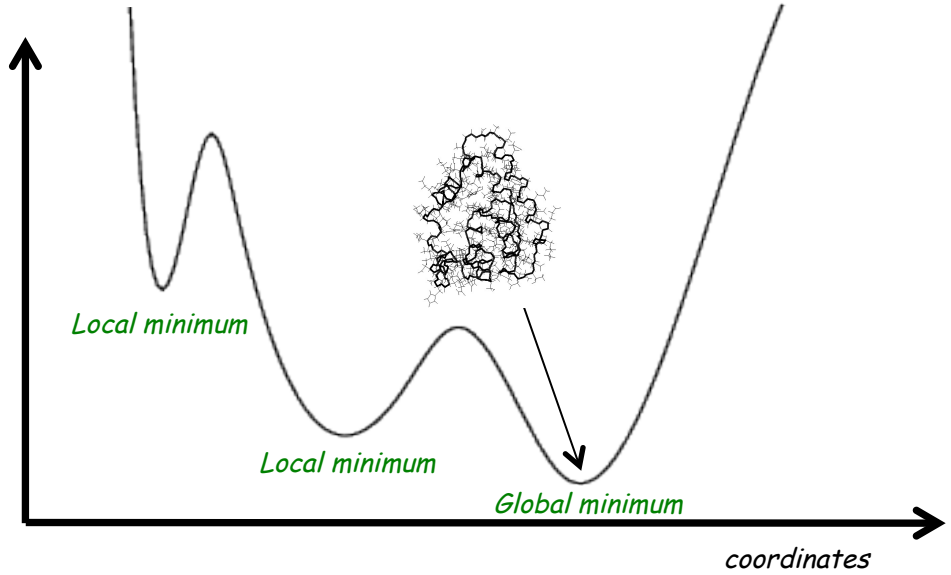
- each diffraction intensity in diffraction pattern
- each ^1H-^1H distance obtained from NMR NOESY data
- each dihedral angle from NMR chemical shifts
- etc….

A_{calc}: the back-calculated value from the structural model (coordinates, …)

k: weighting factor between the various experimental sources

Solve the structure: a minimization problem

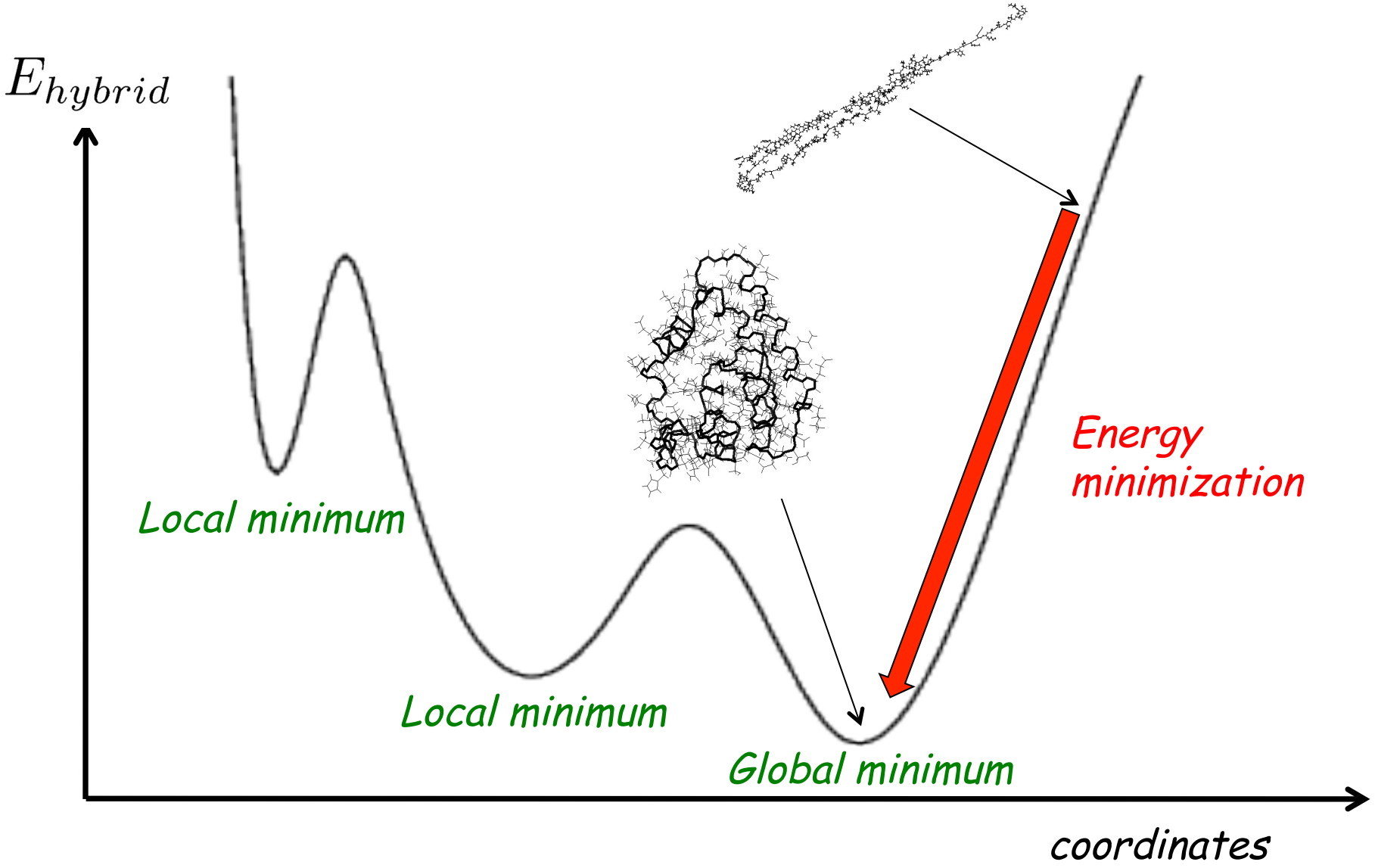
$$E_{hybrid} = E_{phys} + w_{data} E_{data}$$



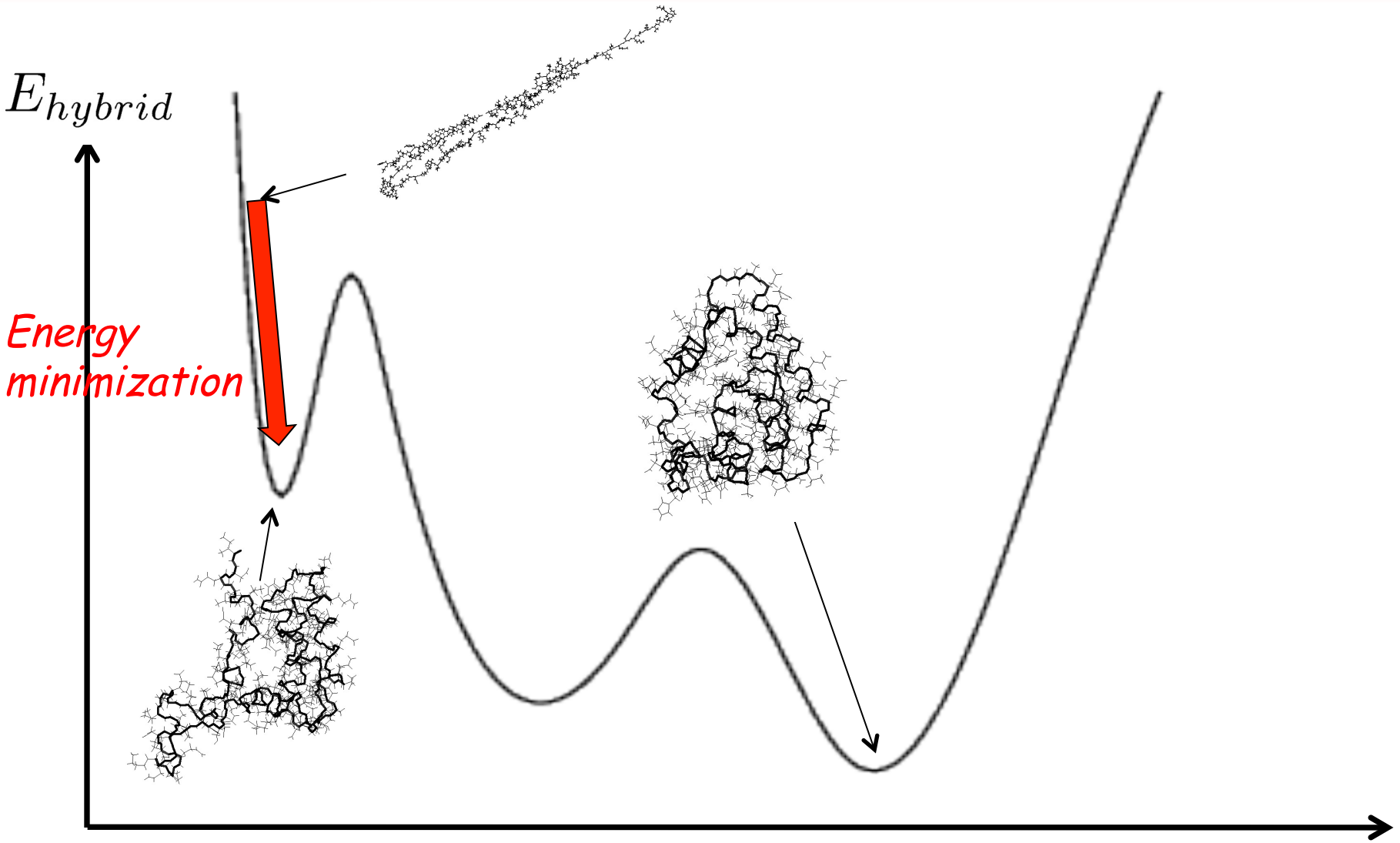
The final structure must both satisfy theoretical chemistry and experimental data

This corresponds to finding the less energy conformers

Minimization algorithm

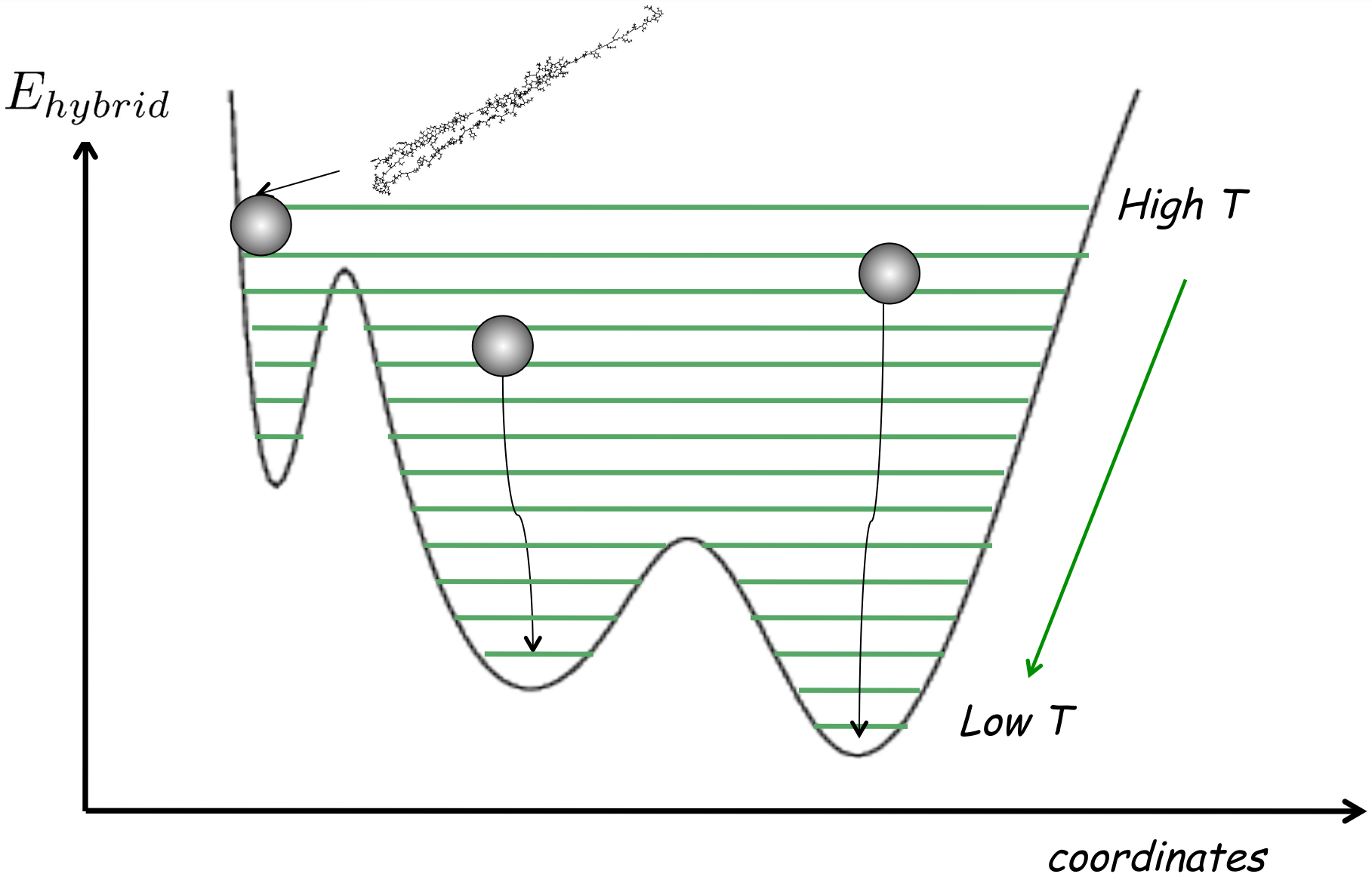


Energy Minimization (EM)

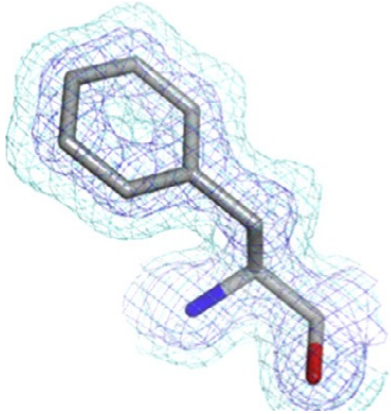


EM converges to the closest local minimum !

Simulated Annealing (SA)

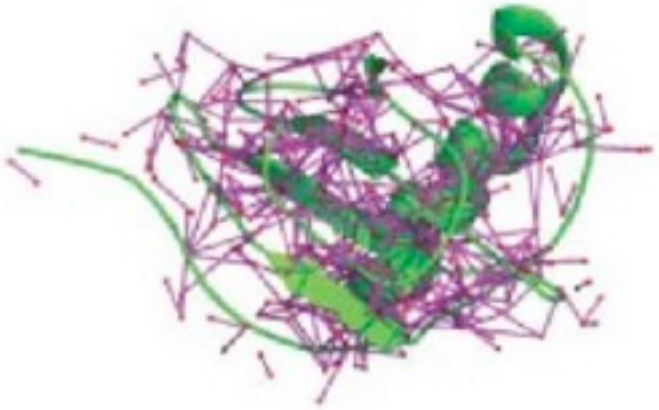


Structures



The final model satisfies both bond length, angles, …

and the electron density (real space) or reciprocal space



The final model satisfies both bond length, angles, …

and the NMR restraints (distances, angles, orientations, …)

NMR derived experimental restraints

NMR provides information about structure:

- ☞ NOE: interproton distances
- ☞ Coupling constants: torsion angles
- ☞ Chemical shifts: local electronic environment

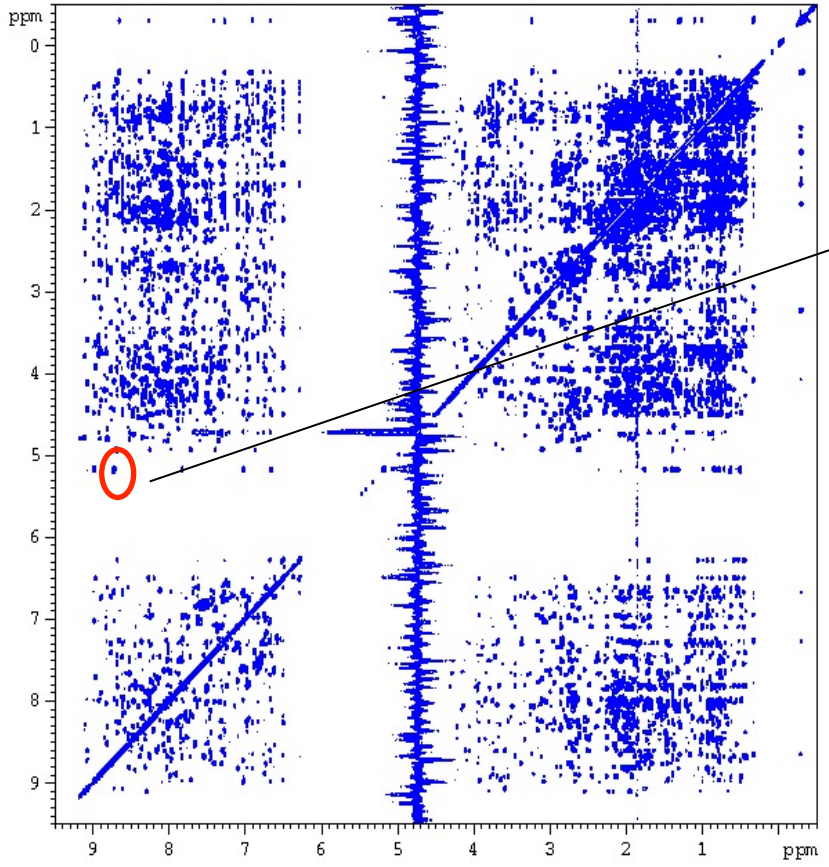
- ☞ Residual Dipolar Couplings: bond orientation
- ☞ Paramagnetic effects

  (pseudo contact shifts, relaxation enchancement):

  Long range distances and orientations

  .....

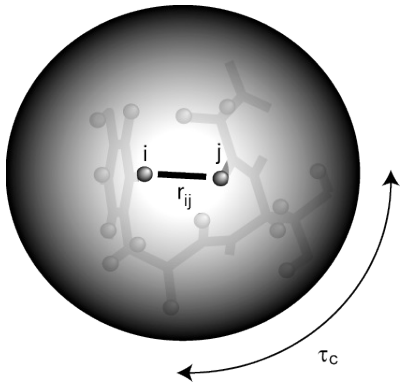
NMR: NOESY to distances



NOESY experiment

NOESY experiment

One off-diagonal crosspeak at (ω_A,ω_B) indicates a short (<5Å) distance between two 1H that resonate at ω_A and ω_B



$$NOE_{ij} \propto r_{ij}^{-6}$$

$$r_{ij} \approx (C_{cal} NOE_{ij})^{-\frac{1}{6}}$$

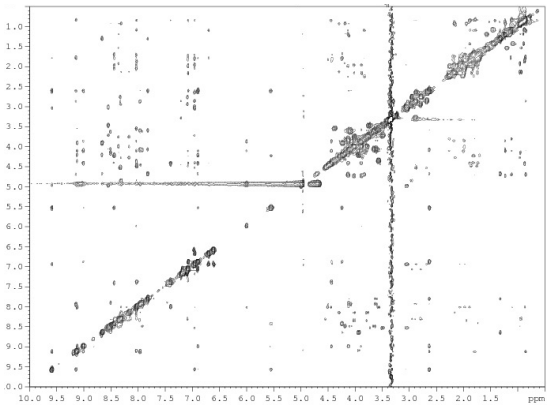
Structure reconstruction

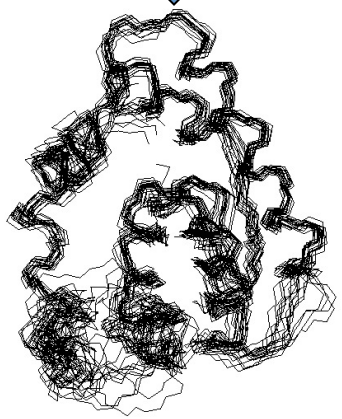
Protein resonance assignment

NOE assignment

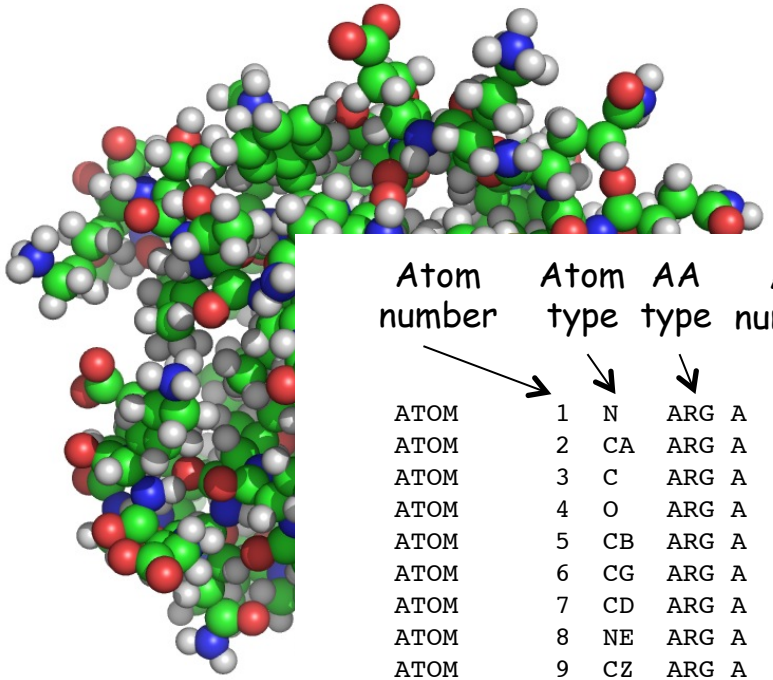
List of distances restraints

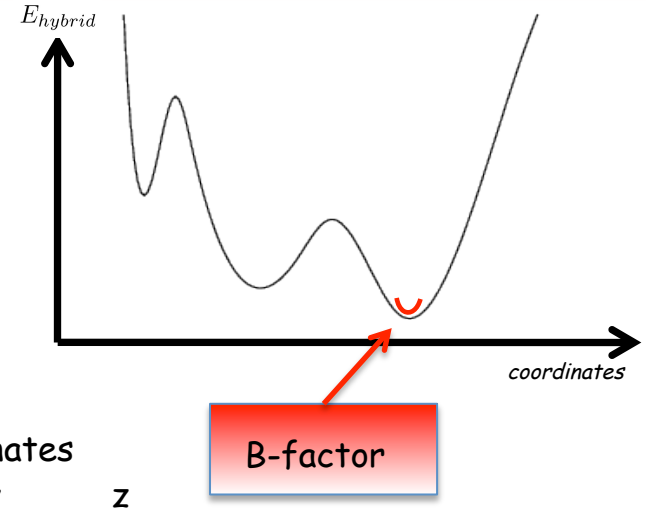
List of NOE peaks (ω_j, ω_j)





What's a structure file?



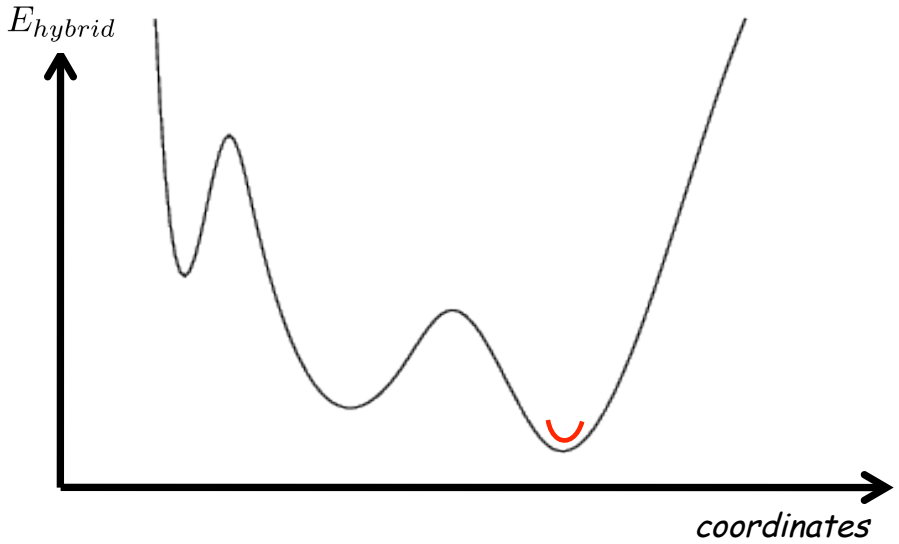


Atom number	Atom type	AA type		AA number	Coordinates x	y	z			
ATOM	1	N	ARG	A 3	8.772	45.207	48.144	1.00	50.93	N
ATOM	2	CA	ARG	A 3	8.715	43.787	48.537	1.00	50.18	C
ATOM	3	C	ARG	A 3	8.439	43.628	50.030	1.00	50.33	C
ATOM	4	O	ARG	A 3	7.303	43.346	50.404	1.00	52.36	O
ATOM	5	CB	ARG	A 3	10.009	43.062	48.096	1.00	46.68	C
ATOM	6	CG	ARG	A 3	11.333	43.643	48.531	1.00	66.73	C
ATOM	7	CD	ARG	A 3	12.359	42.537	48.421	1.00	60.73	C
ATOM	8	NE	ARG	A 3	11.615	41.284	48.391	1.00	65.23	N
ATOM	9	CZ	ARG	A 3	11.714	40.376	49.355	1.00	68.34	C
ATOM	10	NH1	ARG	A 3	12.529	40.581	50.376	1.00	71.04	N
ATOM	11	NH2	ARG	A 3	11.003	39.257	49.297	1.00	63.61	N
ATOM	12	N	PRO	A 4	9.438	43.779	50.953	1.00	46.26	N
ATOM	13	CA	PRO	A 4	9.010	43.629	52.374	1.00	41.86	C
ATOM	14	C	PRO	A 4	7.887	44.569	52.763	1.00	39.72	C
ATOM	15	O	PRO	A 4	8.030	45.766	52.575	1.00	38.27	O
ATOM	16	CB	PRO	A 4	10.305	43.872	53.151	1.00	42.48	C
ATOM	17	CG	PRO	A 4	11.387	43.447	52.216	1.00	45.47	C

…

The PDB (Protein DataBank) file format

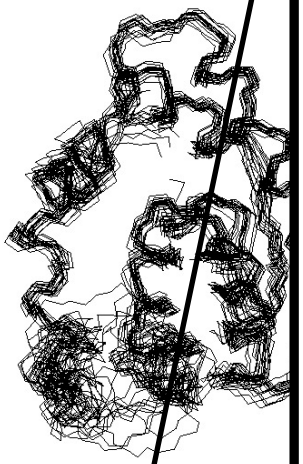
Final NMR structure ensemble



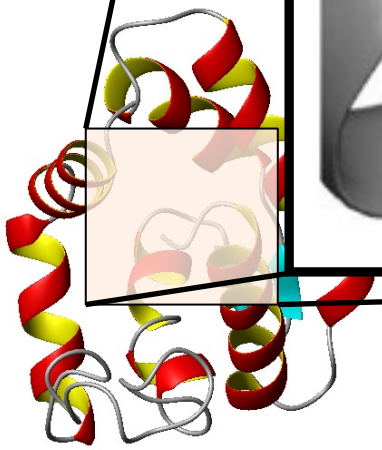
20-100 structures are generated and selected by low energy
They ***all*** satisfy experimental/geometrical energy
The rmsd (root-mean-square-deviation) indicates the precision of the structure

Not well-resolved regions are indicative of a lack of structural information:
-> either dynamic region or technical problems

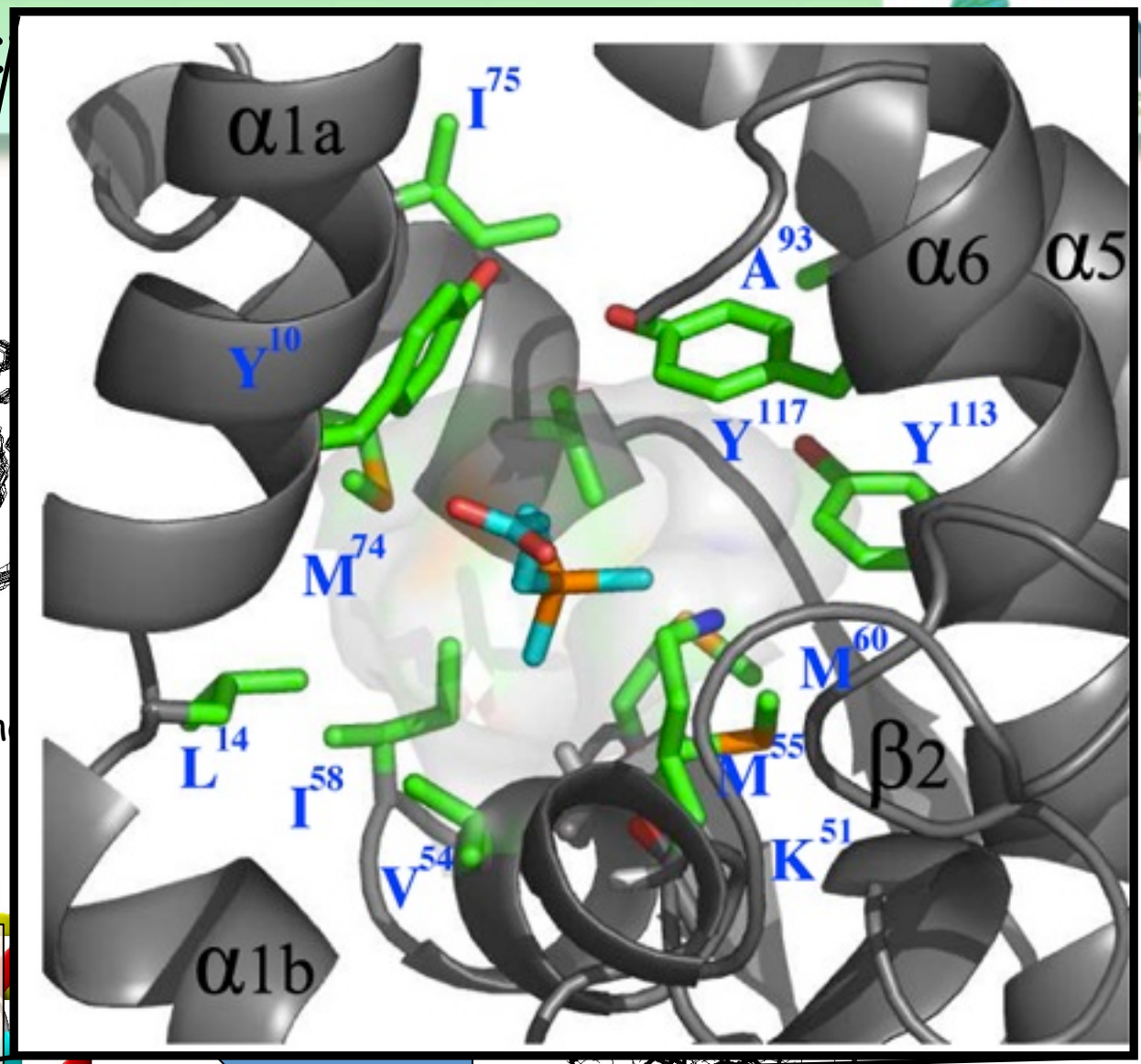
Struc

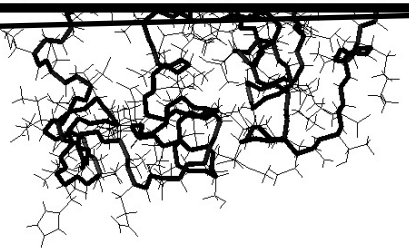


20 structures /
Backbone bonds (no sidech

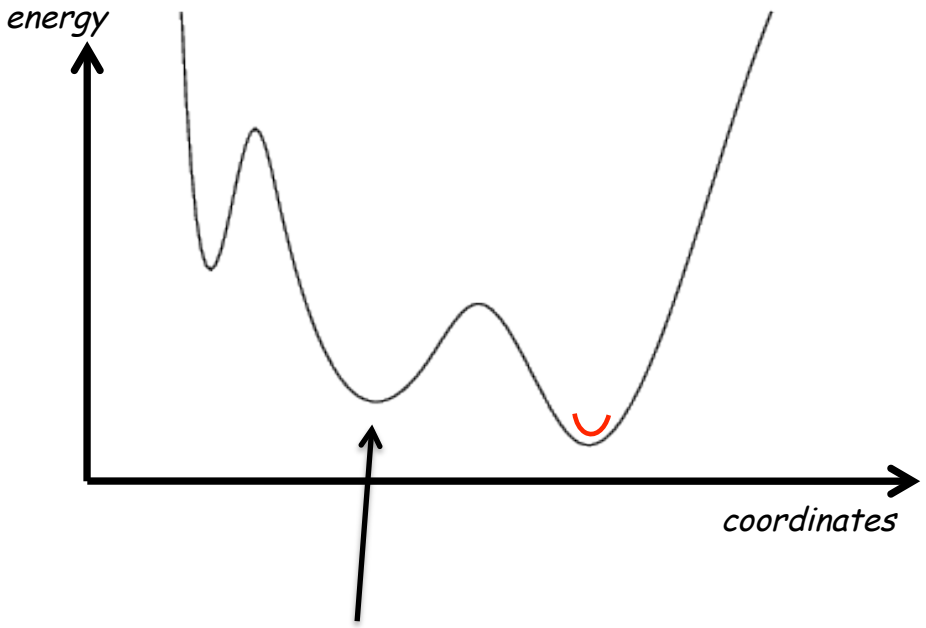


Ribbons
(secondary structures)





Alternative conformations?



What if a second conformer is significantly populated?

How observed experimental data are affected by multiple conformations?

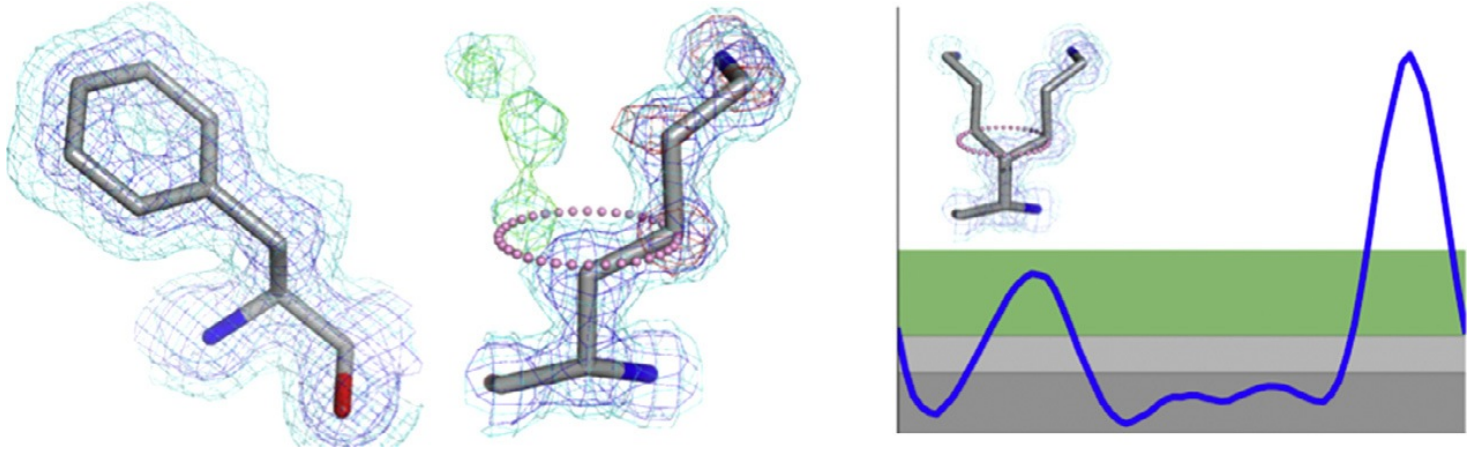
Alternative conformations?

Techniques that give a distribution of experimental data:

- Xray diffraction (distribution of electron density)
- NMR (slow exchanging conformations)
- PELDOR (EPR): distance between two radicals

-> structural data are obtained directly for each conformer

Conformational heterogeneity in electron density maps



Lowly populated conformers are often overlooked in electron density

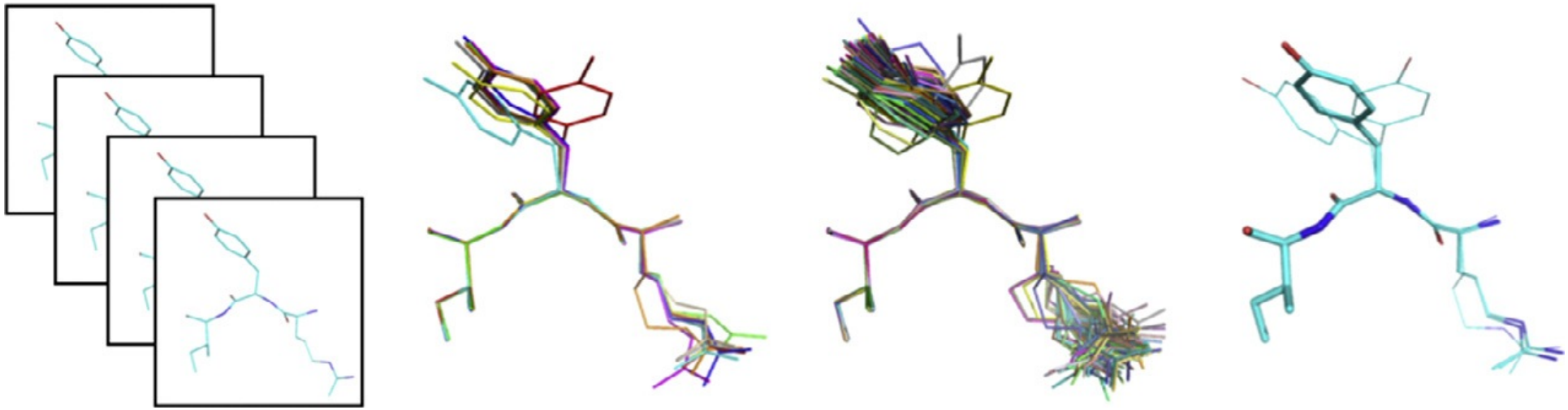
Electron density maps contain ensemble-averaged information about multiple protein conformations. (a) The spread in electron density (blue mesh, 1s high contour; cyan mesh, 0.5 s low contour) around each atom is approximated by a B-factor, which models the thermal motion as a Gaussian displacement about the mean position. (b) An electron density map with multiple maxima (blue mesh, high contour; cyan mesh, low contour) is inadequately modeled by B-factors, resulting in difference map peaks (red mesh, 1.5s Fo–Fc peak underlying the model; green mesh, +1.5s Fo-Fc peak indicating potential placement of alternative conformations). Because the alternative conformation partially overlaps with the primary conformation and is at lower occupancy, it is not visible at high contour. (c) Sampling the electron density around the x1 dihedral angle of the map shown in (b) (pink dots) reveals the presence of a distinct peak at the rotameric angle of -608, providing an anchor point for manual model building of an alternative conformation. Automated model building is further complicated by the potential for backbone motions that can shift these peaks out of rotameric angles

RA Woldeyes, DA Sivak and JS Fraser Current Opinion in Structural Biology 2014, 28:56–62

Working at cryo temperature tends to reduce the number of alternative conformations, ie to restrict the conformational space.

Working at room-temperature allows all « relevant » conformers to be sampled. The problem are radiation damages at RT can be alleviated with XFEL?

Multi-copies or structural ensembles from electron density maps



Refinement of Multiple Independent copies improves the Rfree

Different model types are being used to interrogate conformational heterogeneity. (a) In multiple independent refinements, each copy contributes to a distinct set of model structure factors. The distinct structures, separated here by boxes, yield an estimate of the relative precision of the refinement method. (b) In multi-copy ensemble refinement, a set number of copies of the protein, represented here by different colors, are refined together. (c) Similarly, in time-averaged ensemble refinement, multiple copies of the protein are selected from an MD simulation where the structure factors are averaged over a defined time window. (d) In multiconformer approaches, an optimal combination of between 1 and 4 conformations with associated occupancies (represented here by the thickness of the sticks) is constrained to sum to unit occupancy for each residue. RA Woldeyes, DA Sivak and JS Fraser Current Opinion in Structural Biology 2014, 28:56–62

Alternative conformations?

Techniques whose results depend on image manipulation:

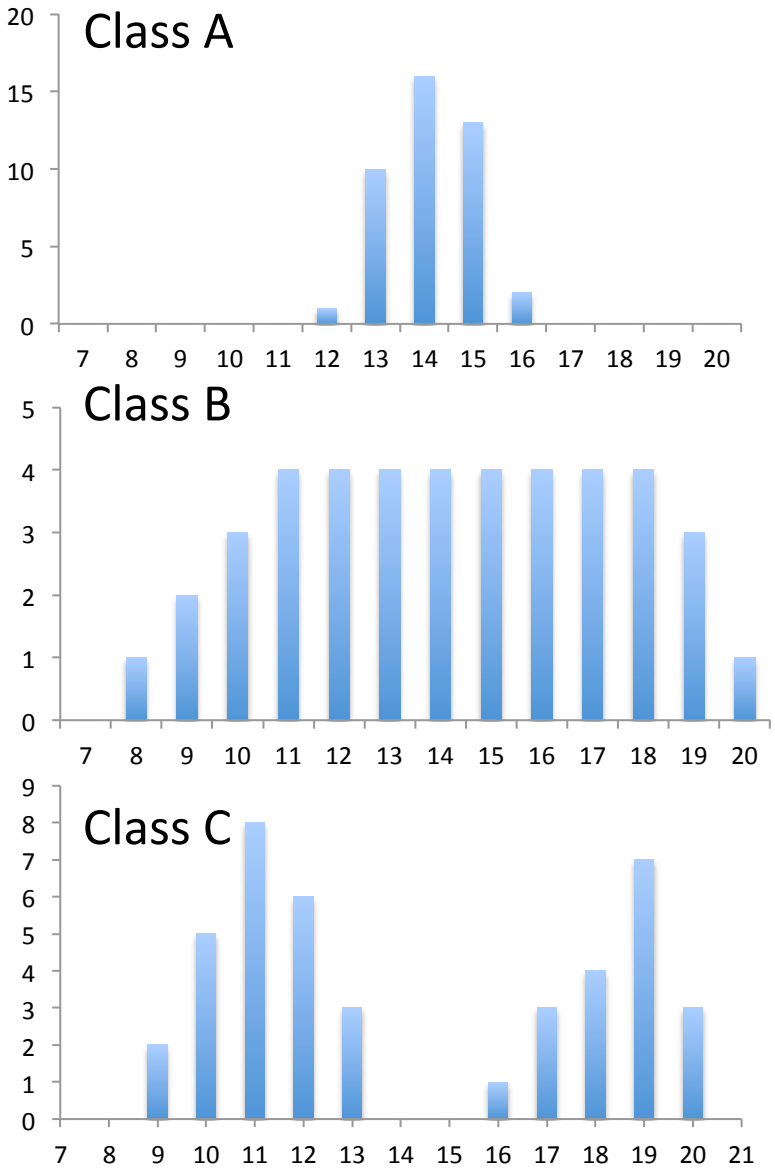
-> Cryo-EM

Techniques that give a population-averaged value:

- SAXS
- NMR (fast exchanging conformations, low energy barrier)
- (bulk) FRET

Difficulty: how to achieve the “deconvolution” without *a priori* model

Average and distribution



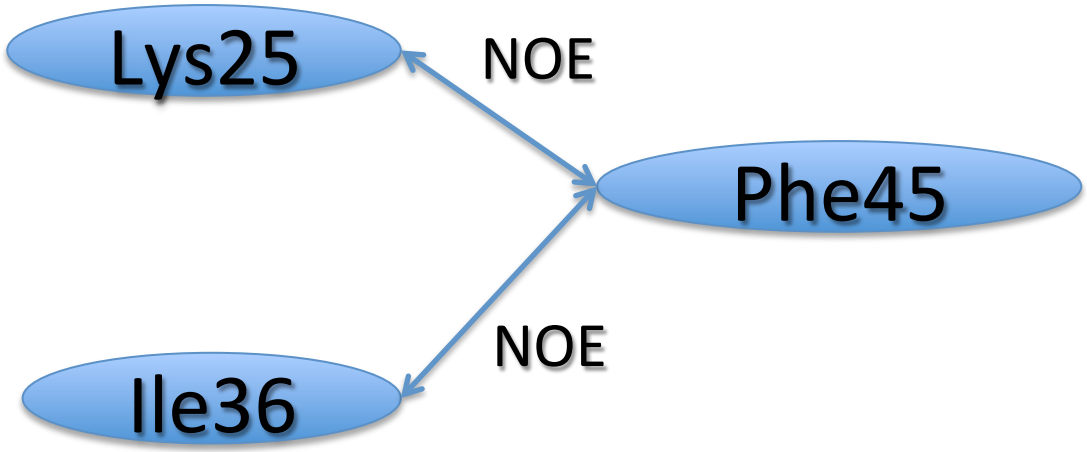
2 classes of 42 students: every student has a grade over 20. Both class has a average of 14/20.

Saying that the grade 14/20 is representative of the class, “average student” makes the **implicit hypothesis of a monomodal distribution**.

Classes A and B have same average but very different dispersion.

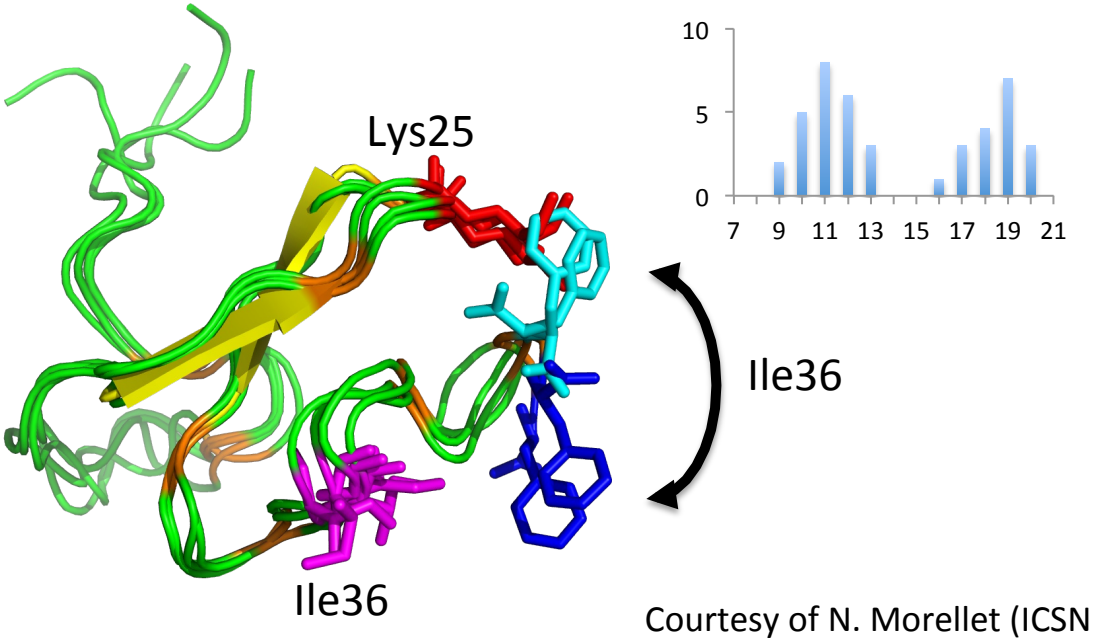
No student has the grade 14/20 in class C.

Ensemble-averaging



Impossible to satisfy both distance in a single conformation.

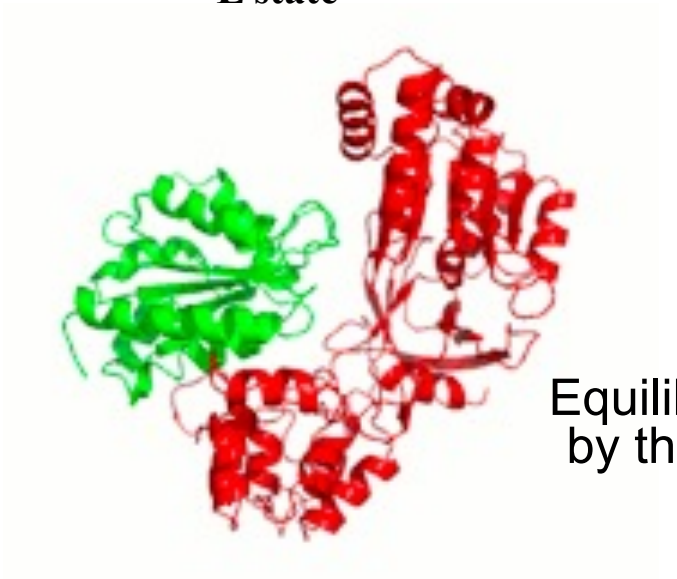
Solution: ensemble averaging. The NMR restraints are averaged out over the ensemble.



Courtesy of N. Morellet (ICSN)

The CPR bidomain enzyme

L state **U state**



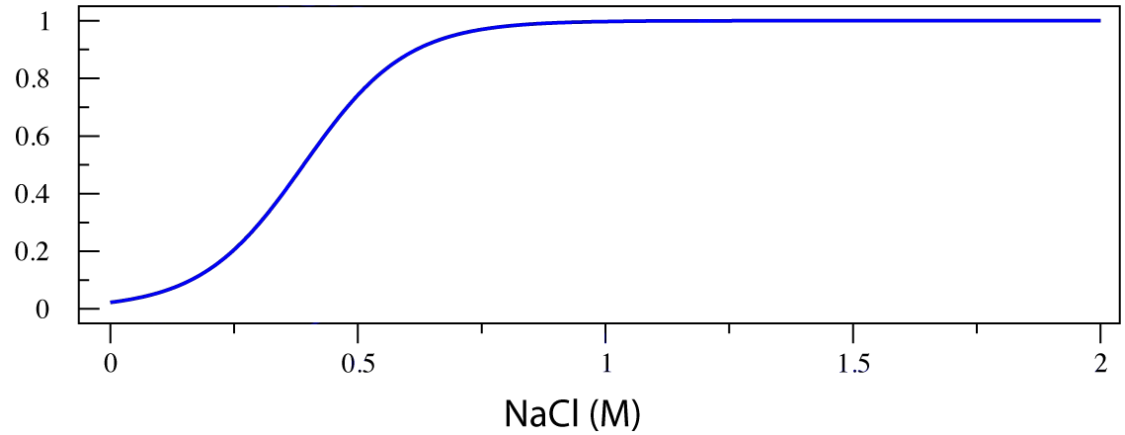
k_{ex} >>> 10^3 s^{-1}

?

Equilibrium controlled by the ionic strength

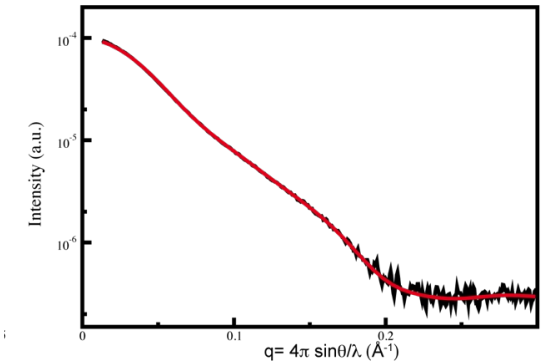
The fraction of U state depends on NaCl concentration

$p_U = 1 - p_L$

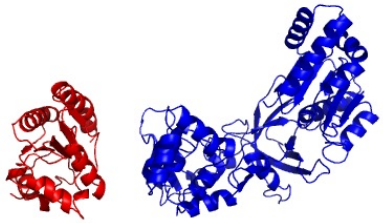


Frances O, Fatemi F, Pompon D, Guittet E, Sizun C, Pérez J, Lescop E, Truan G. Biophys J. 2015;108(6):1527-36

Reproduction of high salt SAXS

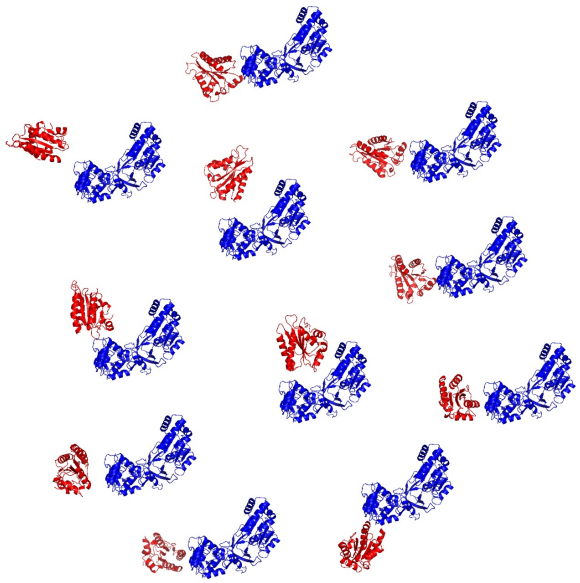


Using DADIMODO



or

EOM-derived ensemble

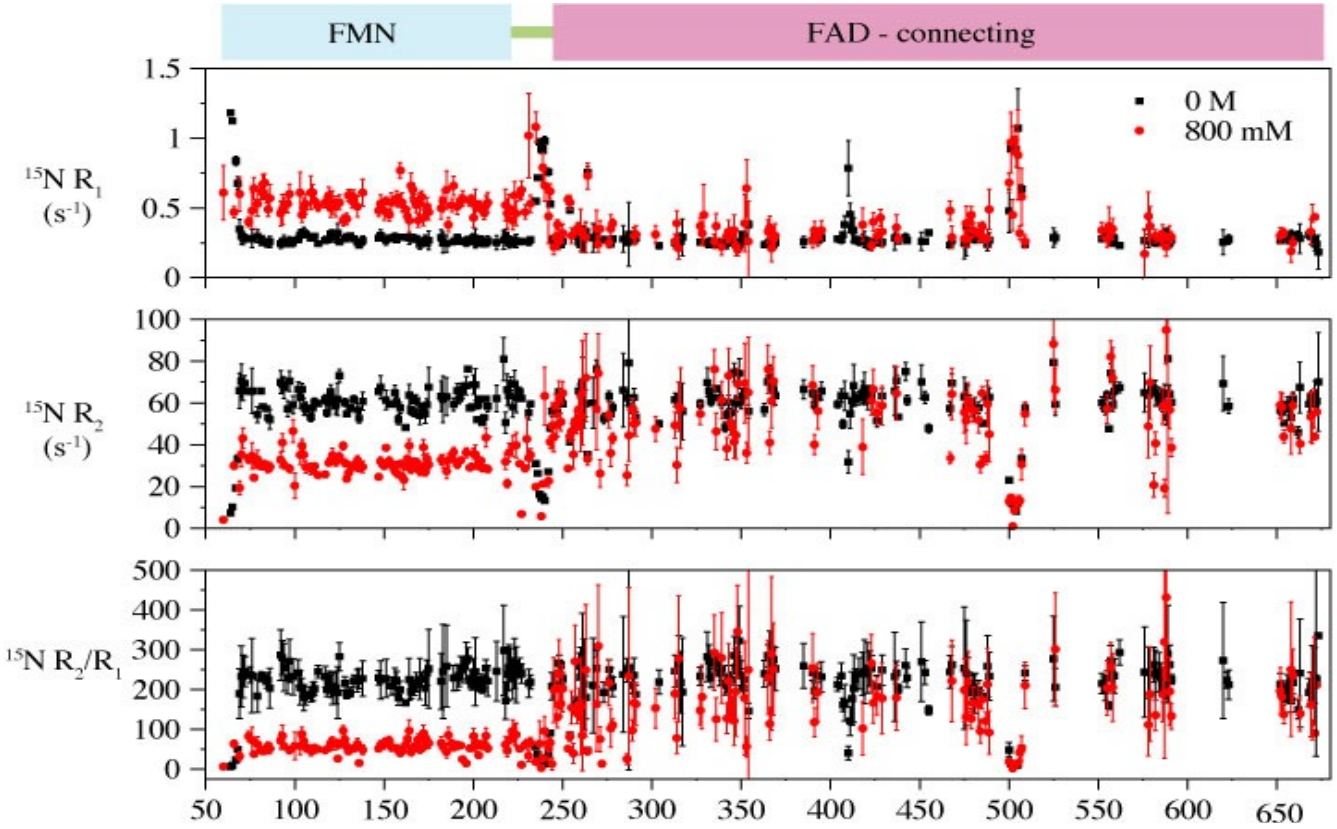


Correctly fits the SAXS data
However, no interdomain interaction!
What stabilizes this conformation?

The ensemble correctly also fits the SAXS data
No further stable interdomain interface
Satisfy the idea of interdomain mobility

Frances O, Fatemi F, Pompon D, Guittet E, Sizun C, Pérez J, Lescop E, Truan G. Biophys J. 2015;108(6):1527-36

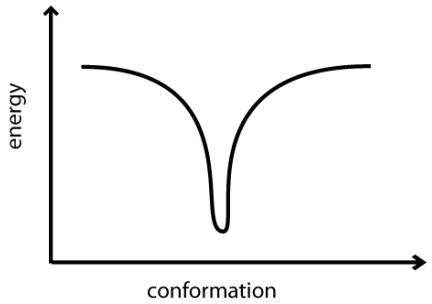
NMR ^{15}N relaxation

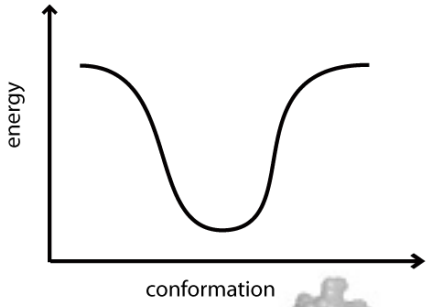


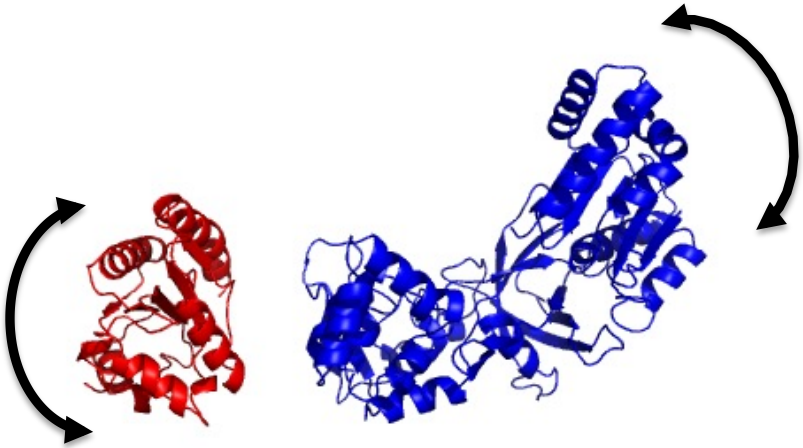
	FAD domain	FMN domain
τ_c (0 M)	30 ns	30 ns
τ_c (0.8 M)	25 ns	15 ns

Ensemble of conformations in dynamic equilibrium (sub tc)

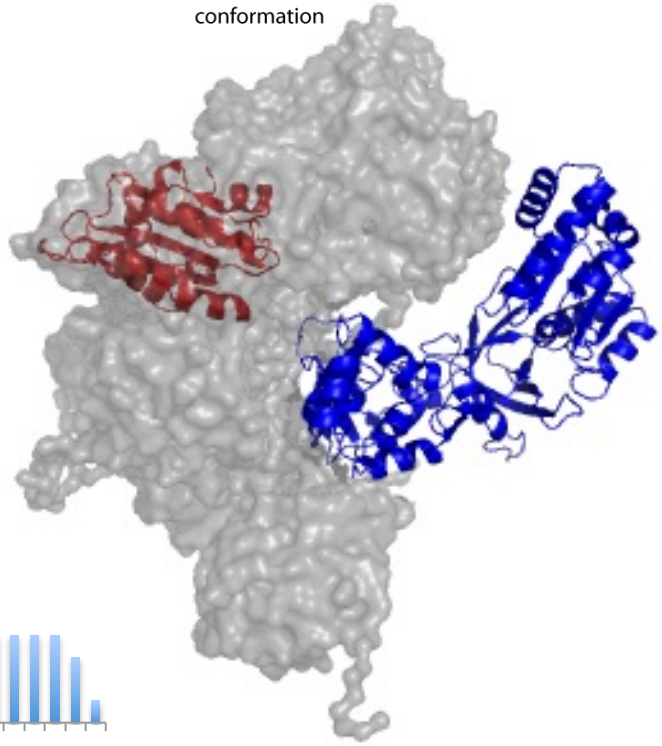
Representation?

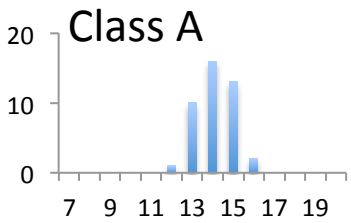


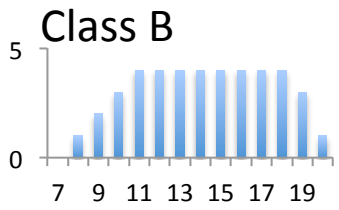




versus



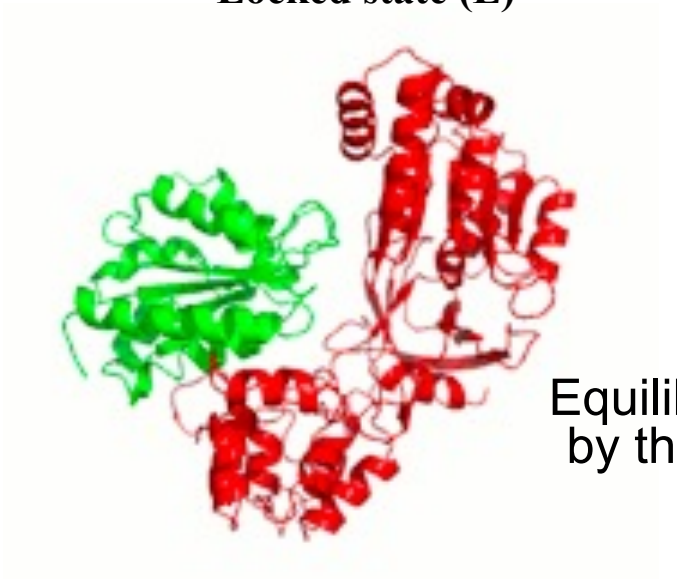




Does the best-fit structure is really representative of the “real” structure?
How to represent the conformational sampling (ensemble)?

The CPR bidomain enzyme

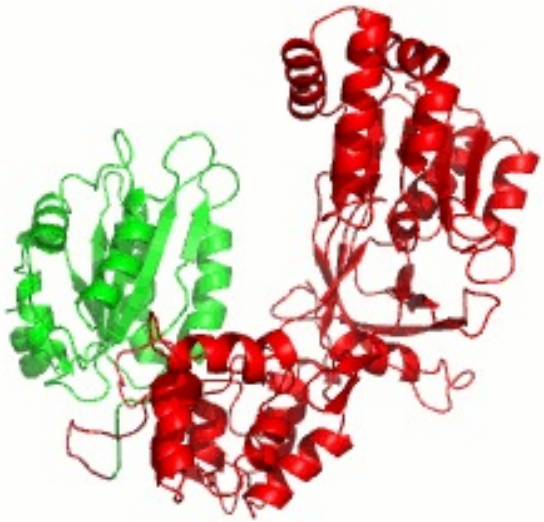
Locked state (L)



k_{ex} >>> 10^3 s^{-1}

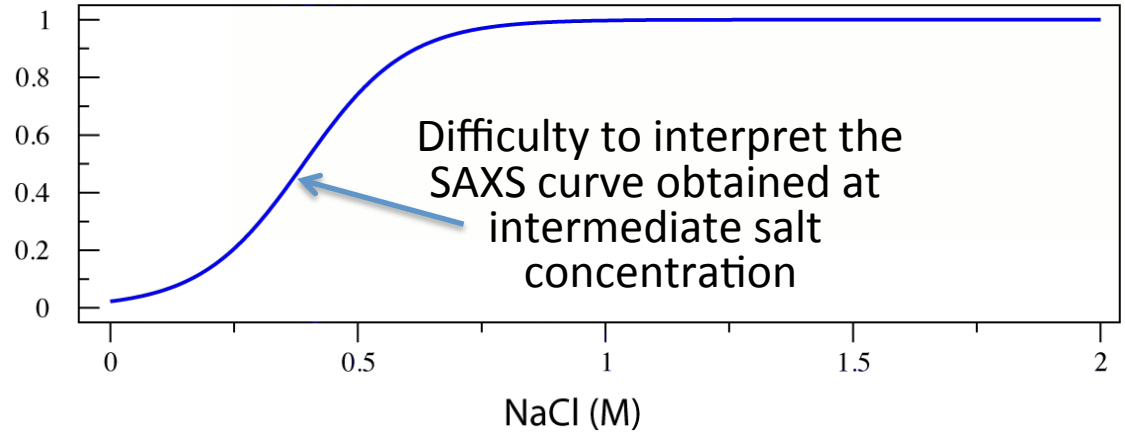
Equilibrium controlled by the ionic strength

Unlocked state (U)



The fraction of U state depends on NaCl concentration

$p_U = 1 - p_L$



Frances O, Fatemi F, Pompon D, Guittet E, Sizun C, Pérez J, Lescop E, Truan G. Biophys J. 2015;108(6):1527-36

Interpretation of SAXS data

Structure refinement makes sense for ***monodisperse samples*** and in absence of any diversity in size or shape (**polymorphism**): ***identical particules***

Polymorphism can be assessed by NMR, EM, …

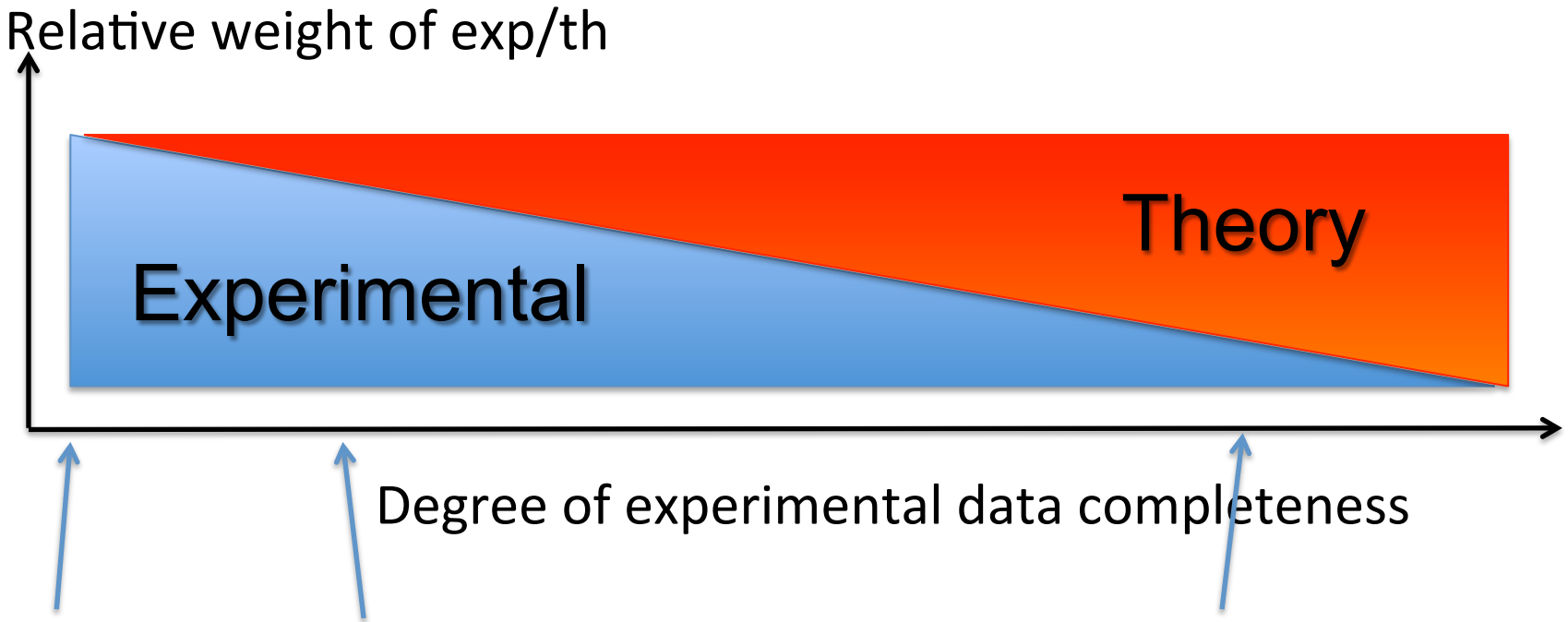
When calculating the ***SAXS envelope***, we do the **implicit hypothesis** of a highly rigid molecule. Be careful!

The best-fit structure is often called the average structure. Take care when using this expression in case of large-scale motion.

Further reading: http://www.embl-hamburg.de/biosaxs/courses/embo2014/slides/flexible-systems-bernado.pdf

Hammel M Eur Biophys J (2012) 41:789–799

Hybrid energy



Very high resolution (<1Å) Xray diffraction data

High resolution (>1Å) Xray diffraction data / NMR / high-res Cryo EM

1. Highly flexible systems (IDP)
2. Sparse data (low-res Cryo EM / NMR / MS / cross link / EPR / …)

Overdetermined experimental data

Underdetermined experimental data

Intrinsically disordered proteins

IDP = ensemble of rapidly interconverting conformations
The conformations are very different from each other.
The averaged structure is meaningless.

Kragelj J, Ozenne V, Blackledge M, Jensen MR. Chemphyschem. 2013 Sep 16;14(13):3034-45

IDP: experimental data

IDP are highly flexible and therefore best studied in solution.
Preferably at room temperature.

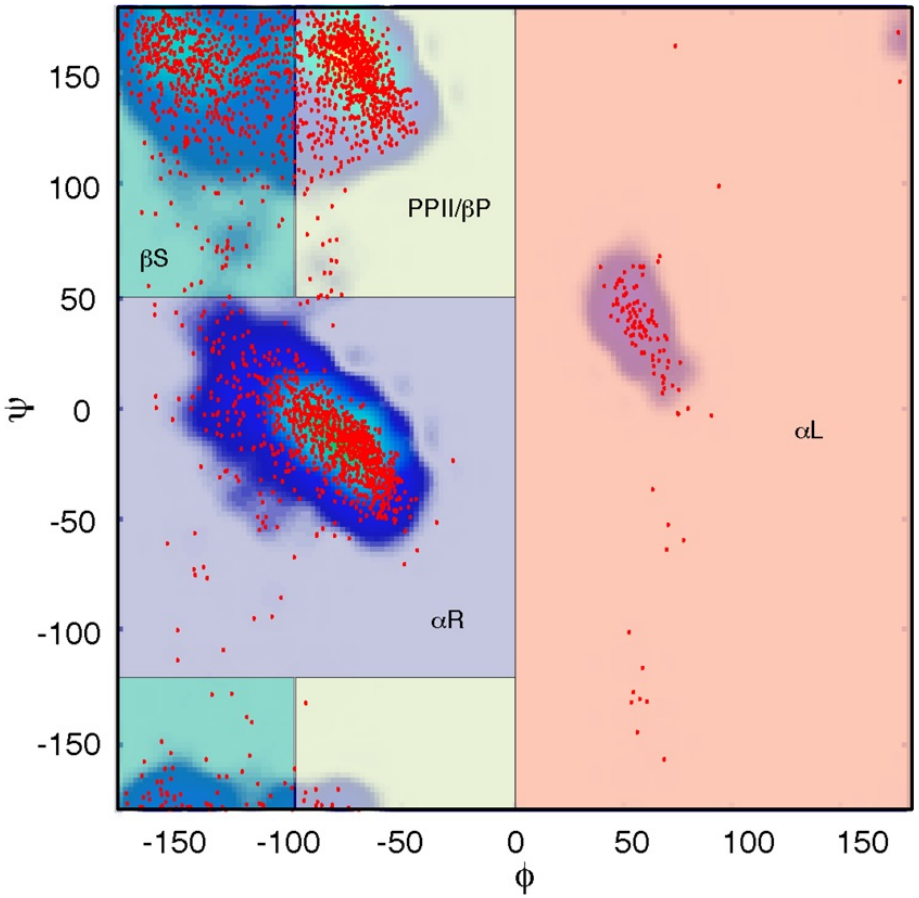
Local structural data:

- NMR: chemical shifts (CSs)
  - residual dipolar couplings, (RDCs)
  - J-couplings
  - hydrogen-exchange protection factor
  - relaxation rates
  - solvent-accessibility

Long range structural order:

- NMR paramagnetic relaxation enhancements (PREs)
  - nuclear Overhauser effects (NOEs)
- hydrodynamic parameters
- SAXS
- and why not FRET, …

IDP



Ramachandran plot showing the statistical coil sampling of threonine (red points and density map from low, blue, to high, red population). The Ramachandran plot is divided into four regions: **α**R (purple), **α**L (salmon), **β**S (green), and **β**P (yellow).

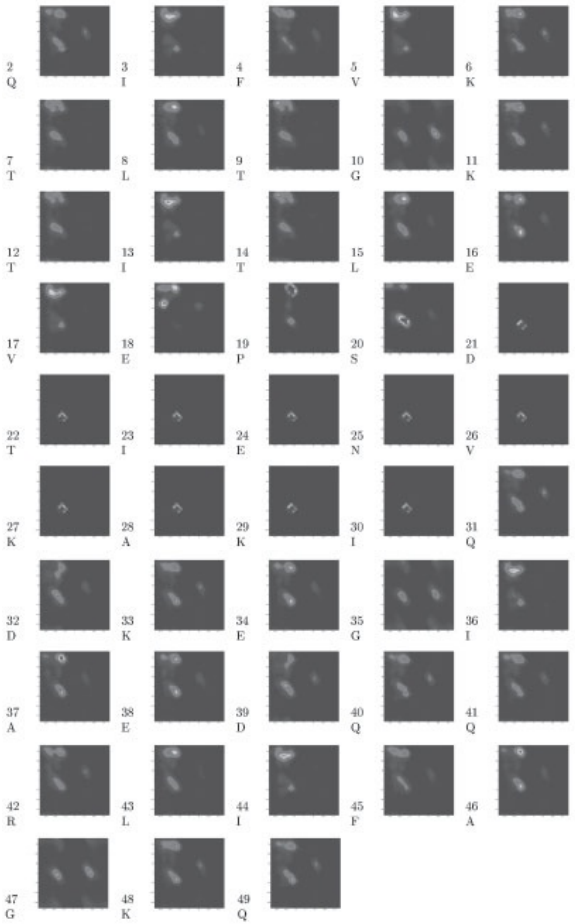
Hypothesis behind IDP:

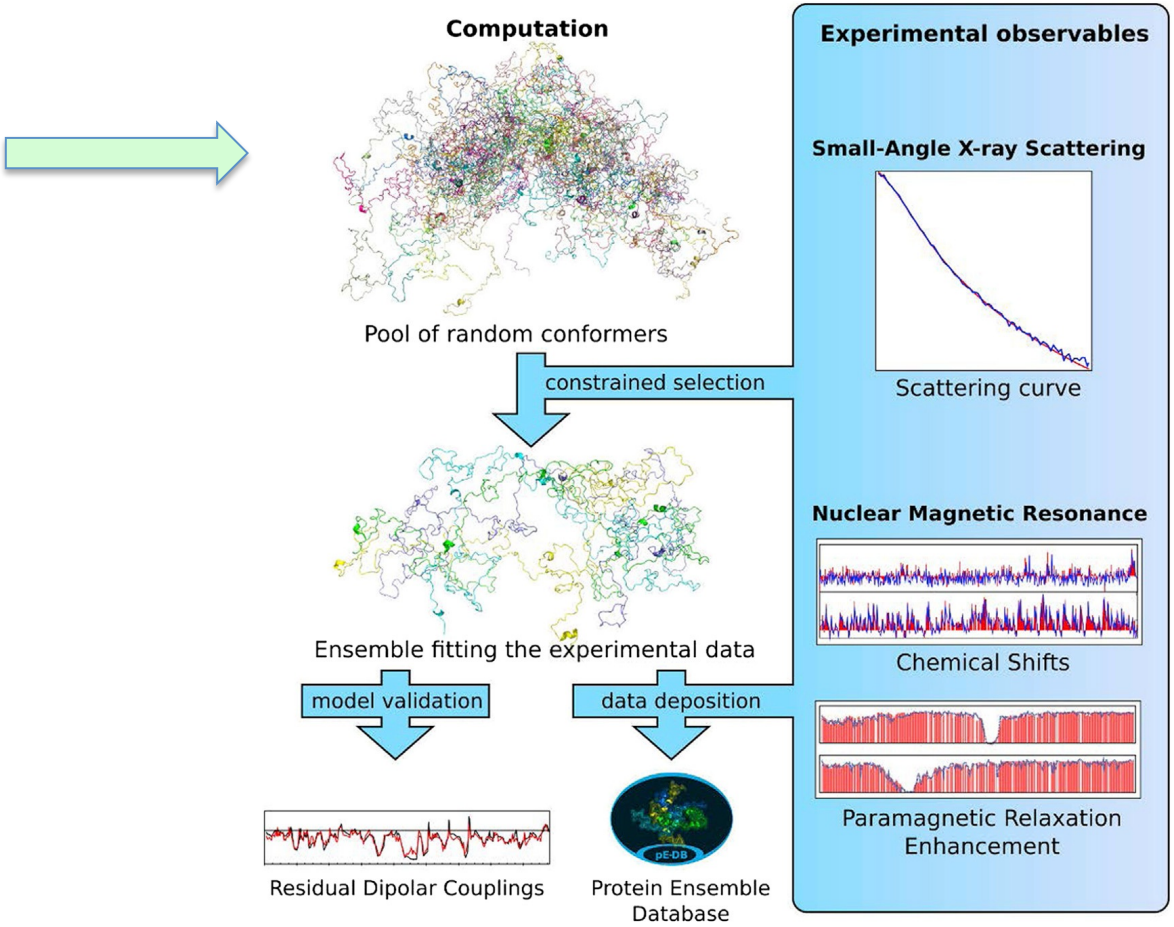
All energetically allowed conformers are rapidly sampled.
Some local transient structural preference may exist.

Ergodic theory:
The time-average properties of one molecule is similar to the ensemble-average of all molecules at a given time

Jensen MR, Zweckstetter M, Huang JR, Blackledge M. Chem. Rev., 2014, 114 (13), pp 6632–6660

Flexible-Meccano



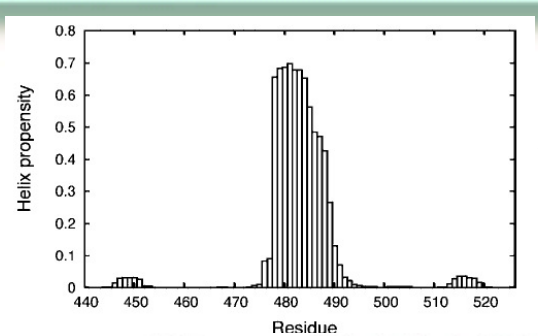


Original or biaised Ramachandran map for every amino acid, corrected for first neighbours

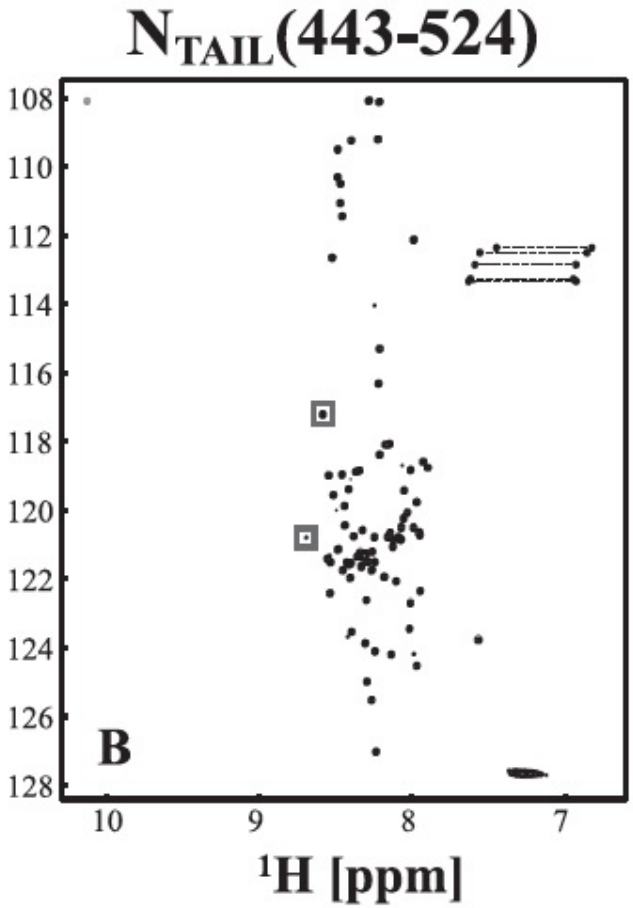
Protein Ensemble DataBase http://pedb.vib.be/

Schwalbe, M., Ozenne, V., Bibow, S., Jaremko, M. Jaremko, L., Gajda, M., Jensen, M.R., Biernat, J., Becker, S., Mandelkow, E., et al. (2014). Structure 22, 238–249

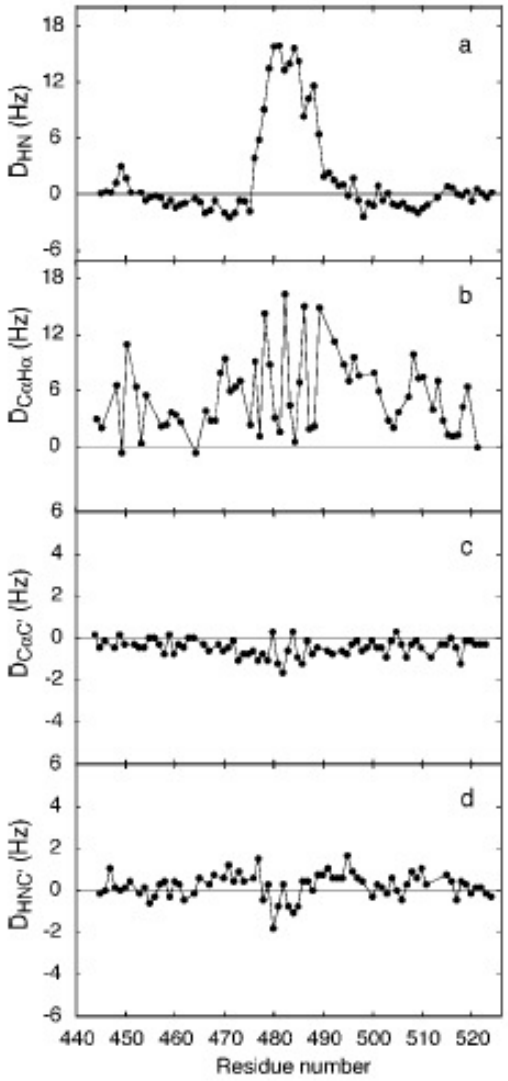
Ntail: an IDP



AGADIR



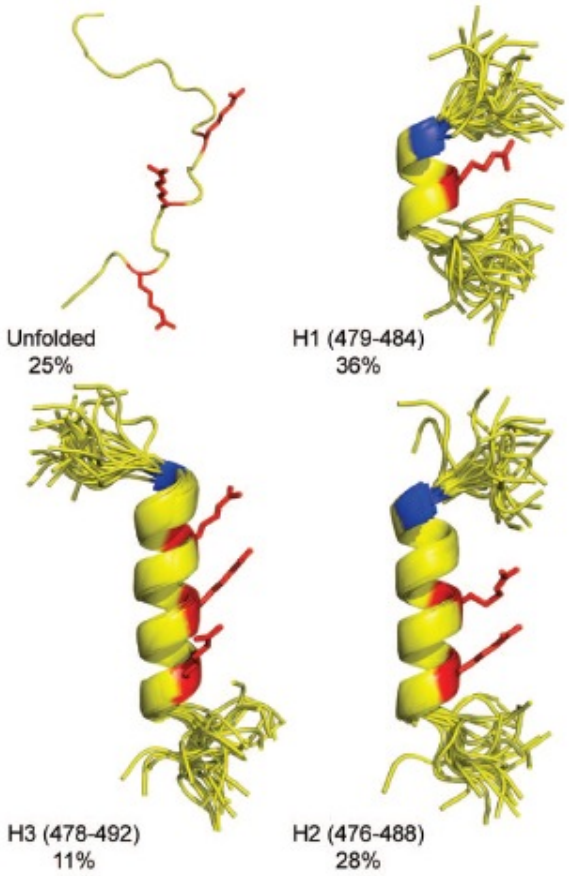
Residual Dipolar Couplings



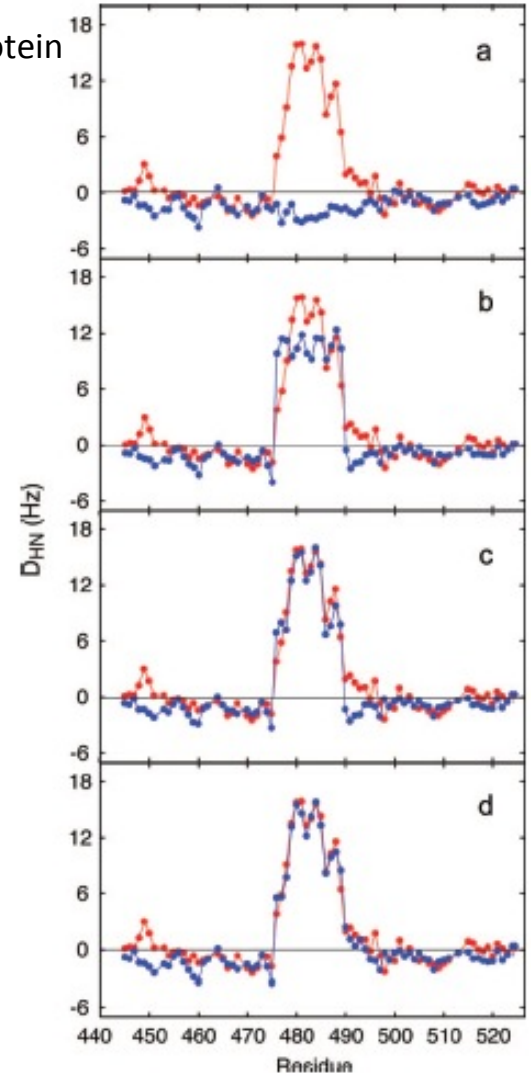
Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, Blackledge M.J Am Chem Soc. 2008;130(25):8055-61.

Ntail

Molecular Recognition Element of Sendai Virus Nucleoprotein



4 ensembles representing Ntail in solution



100% random coil

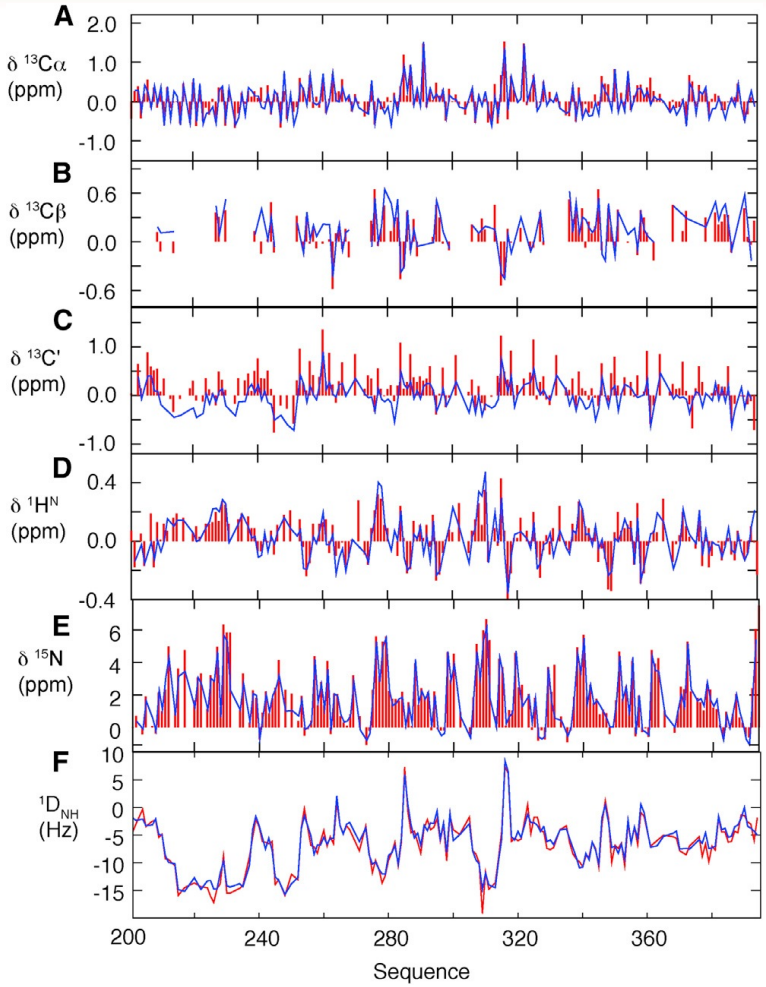
52 % random coil
+ 48% H1 (476-488)

18% random coil +
37% of helix (476-488)
+ 45% of helix (479-484)

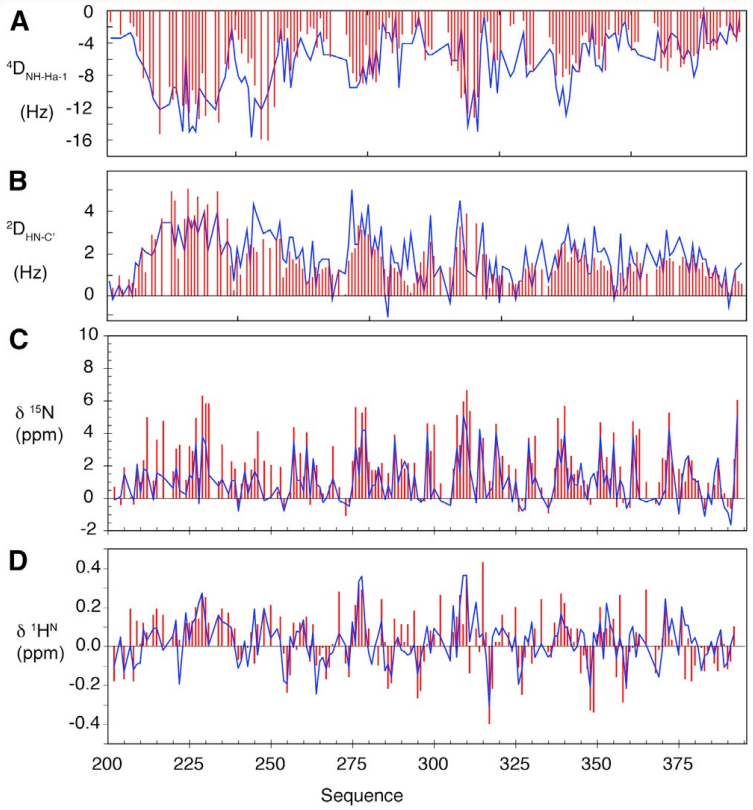
25% random coil;
+28% of helix (476-488)
+36% of helix (479-484)
+11% of helix (478-492)

Jensen MR, Houben K, Lescop E, Blanchard L, Ruigrok RW, Blackledge M.J Am Chem Soc. 2008;130(25):8055-61.

Cross-validation



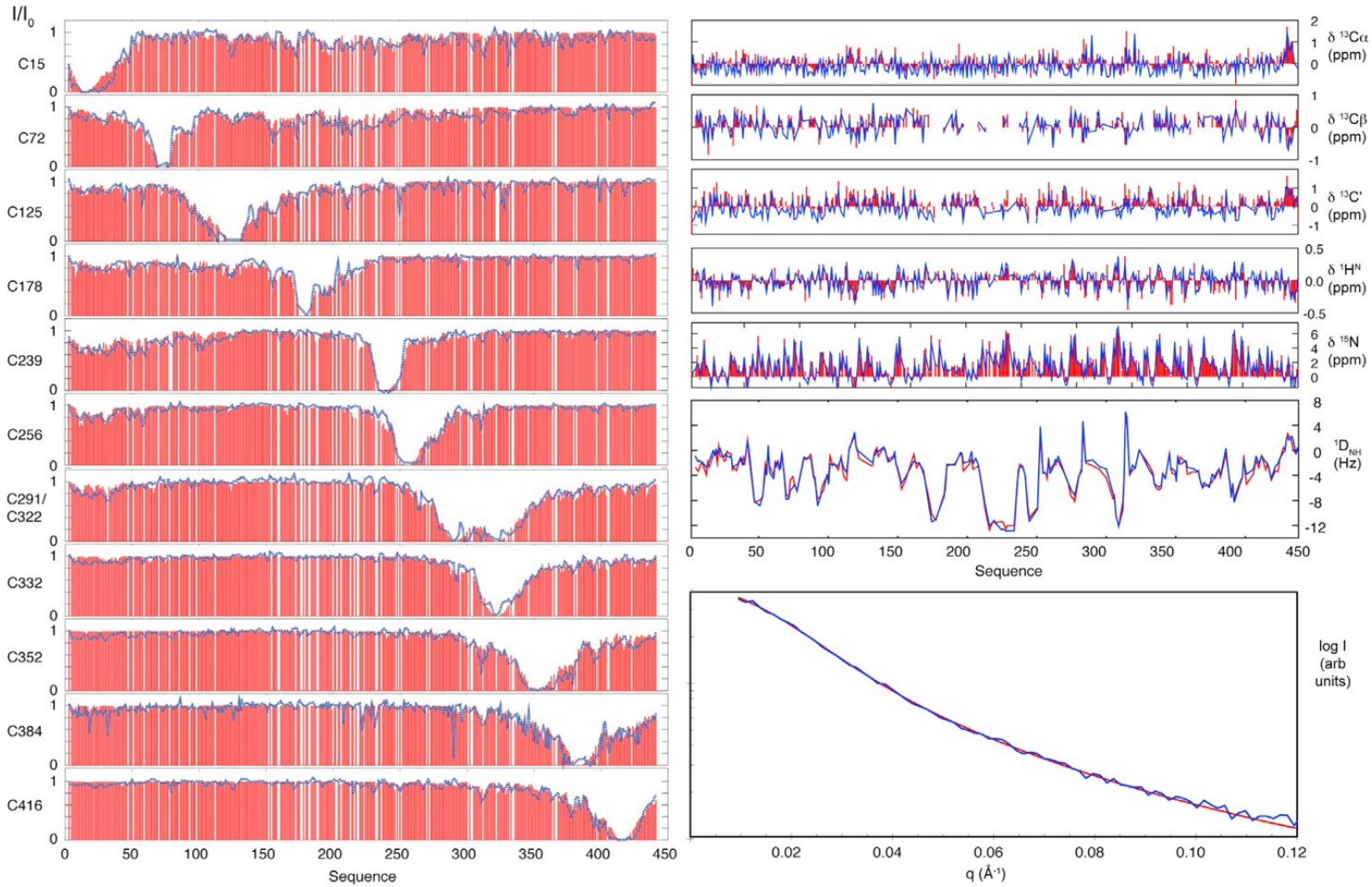
Reproduction of Experimental Data from K32 when Included as Active Data in the ASTEROIDS Target Selection Function
(A–E) Experimental (red bars) and calculated (blue lines) CS data.
(F) Experimental RDCs (red) and values calculated from ASTEROIDS selection (blue).



Cross-Validation of K32 Ensemble
Reproduction of experimental (red) data from K32 using ASTEROIDS
ensemble selection (blue) when not included as active data in the ASTEROIDS target selection function.

Schwalbe M, Ozenne V, Bibow S, Jaremko M, Jaremko L, Gajda M, Jensen MR, Biernat J, Becker S, Mandelkow E, Zweckstetter M, Blackledge M. Structure. 2014 Feb 4;22(2):238-49.

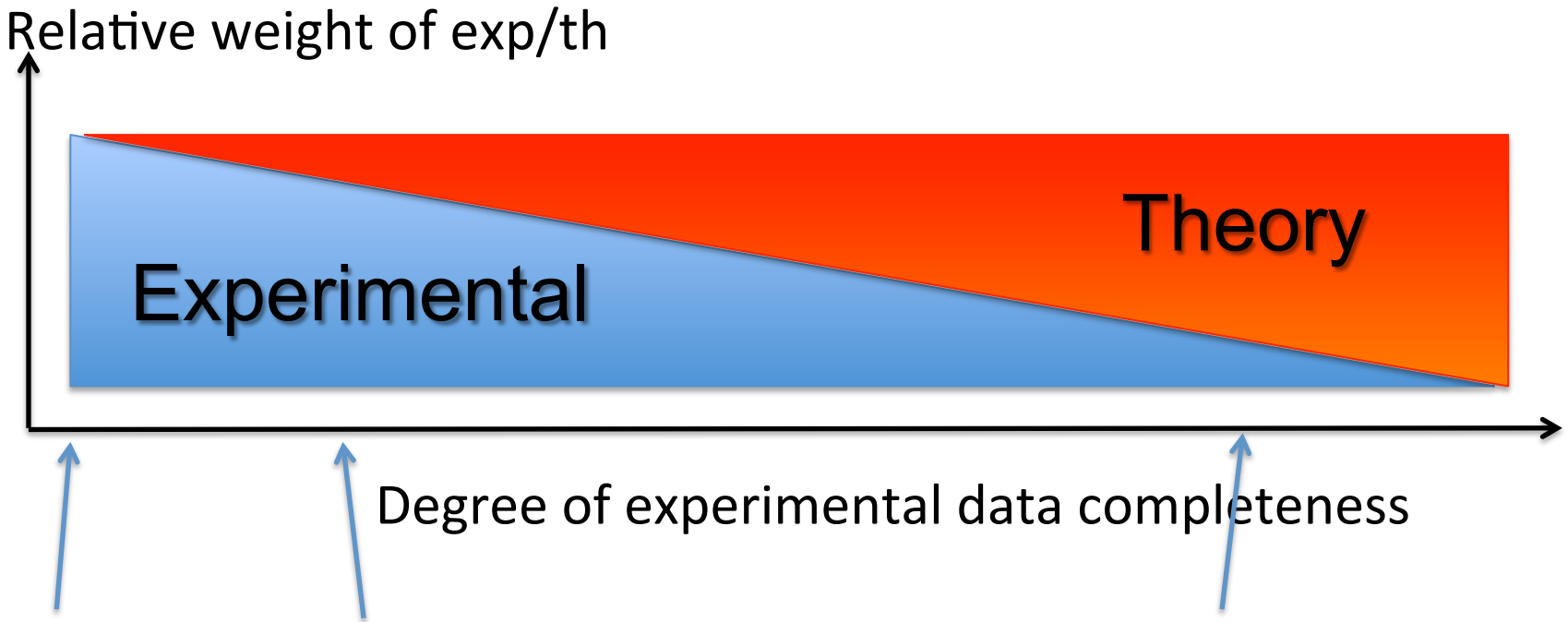
Tau protein



Reproduction of Experimental Data from 441-Residue Tau Using ASTEROIDS Left: experimental (red) and calculated (blue) PRE data. Top right: experimental (red) and calculated (blue) CS data. Mid-right: experimental (red) and calculated (blue) RDCs.
Bottom right: experimental (red) and calculated (blue) SAXS data.

Schwalbe M, Ozenne V, Bibow S, Jaremko M, Jaremko L, Gajda M, Jensen MR, Biernat J, Becker S, Mandelkow E, Zweckstetter M, Blackledge M. Structure. 2014 Feb 4;22(2):238-49.

Hybrid energy



Very high resolution (<1Å) Xray diffraction data

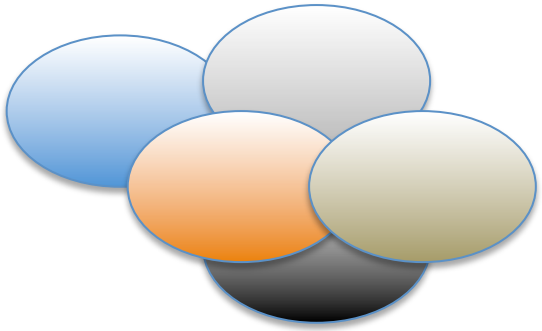
High resolution (>1Å) Xray diffraction data / NMR / high-res Cryo EM

1. Highly flexible systems (IDP)
2. Sparse data (low-res Cryo EM / NMR / MS / cross link / EPR / …)

Overdetermined experimental data

Underdetermined experimental data

Molecular assembly



might resist to crystallization
CryoEM has not been tried
Too large for *ab initio* NMR determination

Low or high-resolution structural data on *subunits* (crystal or NMR structure, SAXS envelop, EM map)

Collect a few restraints on the assembly

Modelling

A structural model of the assembly

Validation?

Experimental data

mutagenesis: binding assay

- ⊕ - residue level
- ⊖ - loss of structure should be checked

cross-linking: MS

- ⊕ - distance information
- ⊖ - difficult to attach and to detect

H/D exchange: MS/NMR

- ⊕ - residue level
- ⊖ - direct vs indirect effects
  - labeling needed for NMR

CSP: NMR

- ⊕ - atomic level
- ⊖ - direct vs indirect effects
  - labeling needed

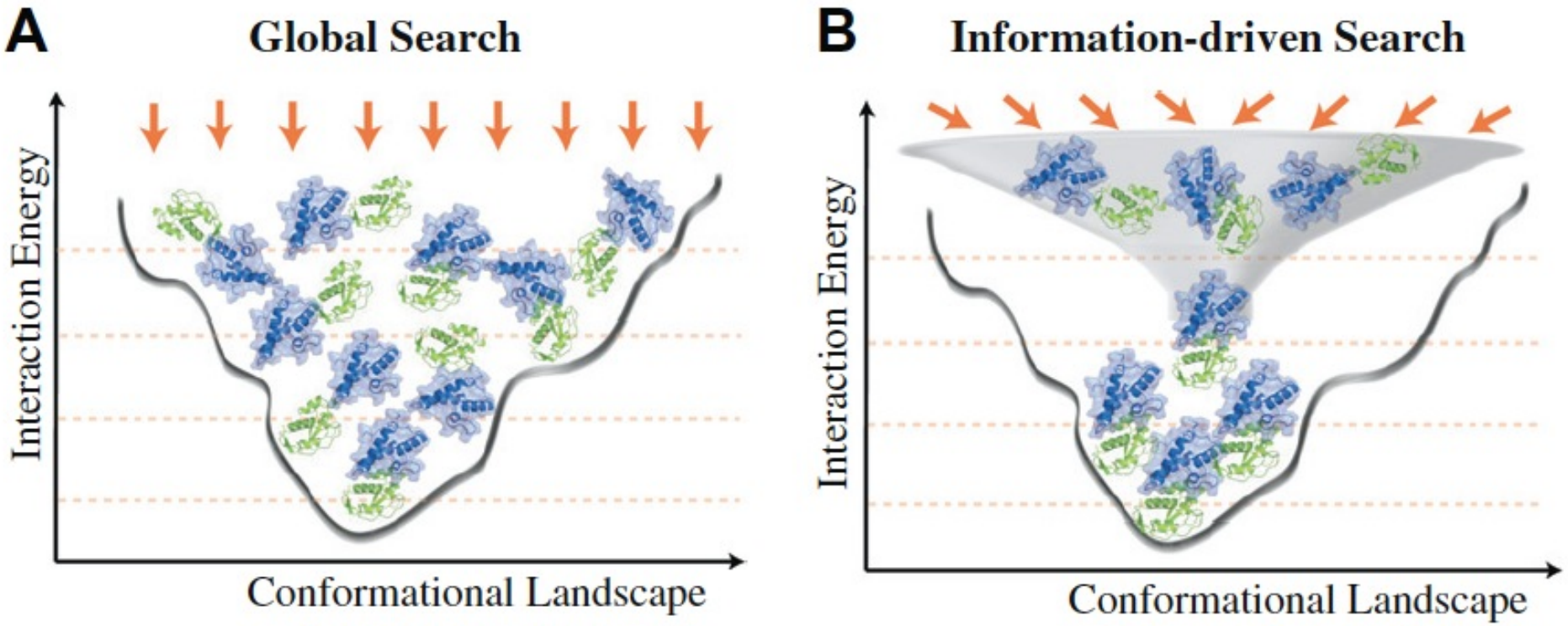
RDC, relaxation: NMR

- ⊕ - orientation information
- ⊖ - labeling needed

Experimental data

Information content	Resolution	Data types	Remarks
Interface	Residues	Mutagenesis	Readout by any binding essay; can give false positives
	Residues	H/D exchange	Read-out by MS or NMR
	Residues	chemical footprinting	Read-out by biochemistry
	Residues	NMR chemical shift mapping	Read-out by NMR
	Atoms	Solvent Paramagnetic Relaxation Enhancement	Read-out by NMR
Distance	Atom pairs	Crosslinking/MS	Long-range; linker length/flexibility
	Residue pairs	Correlated mutations	e.g., exchanges charges in a salt bridge
	Residue pairs	Label pairs FRET	Long-range, <80–100 Å
	Residue pairs	Label pairs EPR (PELDOR, DEER)	Long-range, <60–80 Å
	Atom pairs	(1H-1H) NOE	Read-out by NMR, < 5Å
	Atom pairs	(1H-X) Paramagnetic Relaxation Enhancement(PRE) or Pseudo Contact Shift (PCS)	Read-out by NMR, <40 Å
Orientation	Domain	Residual Dipolar Couplings	Read-out by NMR under anisotropic medium
	Domain	Relaxation anisotropy	Read-out by NMR ^{15}N relaxation
	Domain	Pseudo Contact Shift (PCS)	Read-out by NMR chemical shifts
Global Shape	Complex	Cryo-EM / Negative stain EM	Overall structure, >10 Å usually
		SAXS/SANS	Molecular envelope, Rg
		IM-MS	Collision cross-section
Intersubunit interaction	Domain	Proteomic, SPR, Pull-down, …	
Subunit position	Domain	Immuno-EM, Gold or GFP-labeling	Read-out by cryoEM

Global or data-driven search



E. Karaca, A.M.J.J. Bonvin / Methods 59 (2013) 372–381

Hybrid energy

$$E_{hybrid} = E_{phys} + w_{data} E_{data}$$

Physical interaction

Use structure of different origins

Rigid body approach:

- intra-subunit energy not calculated
- inter-subunit interaction

Experimental data

Heterogeneous sources of restraints

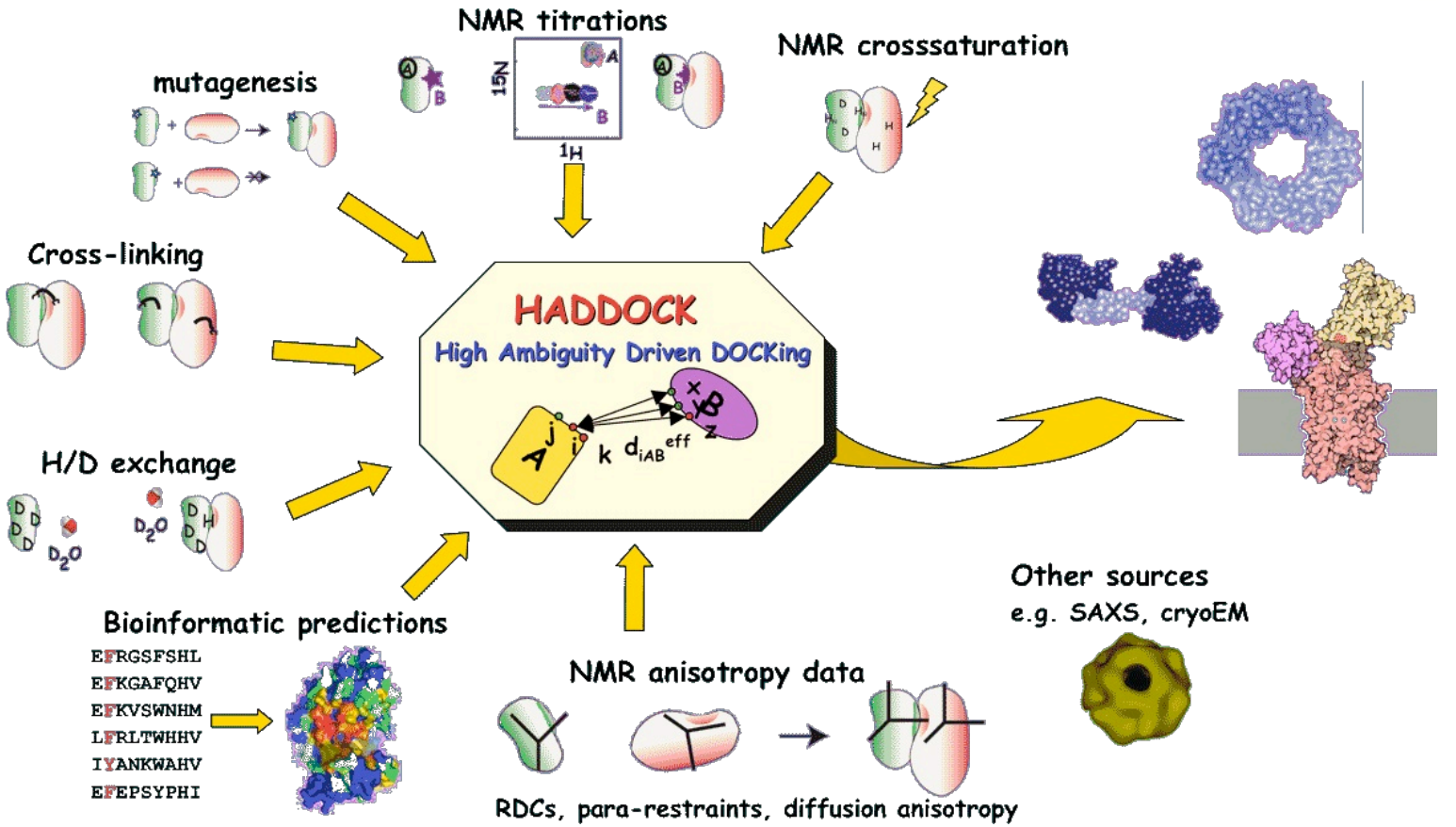
Convert into a pseudo-energy term (weight)

Limitations: conformational changes upon complexation, uniqueness, wrong conversion of data into restraints

Solution: local (small amplitude) rearrangement allowed in final step

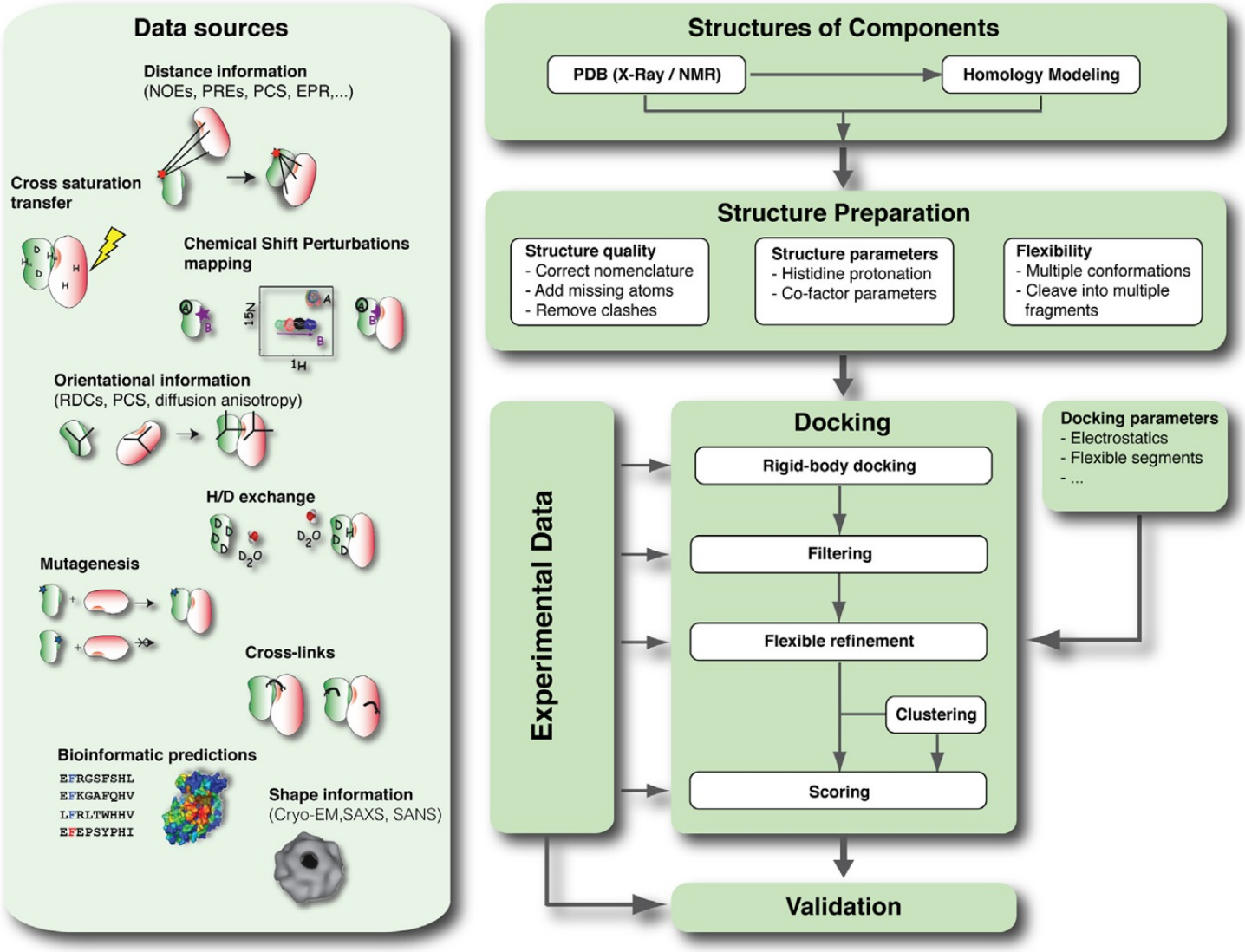
Haddock (A.M. Bonvin)

http://haddocking.org/



H. van Ingen, A.M.J.J. Bonvin / Journal of Magnetic Resonance 241 (2014) 103–114

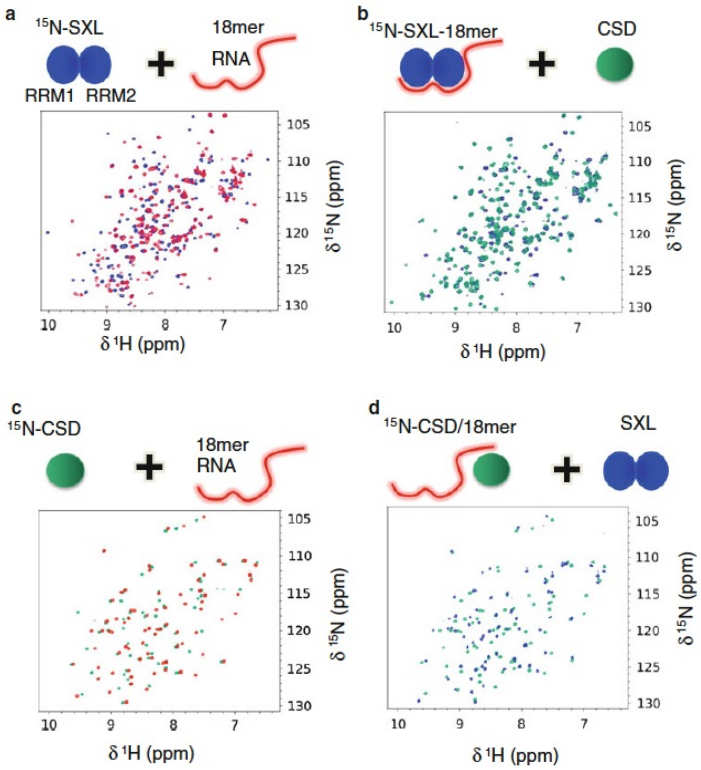
Haddock protocol



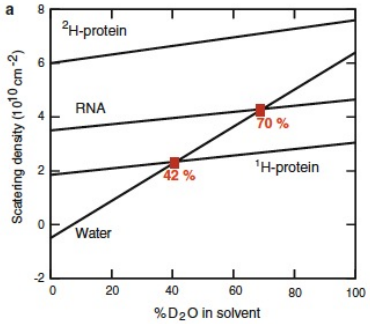
H. van Ingen, A.M.J.J. Bonvin Journal of Magnetic Resonance 241 (2014) 103–114

Example: a ternary protein-protein-RNA

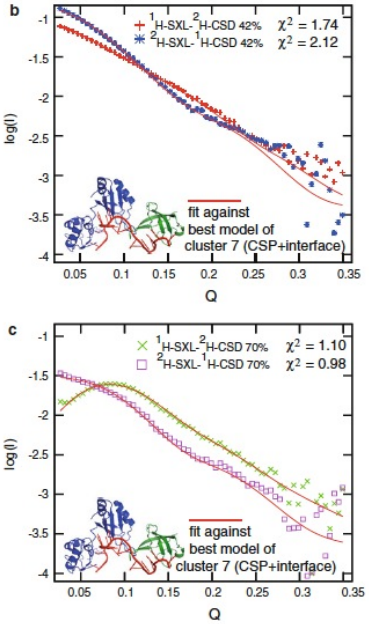
SXL: Two RRM domains
CSD: cold shock domain 1 from UNR (CSD)
18-mer RNA

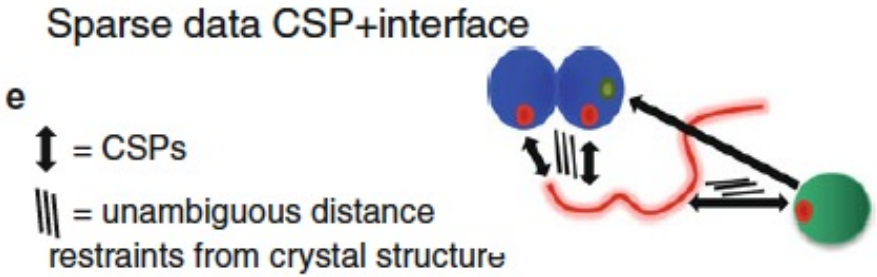


NMR



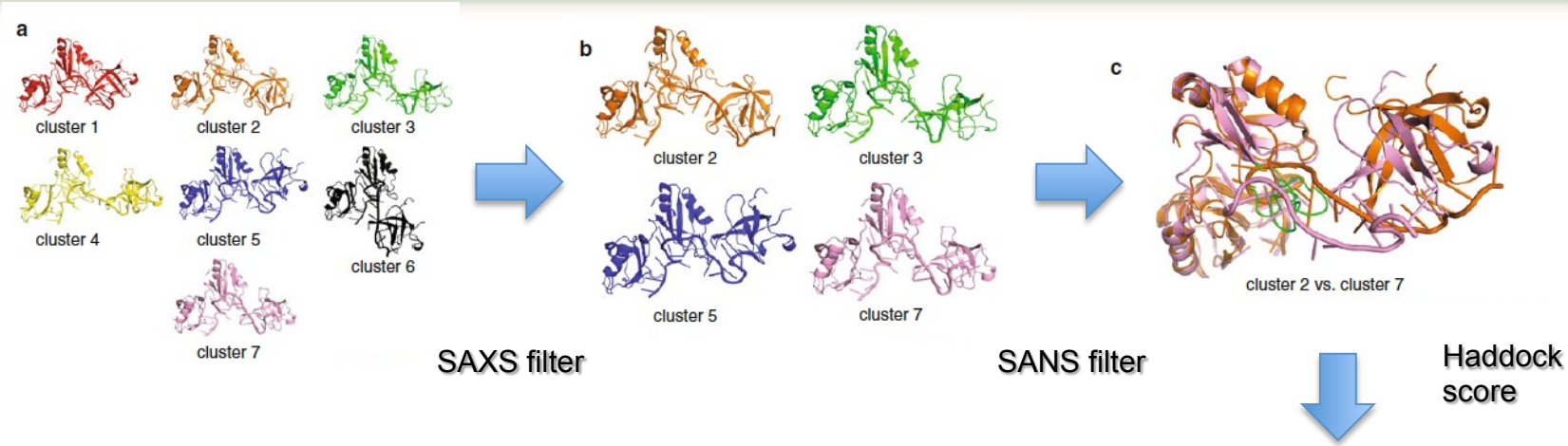
SANS
SAXS

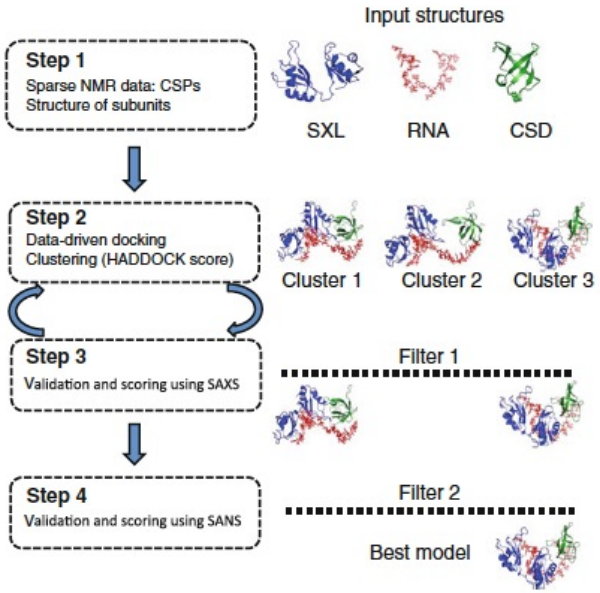


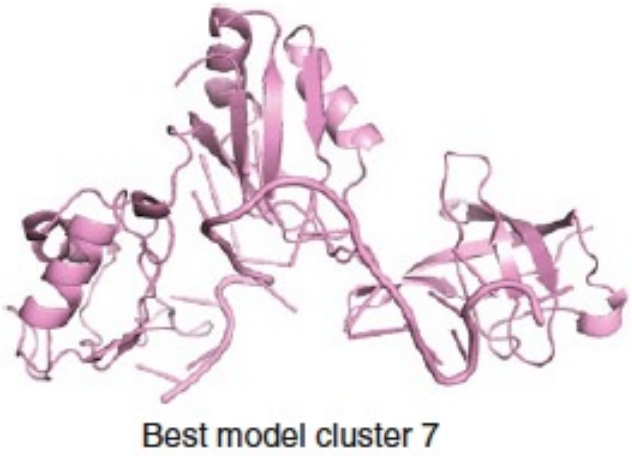


Hennig J, Wang I, Sonntag M, Gabel F, Sattler M. J Biomol NMR. 2013;56(1):17-30.

Example: a ternary protein-protein-RNA





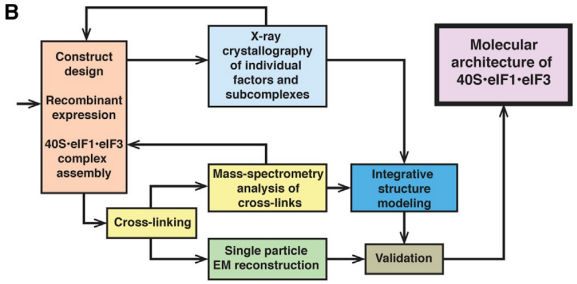


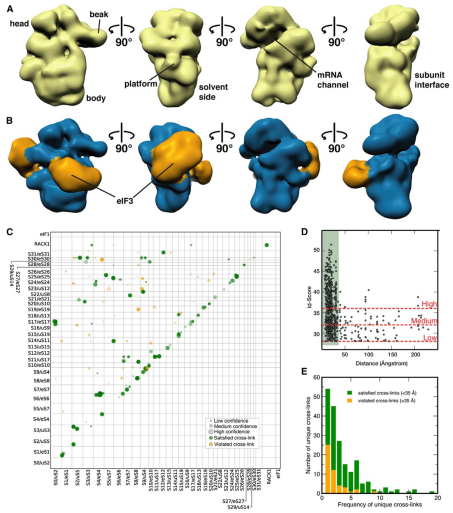
Hennig J, Wang I, Sonntag M, Gabel F, Sattler M. J Biomol NMR. 2013;56(1):17-30.

IMP software (A. Sali)

http://integrativemodeling.org

Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A.

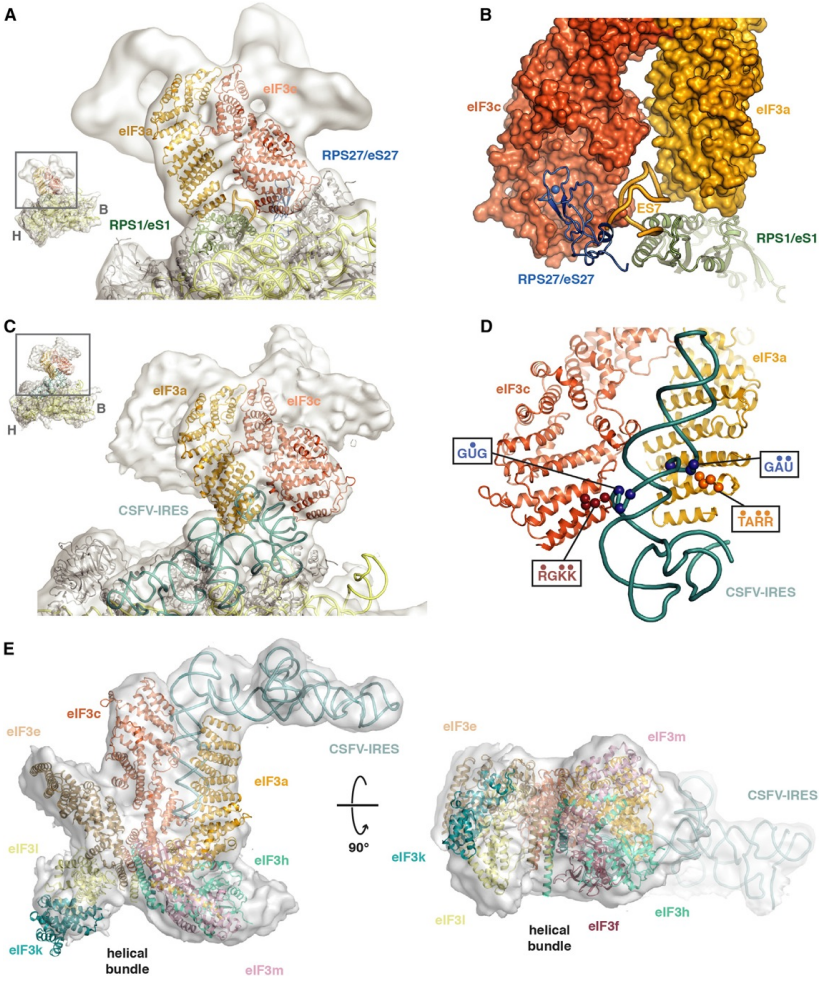




CryoEM

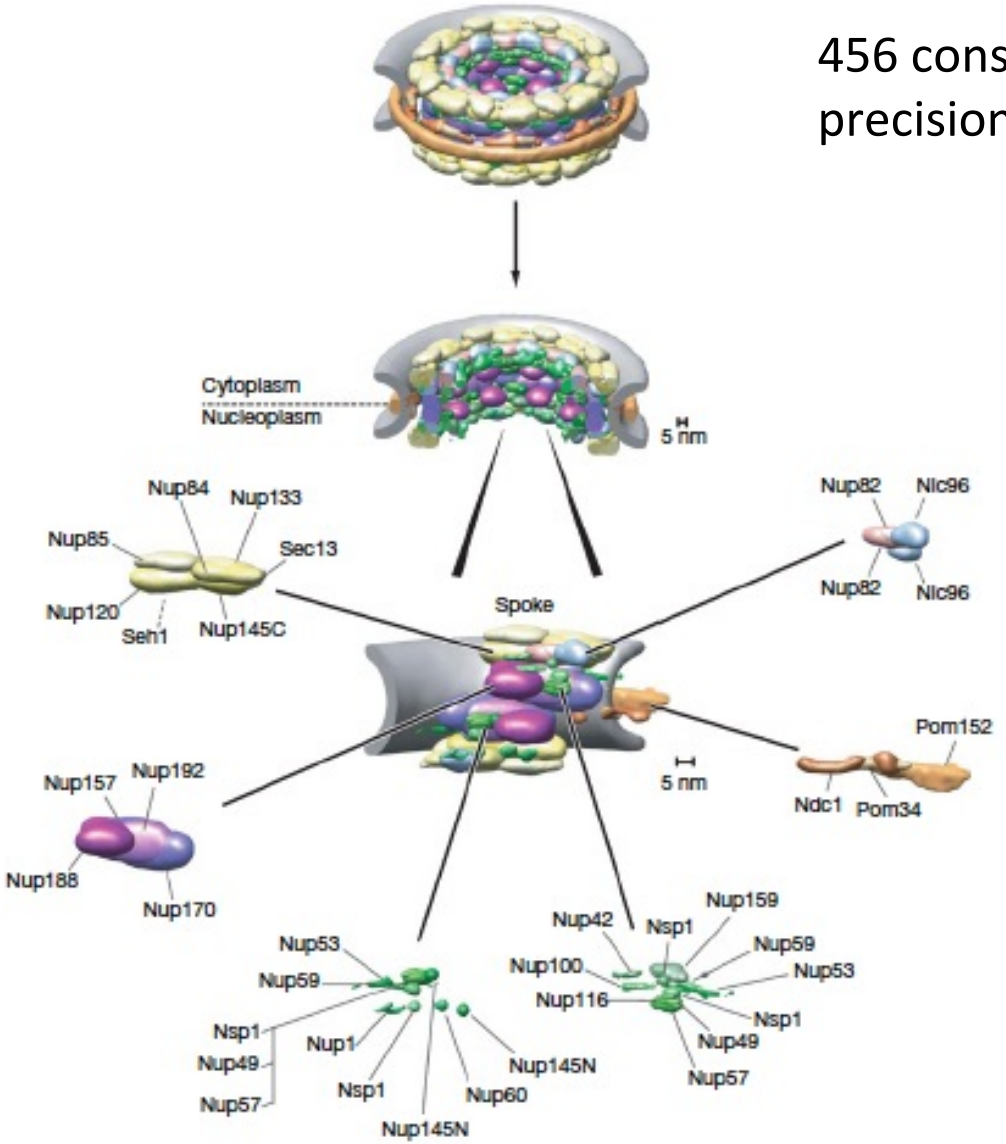
Crosslink

Molecular architecture of the 40S · eIF1 · eIF3 translation initiation complex.



Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CH, Cimermančič P, Boehringer D, Sali A, Aebersold R, Ban N. Cell. 2014;158(5):1123-35

Yeast Nuclear Pore Complex



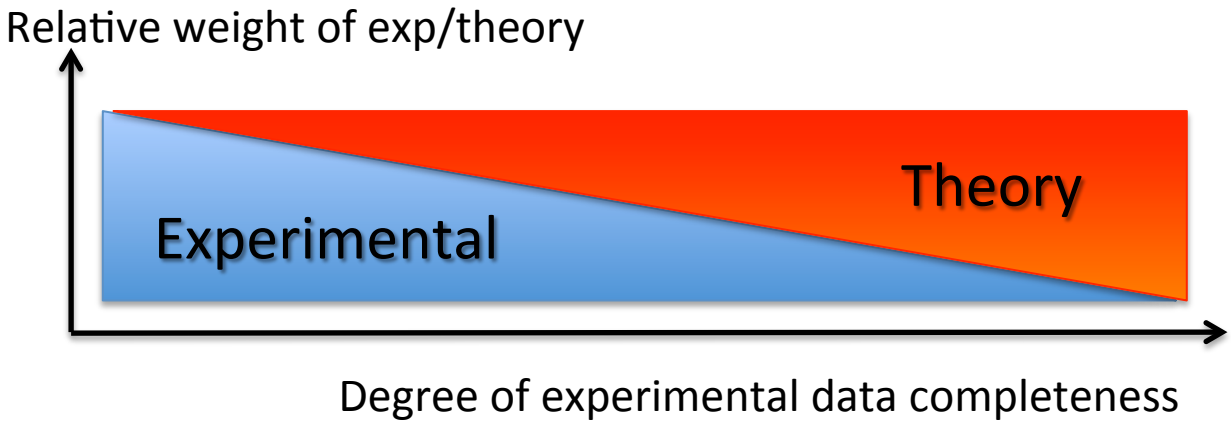
456 constituent proteins with an average precision of approximately 5 nm

Stoichiometry: protein quantification,
Protein proximities: subcomplex purification
Protein positions: immuno-EM
Subcomplex shapes: sedimentation analysis
Overall NPC shape: EM

Alber F, Dokudovskaya S, Veenhoff L, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait B, Rout M, and Sali A (2007) Determining the architectures of macromolecular assemblies, Nature 450, 683–694.

Take-home messages

(1) The final result of a technique depends on the choice of the (user-dependent) model.



(2) Proteins are flexible and heterogeneous. Deciding that they are represented by a simple model is a strong hypothesis.

(3) Do not forget that the final model is supposed to have the lowest energy

Keep a critical view on your result

Fin