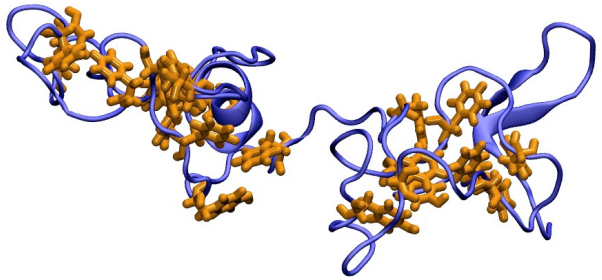
Tools for biophysical characterization of IDRs

Bio5469 (Washington University in St. Louis)
Sept 20th 2022
Alex Holehouse
alex.holehouse@wustl.edu

Recap from Mon.

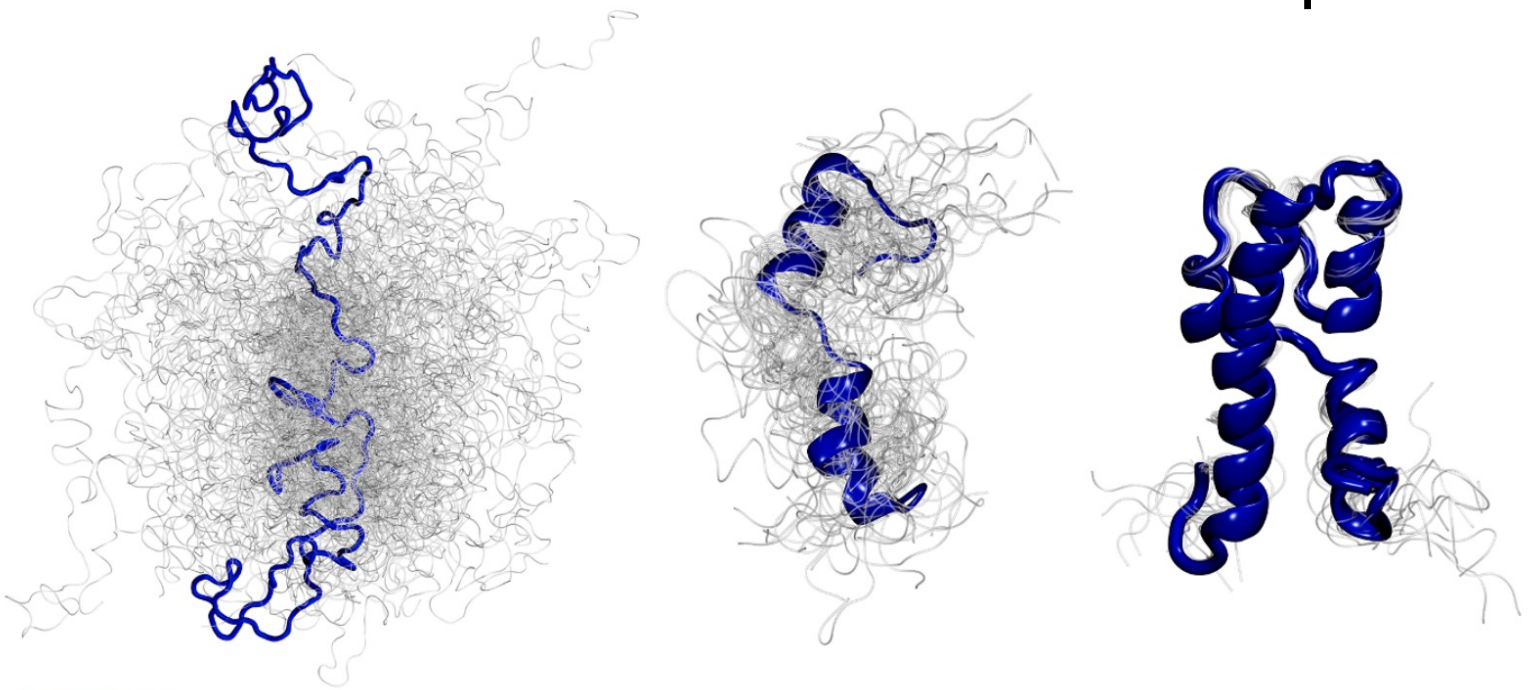
Recap from Mon.

- (1) IDRs are defined by lacking a fixed 3D structure



“IDP”

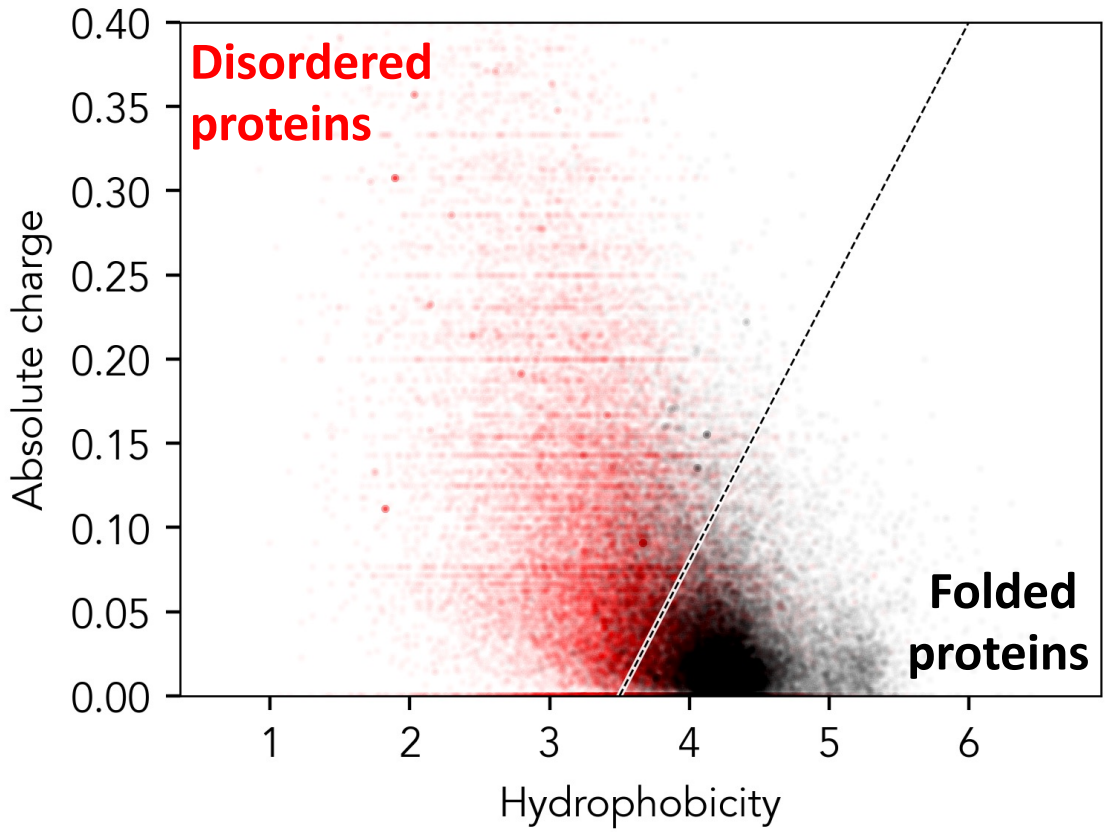
Folded protein



Structural heterogeneity

Recap from Mon.

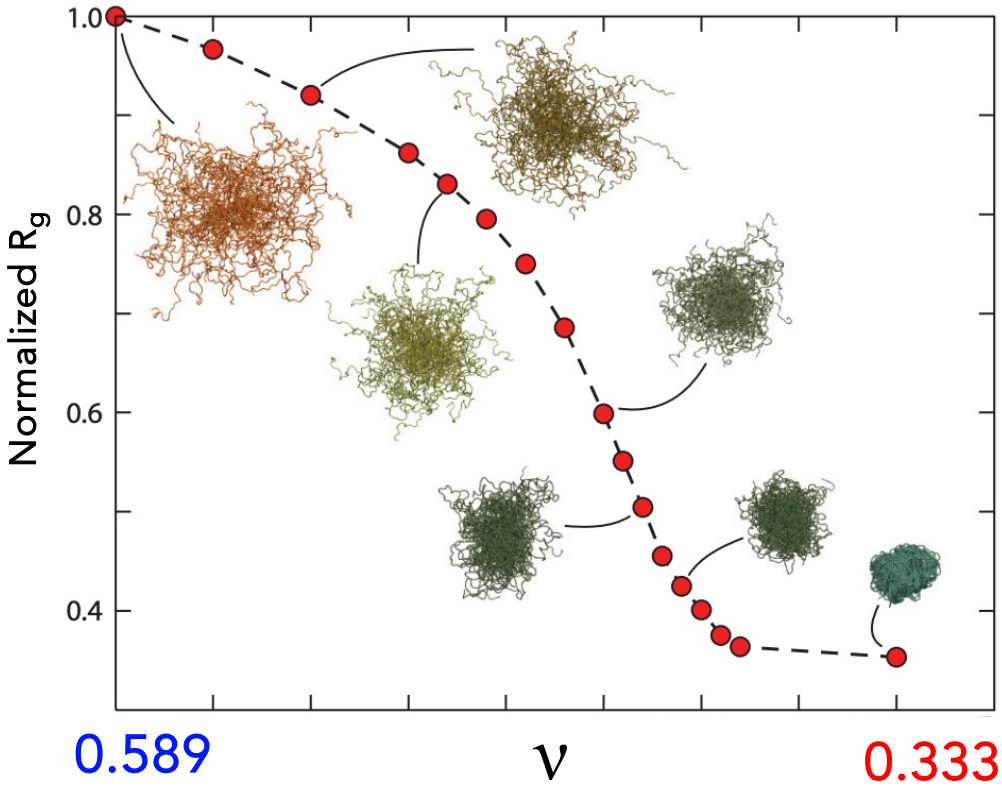
- (1) IDRs are defined by lacking a fixed 3D structure
- **(2) IDRs are depleted in hydrophobic residues and enriched in charged residues**


Disordered proteins
Folded proteins
Absolute charge
Hydrophobicity
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6

Recap from Mon.

- (1) IDRs are defined by lacking a fixed 3D structure
- (2) IDRs are depleted in hydrophobic residues and enriched in charged residues
- **(3) The lens of polymer physics offers a convenient reference frame to think about IDRs through**

Polymer scaling theory gives us tools to describe this



$$R_g = B_0 N^{\nu}$$

N = number of residues

B_0 = prefactor
(chain width/stiffness)

ν (nu) = scaling exponent

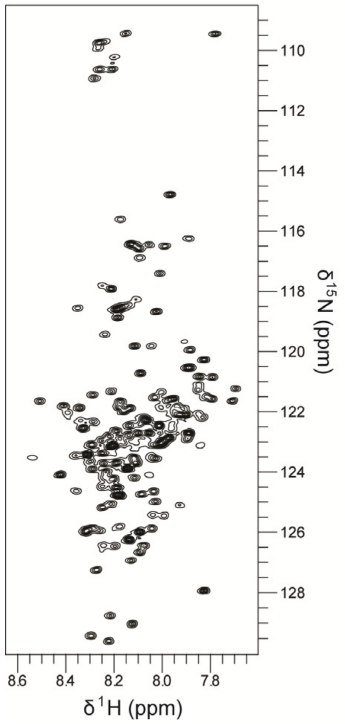
Today:

Approaches for the biophysical characterization of disordered proteins

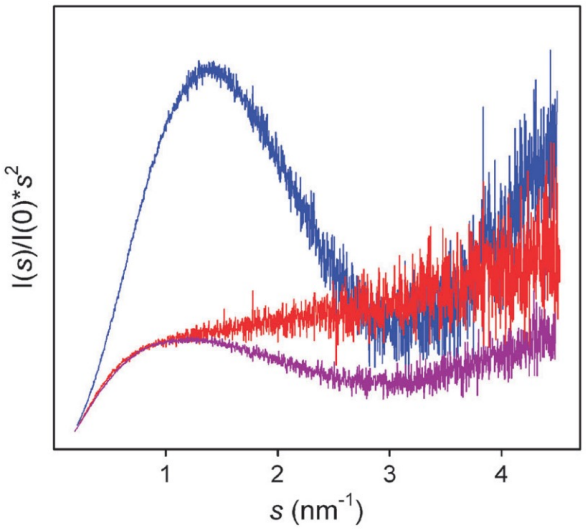
Part I : NMR, SAXS and smFRET

Three key experimental methods for characterizing IDRs

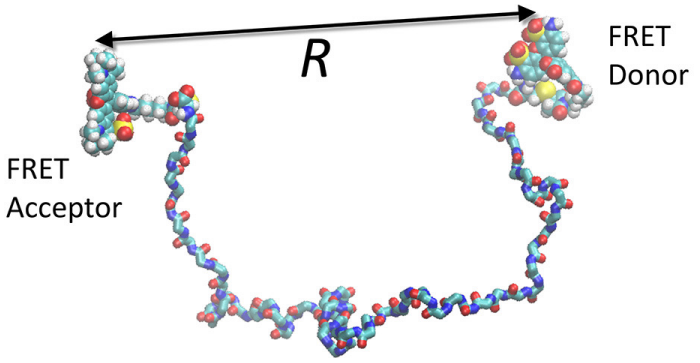
Nuclear Magnetic Resonance (NMR) spectroscopy)



Small

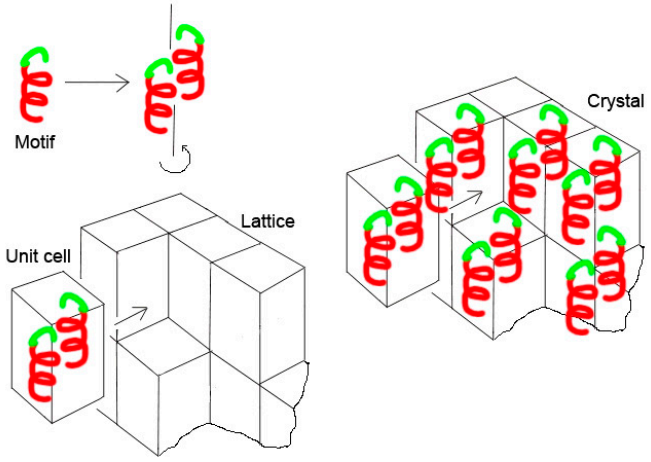


(Single molecule)
Förster Resonance Energy Transfer

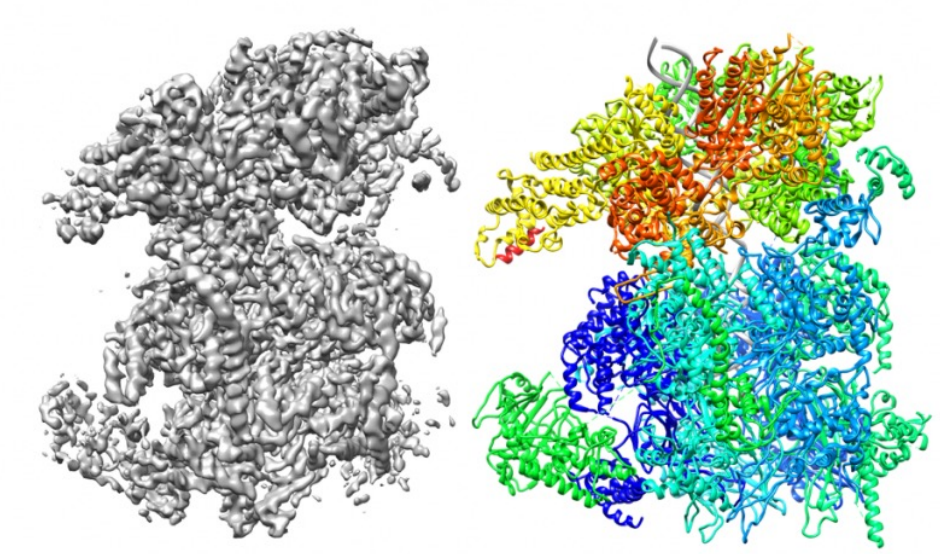


Conceptual challenge

You can study folded proteins as static entities



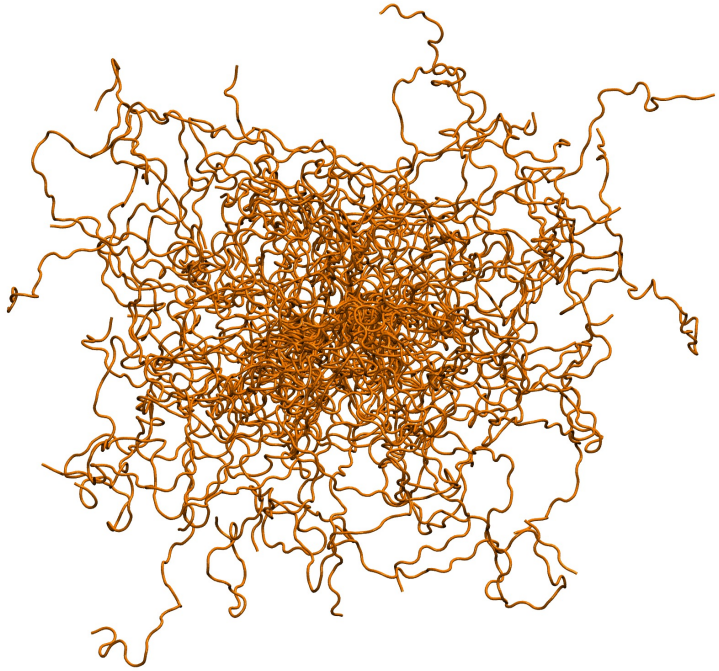
X-ray crystallography



CryoEM

Conceptual challenge

You cannot (currently) study disordered proteins as static entities



1. Nuclear Magnetic Resonance (NMR) spectroscopy for studying disordered proteins

NMR is a spectroscopic technique based on magnetization

NMR gives you local, residue-specific information

NMR

NMR gives you local, residue-specific information

- Treats each residue like a tiny bar magnet (by looking at certain atomic nuclei)

NMR

NMR gives you local, residue-specific information

- Treats each residue like a tiny bar magnet (by looking at certain atomic nuclei)
- Each bar magnet is spinning around!

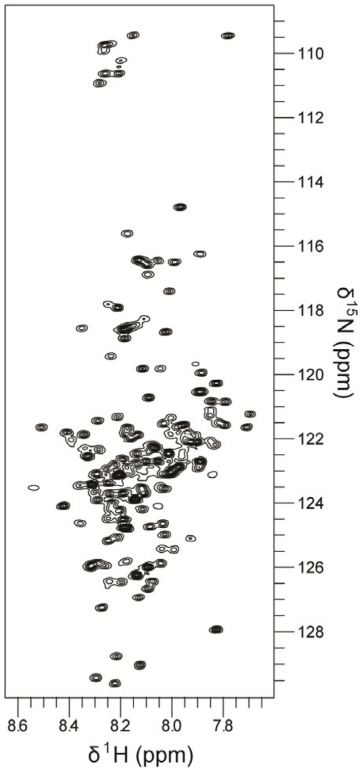
NMR

NMR gives you local, residue-specific information

- Treats each residue like a tiny bar magnet (by looking at certain atomic nuclei)
- Each bar magnet is spinning around!
- The rate that each bar magnet spins depends on the chemical environment it's in

NMR

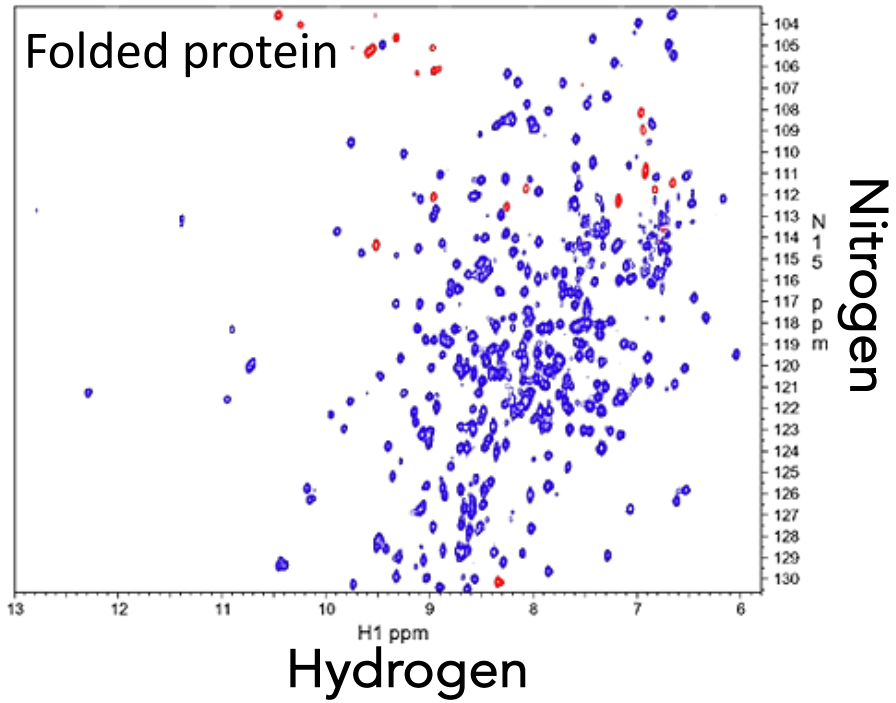
NMR gives you local, residue-specific information



- Treats each residue like a tiny bar magnet (by looking at certain atomic nuclei)
- Each bar magnet is spinning around!
- The rate that each bar magnet spins depends on the chemical environment it's in

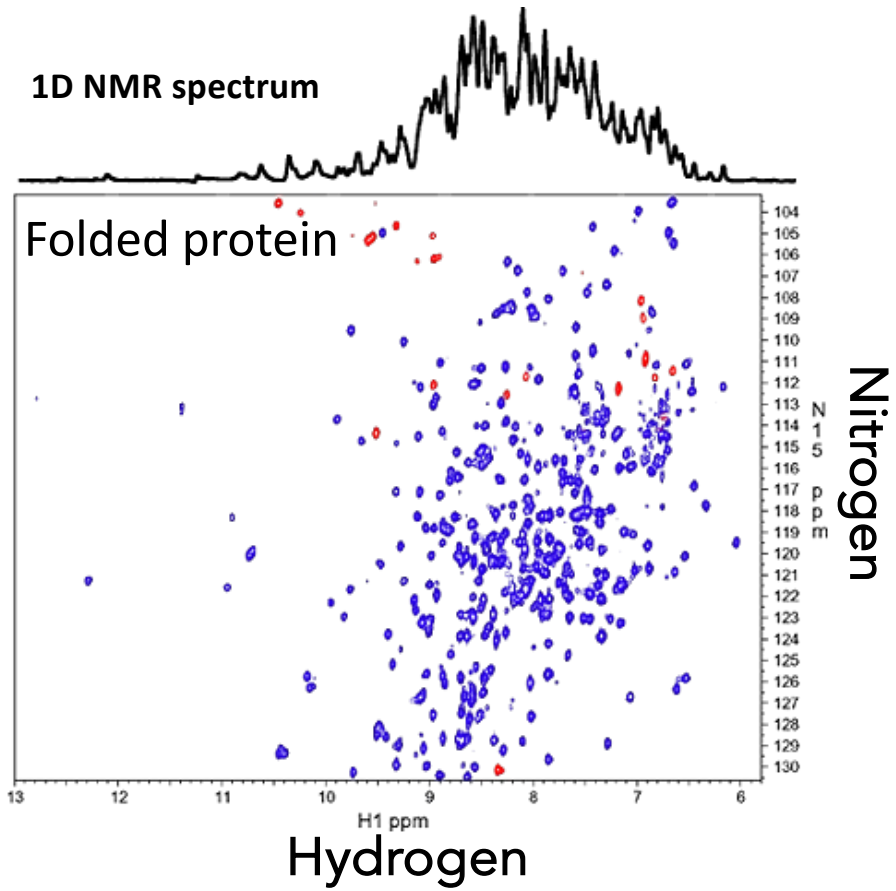
NMR

NMR gives you local, residue-specific information



NMR

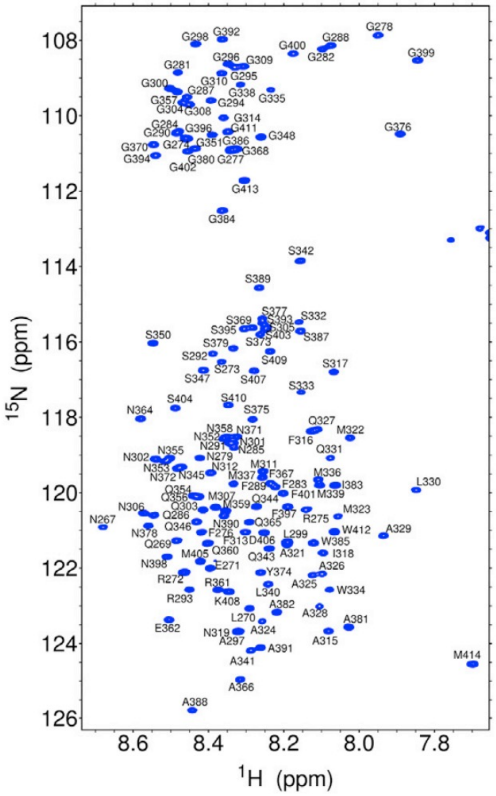
NMR gives you local, residue-specific information



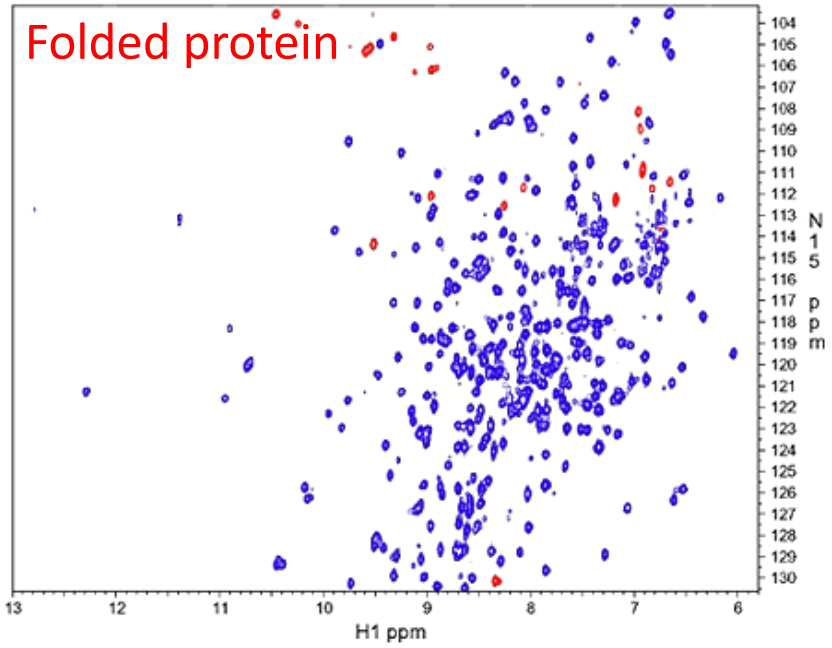
NMR

NMR gives you local, residue-specific information

Disordered protein



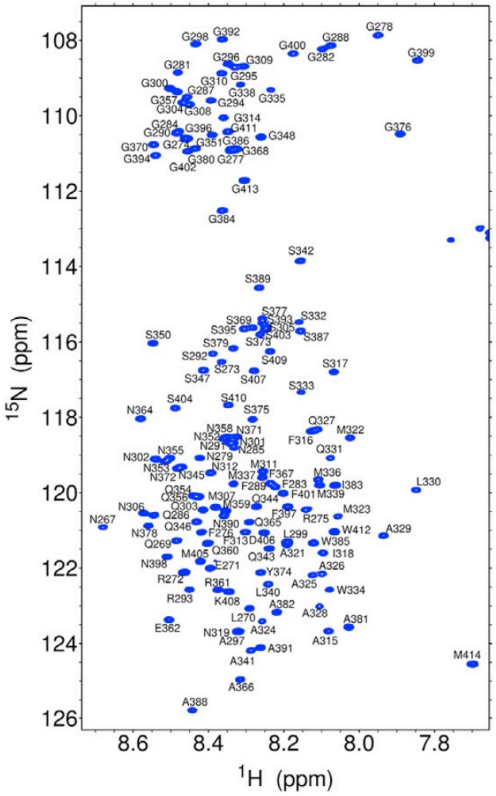
Folded protein



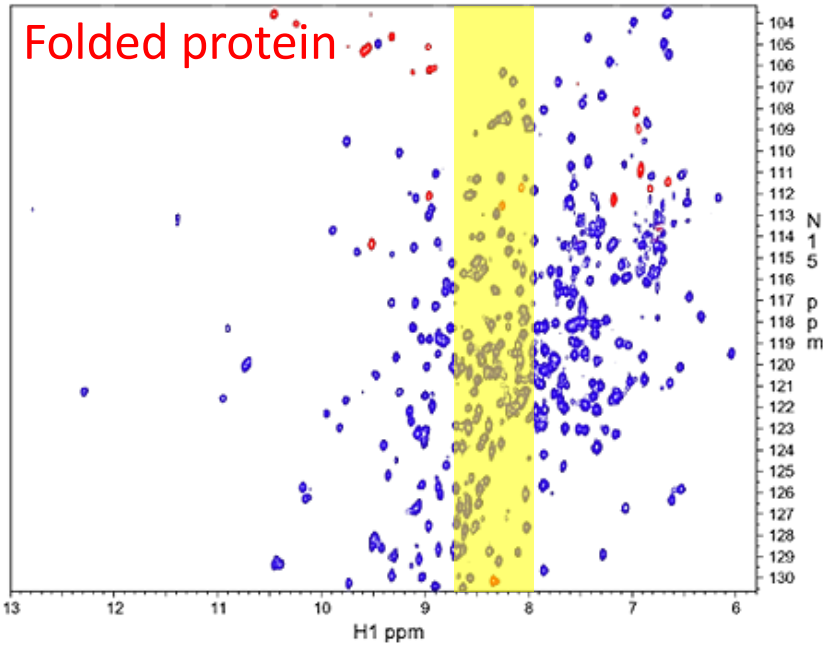
NMR

NMR gives you local, residue-specific information

Disordered protein



Folded protein

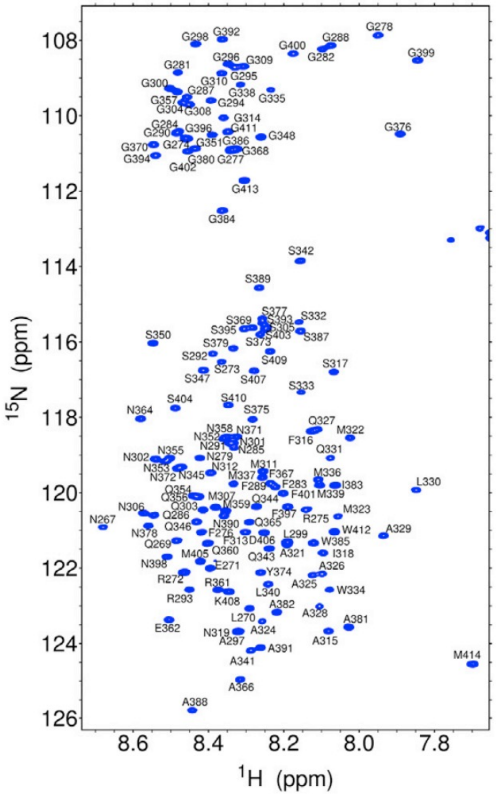


IDRs show “poor peak dispersion”

NMR

NMR gives us a way to see local residual structure in IDRs

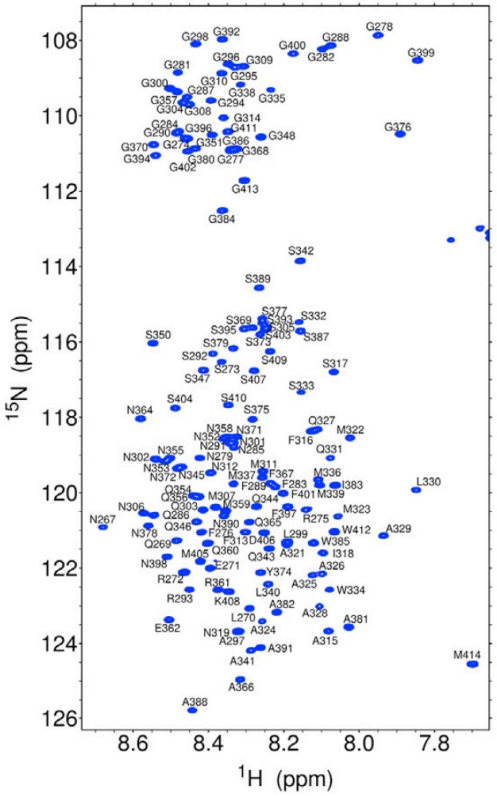
Disordered protein



NMR

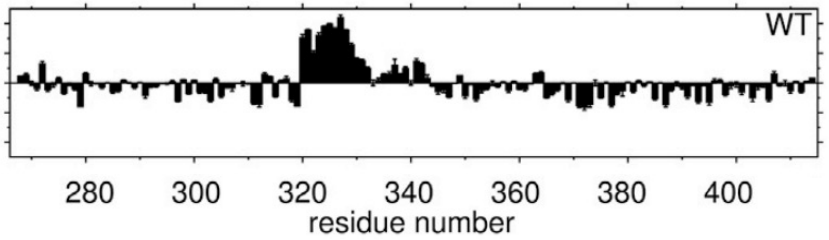
NMR gives us a way to see local residual structure in IDRs

Disordered protein



Difference compared to "**random coil**" chemical shifts

Chemical shift difference

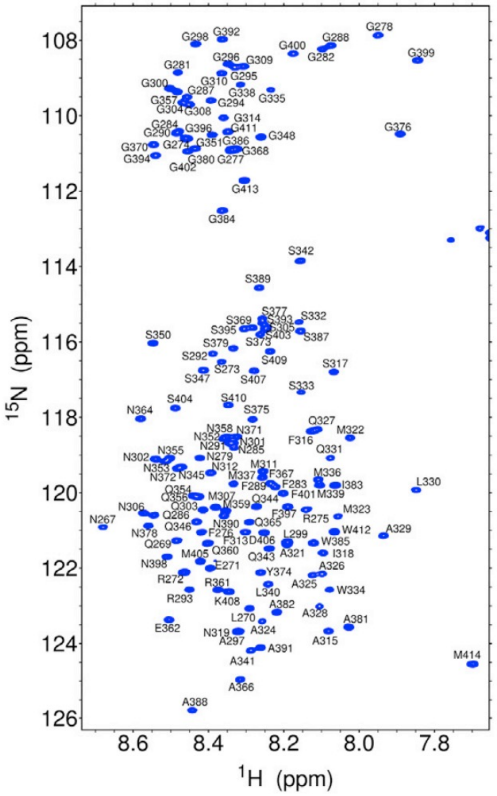


NMR

Random coil = short GGXGG peptides where X is a specific amino acid

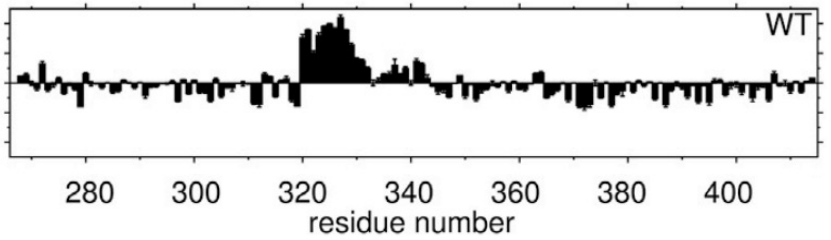
NMR gives us a way to see local residual structure in IDRs

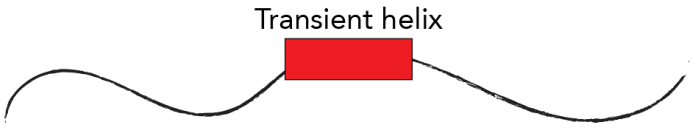
Disordered protein



Difference compared to “random coil” chemical shifts

Chemical shift difference



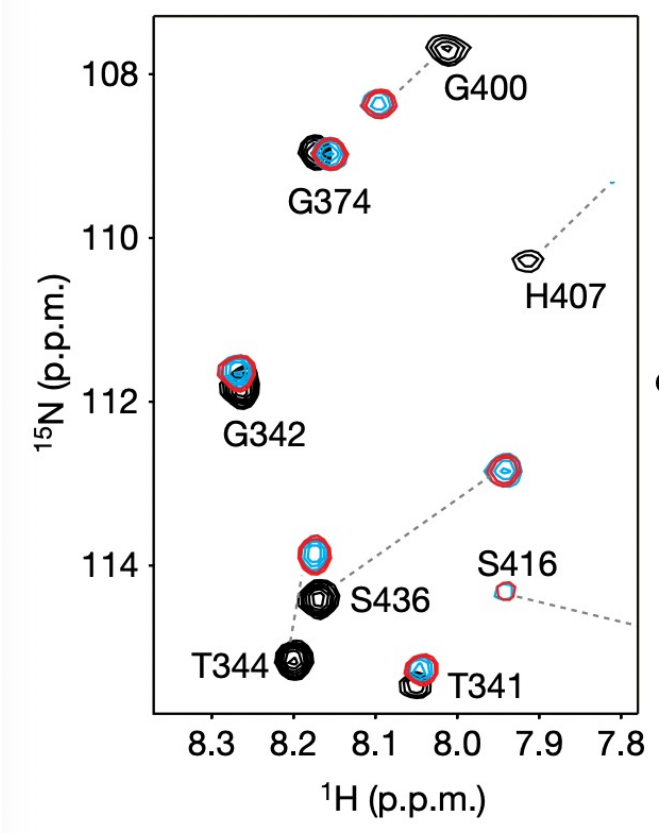


NMR

NMR gives us a way to see residue-specific binding

NMR

NMR gives us a way to see residue-specific binding



In isolation

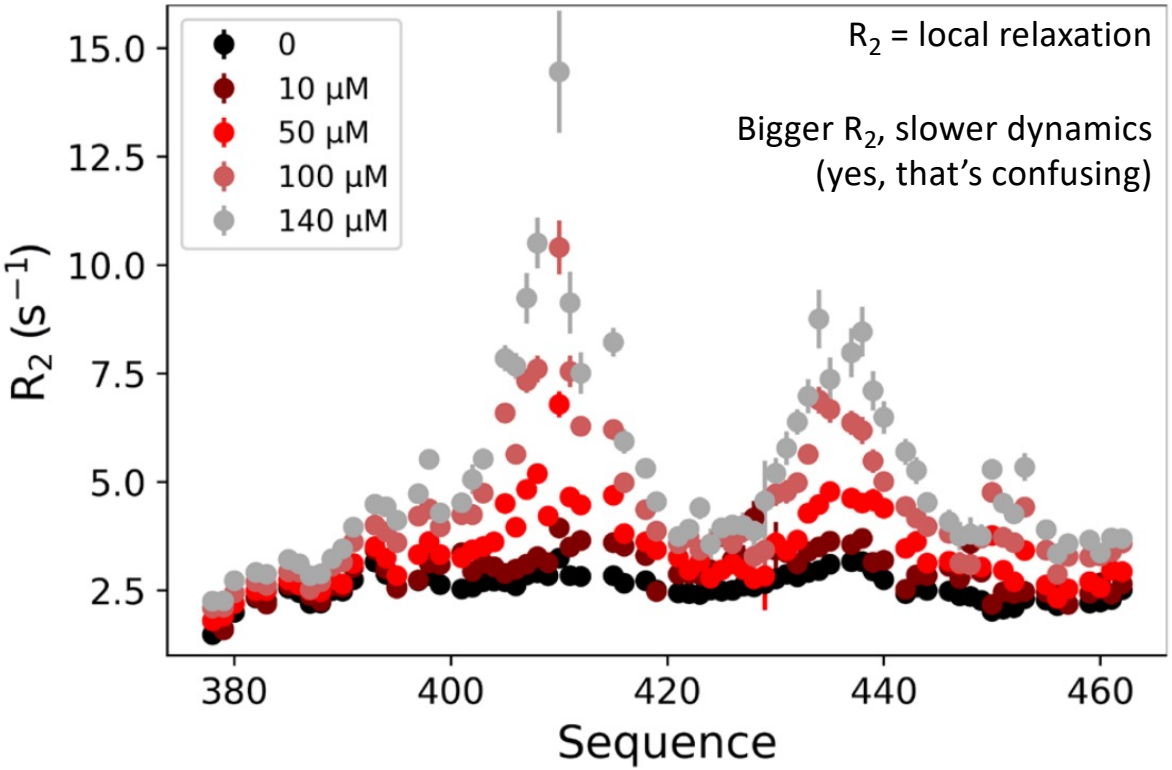
With partner

NMR

NMR also gives us a way to see local dynamics of IDRs

NMR

NMR also gives us a way to see local dynamics of IDRs



NMR

Benefits of NMR

- Provides residue specific information
- Incredibly versatile – can be used to track IDP binding, get local information from NOEs or chemical shifts or global information using pulsed-field gradient NMR
- Label free
- (Generally) interpretation of data is model free

NMR

Drawbacks of NMR

- Technically complicated – not something you can just ‘pick up’
- Instrumentation is expensive to buy and maintain
- Need high protein concentration (signal:noise is not good)
- Data can be hard to interpret

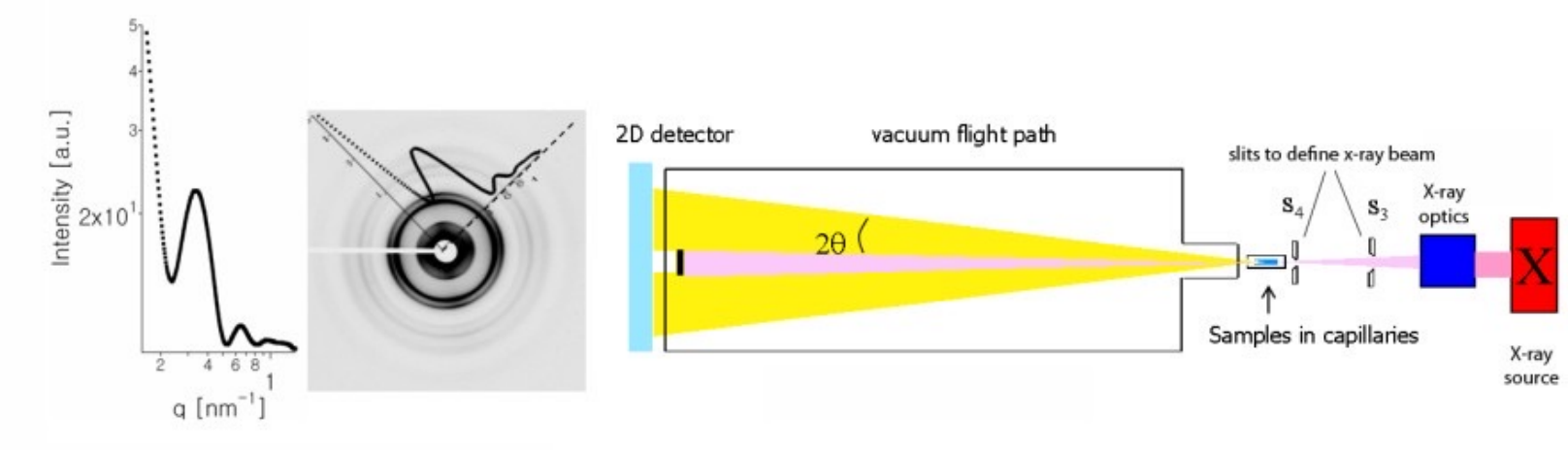
2. Small-angle X-ray scattering (SAXS) for studying disordered proteins

SAXS

SAXS is a solution scattering technique

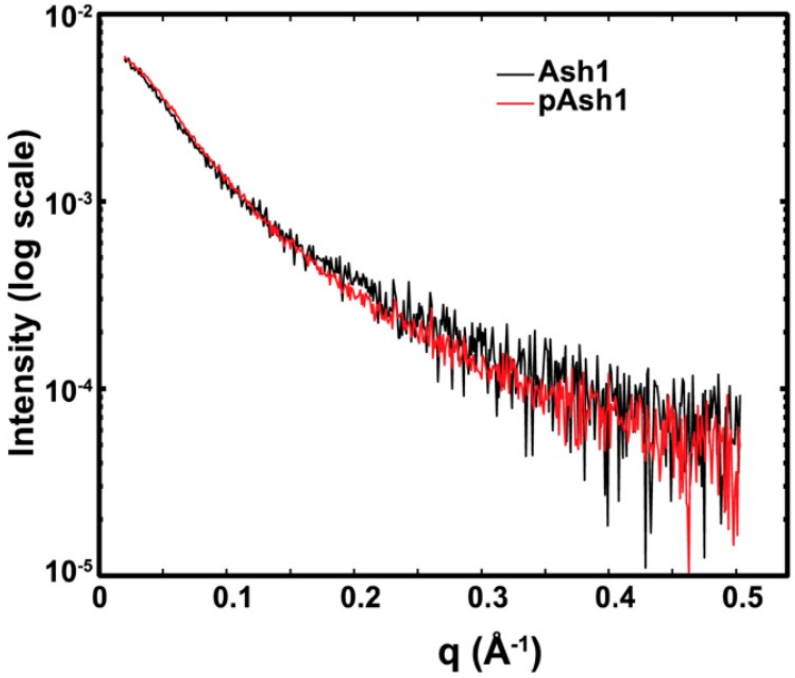
SAXS

SAXS is a solution scattering technique



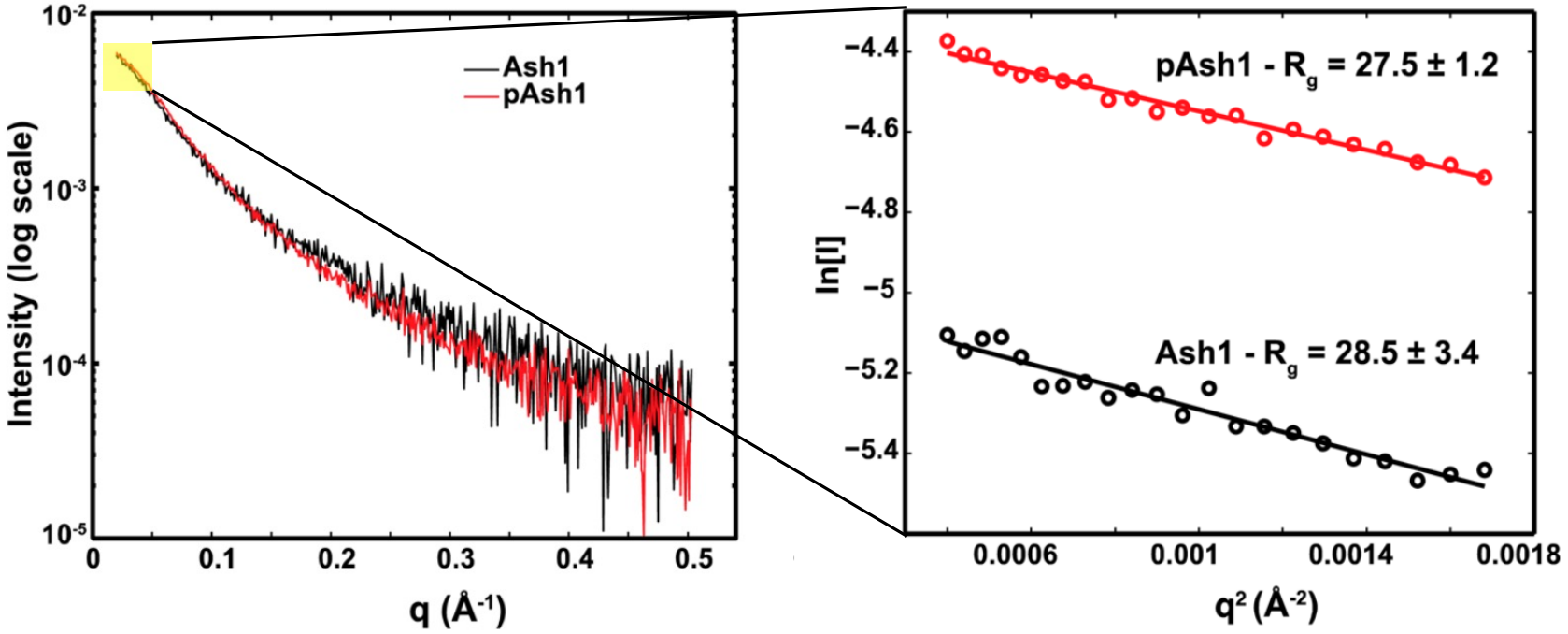
SAXS

SAXS is a solution scattering technique



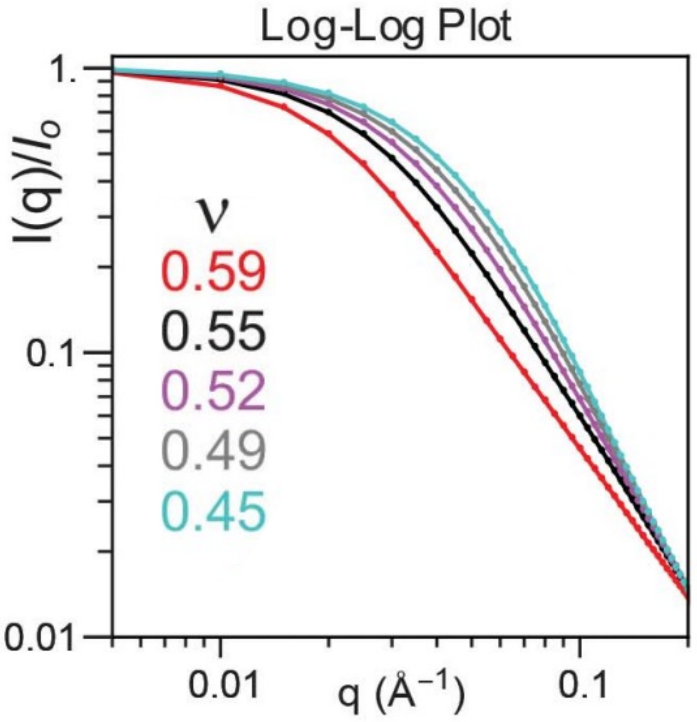
SAXS

SAXS measures global dimensions



SAXS

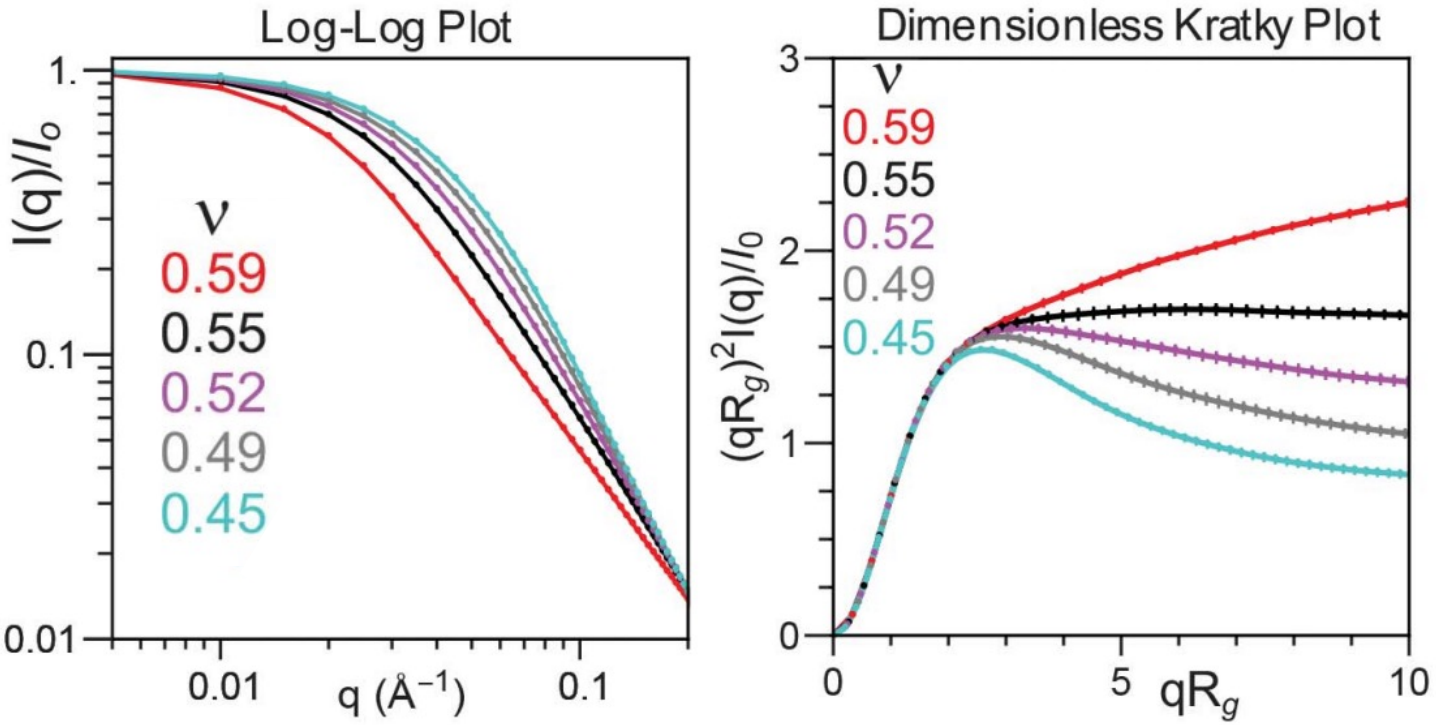
SAXS also provides information on polymer shape



Riback et al. ***Science*** (2017)

SAXS

SAXS also provides information on polymer shape



Riback et al. ***Science*** (2017)

SAXS

Benefits of SAXS

- Relatively easy to do (if you can get time on a beamline and protein behaves)
- Label free
- Model free (for radius of gyration)
- Can compare directly with simulations

SAXS

Drawbacks of SAXS

- Need high protein concentration (SEC-coupled SAXS helps)
- Is a low-resolution technique – can be misleading…
- Not particularly versatile in terms of what you can learn
- Mostly need a synchrotron (although can be done on a home source…)

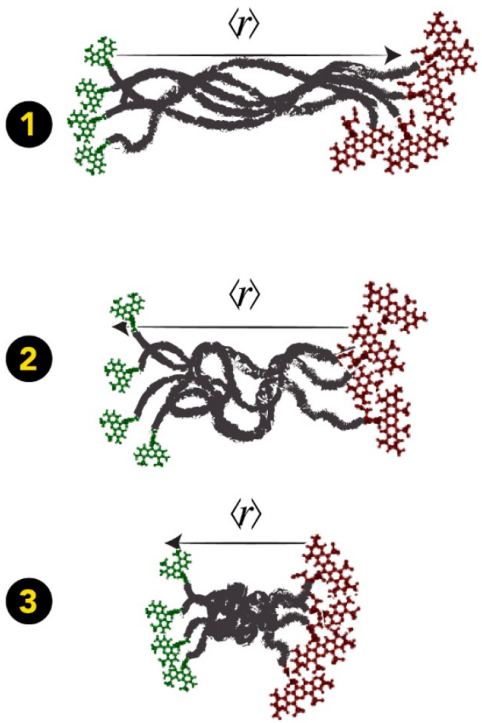
Synchrotrons are really expensive….

SAXS

3. Single-molecule Förster Resonance Energy Transfer (smFRET) for studying disordered proteins

smFRET

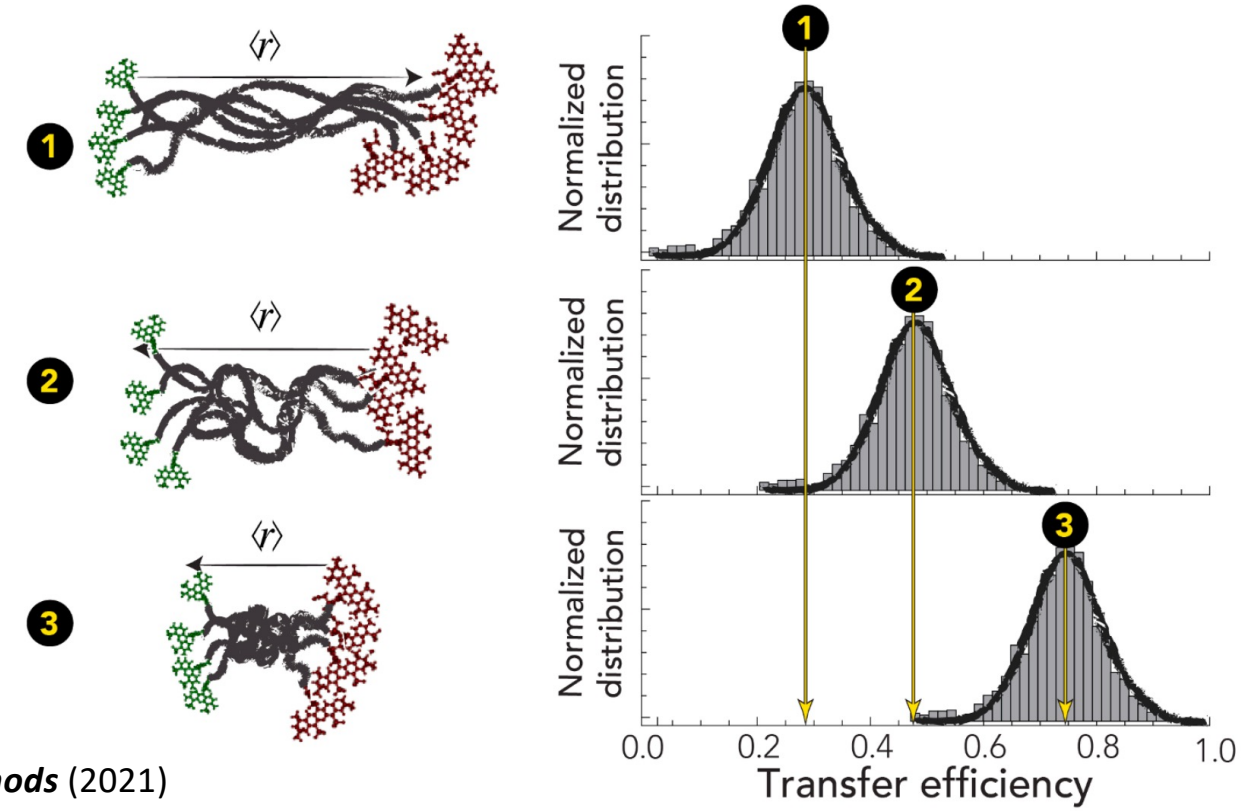
Single-molecule FRET measures residue-residue distances



Alston et al. ***Methods*** (2021)

smFRET

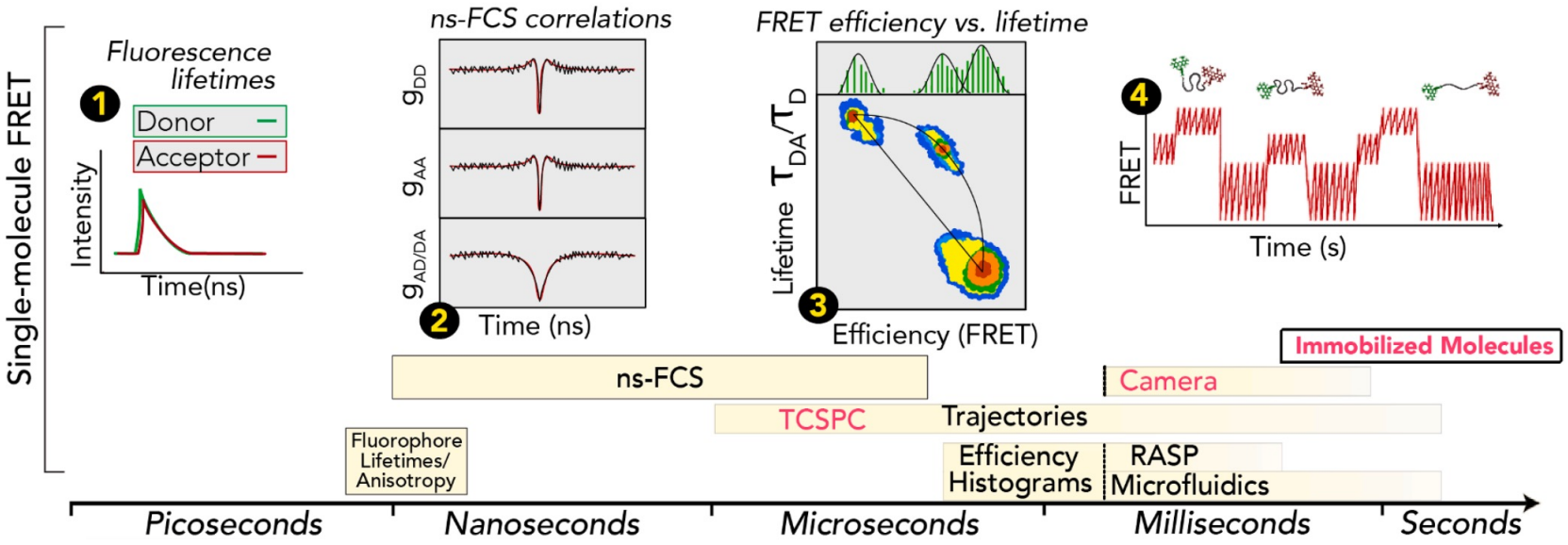
Single-molecule FRET measures residue-residue distances



Alston et al. ***Methods*** (2021)

smFRET

Single-molecule spectroscopy involves a wide range of tools



Alston et al. ***Methods*** (2021)

smFRET

Benefits of smFRET (or more broadly single molecule spectroscopy)

- Operates at extremely low protein concentrations
- Can simultaneously obtain information on protein conformations and and dynamics
- Can use both inter-molecular and intra-molecular smFRET to examine conformational changes or binding
- Can examine interaction between arbitrary types of biomolecules as long as labels can be attached

smFRET

Drawbacks of smFRET (or more broadly single molecule spectroscopy)

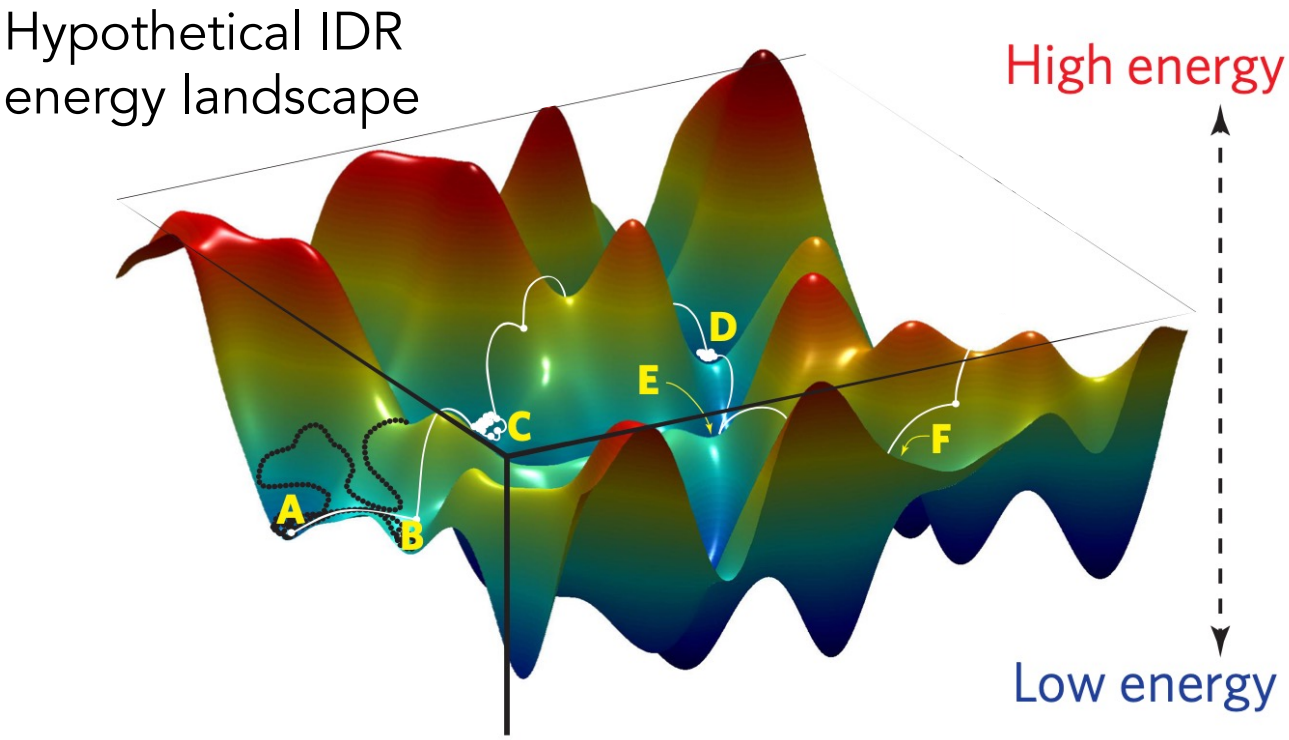
- Requires the addition of labels
- Interpretation often involves some kind of polymer model
- Global biases are inherently inferred
- Technically challenging

smFRET

4. Molecular simulations for studying disordered proteins

Simulations

Molecular simulations allow us to explore the an IDP's energy landscape



Simulations

Molecular simulations involve two components

Simulations

Molecular simulations involve two components

- **1. A way to represent the molecule(s) of interest:**
  - How do we represent our protein in the computer?
  - Sets of parameters that define the underlying physics of the system
  - Known as **a forcefield**

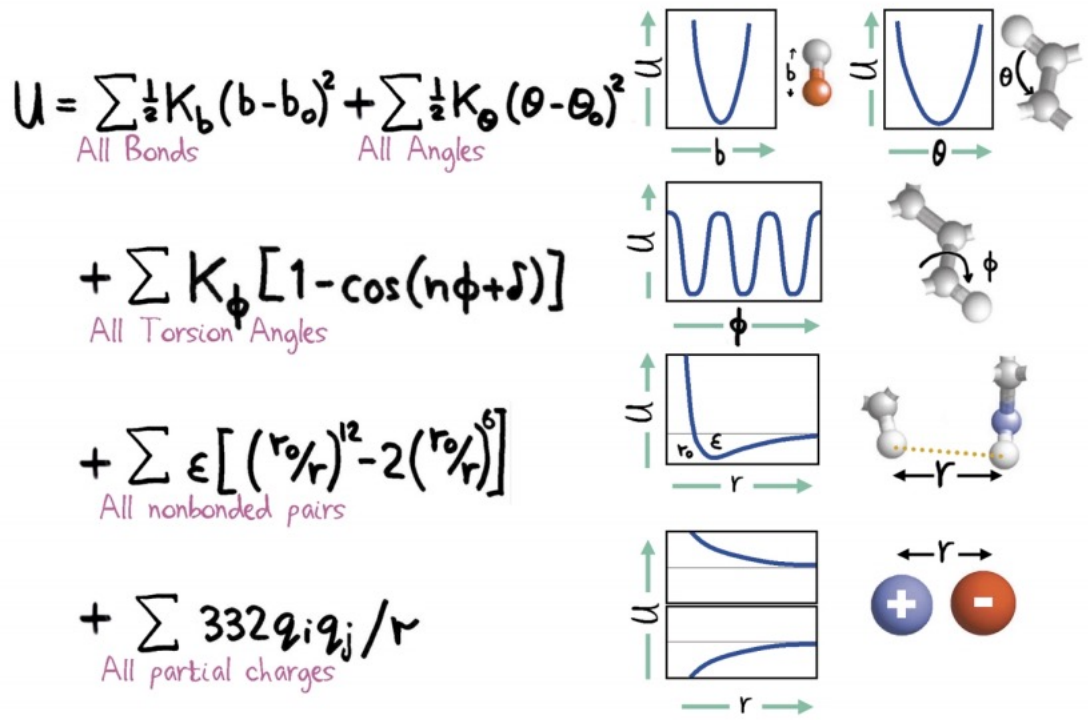
Molecular simulations involve two components

- **1. A way to represent the molecule(s) of interest:**
  - How do we represent our protein in the computer?
  - Sets of parameters that define the underlying physics of the system
  - Known as **a forcefield**

- **2. A way to update the configuration of the system:**
  - Biomolecules move
  - Need some way to 'sample' energetically relevant states

Simulations

1. Forcefields describe the underlying physics



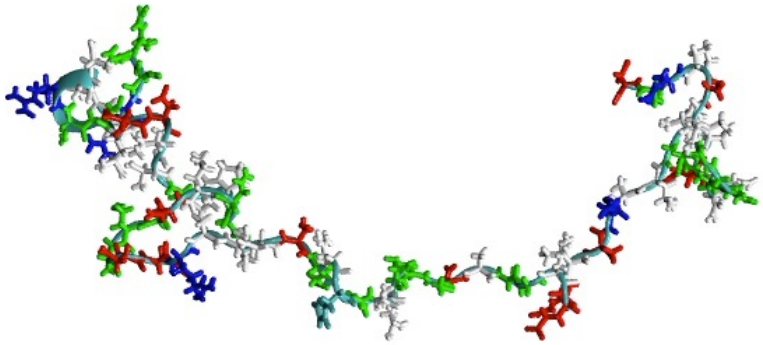
Simulations

Two main ways of updating the system configurations

Simulations

Approach 1: Molecular dynamics

Molecular dynamics (MD) generates dynamical trajectories evolving the system through time



Simulations

MD allows you to …

Simulations

MD allows you to …

- Compute ensembles using only sequence as input

Simulations

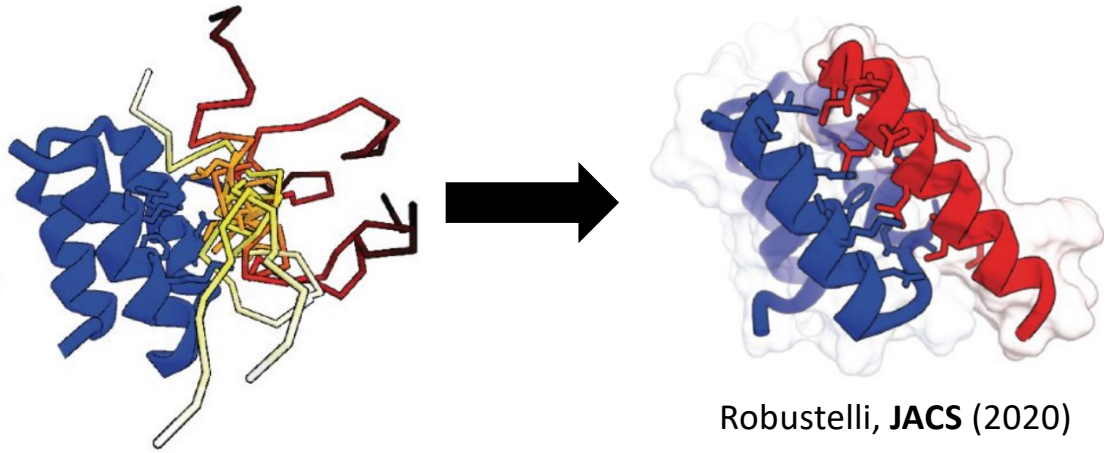
MD allows you to …

- Compute ensembles using only sequence as input
- Measure how quickly IDRs re-arrange

Simulations

MD allows you to …

- Compute ensembles using only sequence as input
- Measure how quickly IDRs re-arrange
- Assess binding of IDRs

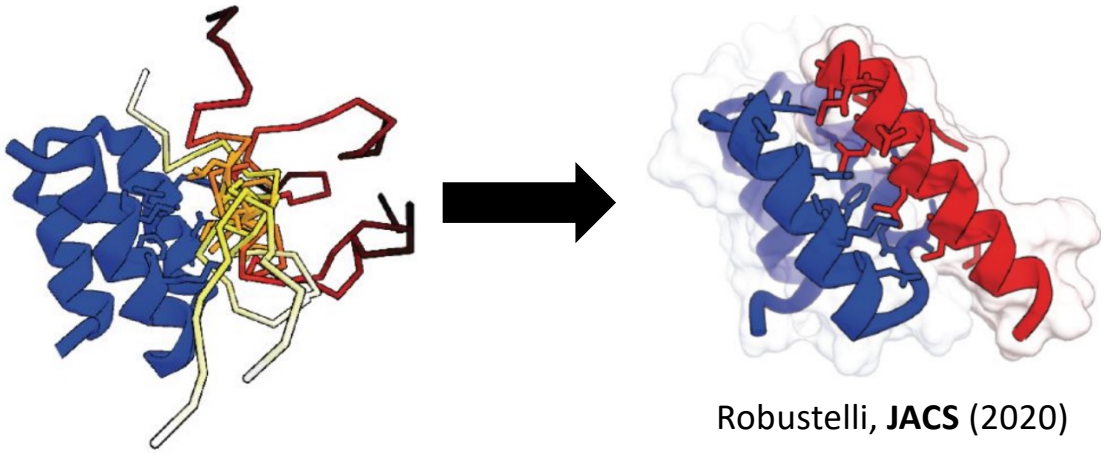


Robustelli, **JACS** (2020)

Simulations

MD allows you to …

- Compute ensembles using only sequence as input
- Measure how quickly IDRs re-arrange
- Assess binding of IDRs
- Much more…

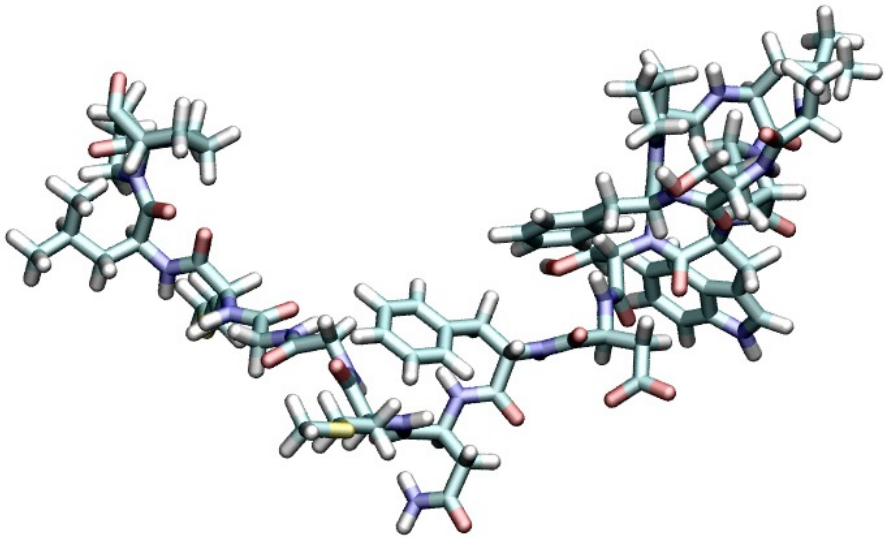


Robustelli, **JACS** (2020)

Simulations

Approach 2: Monte Carlo

Monte Carlo (MC) simulations sample configurational space without considering time



Simulations

Monte Carlo algorithm

Simulations

Monte Carlo algorithm

- 0. Calculate the current potential energy of the system

Monte Carlo algorithm

- 0. Calculate the current potential energy of the system
- 1. Randomly pick a degree of freedom

Monte Carlo algorithm

- 0. Calculate the current potential energy of the system

- 1. Randomly pick a degree of freedom

Location	Degree of freedom
Molecule	Rigid body coordinate (position and orientation)
Backbone	ω angle (CA_{i-1}, C_{i-1}, N_i, CA_i)
	ϕ angle (C_{i-1}, N_i, CA_i, C_i)
	ψ angle (N_i, CA_i, C_i, N_{i+1})
	Proline (has seven non-redundant degrees of freedom to facilitate puckering)
Sidechain	Depending on residue has ≥ 0 χ_1, χ_2, χ_3, χ_4 angles

Simulations

Monte Carlo algorithm

- 0. Calculate the current potential energy of the system

- 1. Randomly pick a degree of freedom
- 2. Change the degree of freedom by a random amount

Monte Carlo algorithm

- 0. Calculate the current potential energy of the system

- 1. Randomly pick a degree of freedom
- 2. Change the degree of freedom by a random amount
- 3. Calculate the NEW energy of the system

Monte Carlo algorithm

- 0. Calculate the current potential energy of the system

- 1. Randomly pick a degree of freedom
- 2. Change the degree of freedom by a random amount
- 3. Calculate the NEW energy of the system
- 4. Accept/reject the new configuration

Monte Carlo algorithm

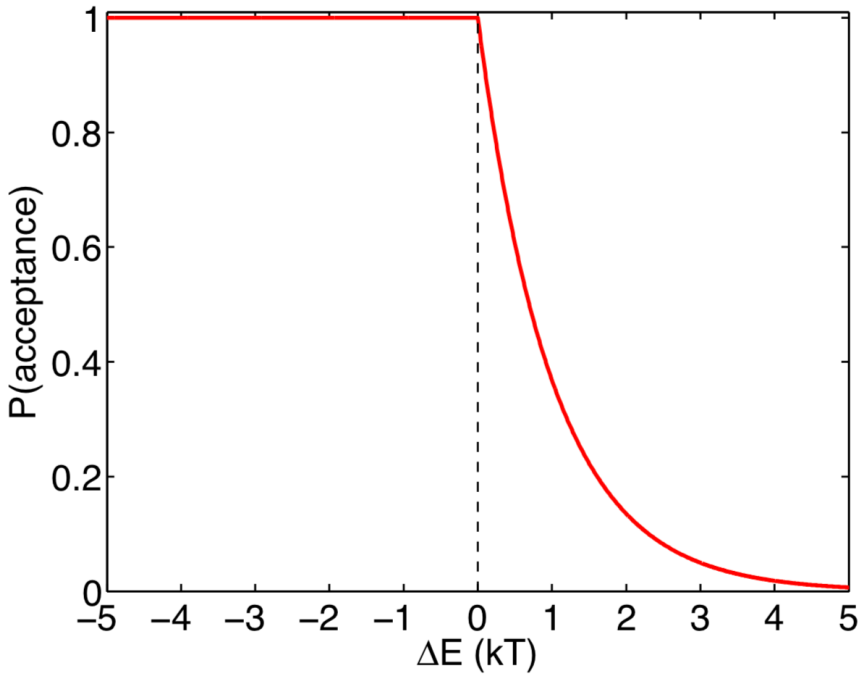
- 0. Calculate the current potential energy of the system

- 1. Randomly pick a degree of freedom
- 2. Change the degree of freedom by a random amount
- 3. Calculate the NEW energy of the system
- 4. Accept/reject the new configuration
- 5. Repeat steps 1-5 a billion times *(writing the current configuration to file at some interval)*

Simulations

We usually use the Metropolis-Hastings acceptance criterion to accept MC moves for simulations

$$p = \min\left\{1, \exp\left[-\left(\frac{1}{k_B T}\right) \times (E_B - E_A)\right]\right\}$$



E_B – potential energy of end state

E_A – potential energy of starting state

Metropolis-Hastings (1950s)

Simulations

MC allows you to …

Simulations

MC allows you to …

- Compute ensembles using only sequence as input

Simulations

MC allows you to …

- Compute ensembles using only sequence as input
- Assess binding of IDRs

Simulations

MC allows you to …

- Compute ensembles using only sequence as input
- Assess binding of IDRs
- Much more…

Simulations

Benefits of using simulations for studying IDPs

Simulations

Benefits of using simulations for studying IDPs

- Relatively cheap

Simulations

Benefits of using simulations for studying IDPs

- Relatively cheap
- Offers high-resolution predictive power

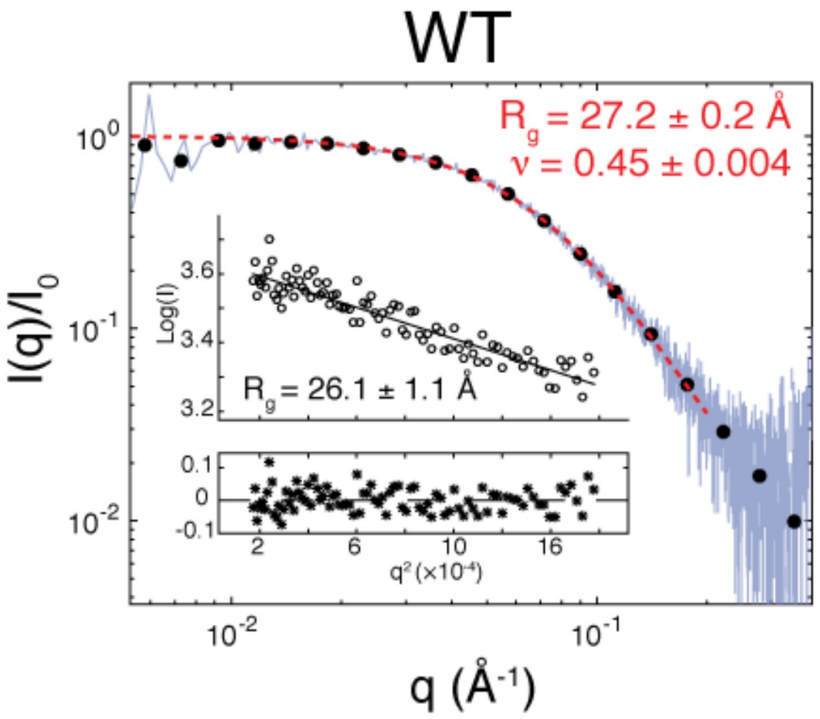
Simulations

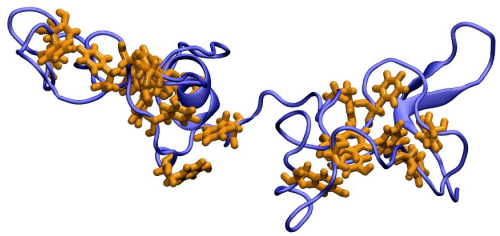
Benefits of using simulations for studying IDPs

- Relatively cheap
- Offers high-resolution predictive power
- Enables high-resolution interpretation of (most) experimental data

Simulations

Benefits of using simulations for studying IDPs





Simulations

Benefits of using simulations for studying IDPs

- Relatively cheap
- Offers high-resolution predictive power
- Enables high-resolution interpretation of (most) experimental data
- Have absolutely control and understanding of the underlying chemical physics

Simulations

Drawbacks of using simulations for studying IDPs

Simulations

Drawbacks of using simulations for studying IDPs

- Forcefields suffers from limitations and inaccuracies

Simulations

Drawbacks of using simulations for studying IDPs

- Forcefields suffers from limitations and inaccuracies
  - Forcefields are (typically) a computationally tractable implementation of a simplified version of our understanding of physical chemistry

Simulations

Drawbacks of using simulations for studying IDPs

- Forcefields suffers from limitations and inaccuracies
  - Forcefields are (typically) a computationally tractable implementation of a simplified version of our understanding of physical chemistry
    - Three layers of inaccuracies
    - Most ”standard” forcefields get a lot wrong (some exceptions: AMOEBA/HIPPO being beacons of rigor in an empirically parameterized world)

Simulations

Drawbacks of using simulations for studying IDPs

- Forcefields suffers from limitations and inaccuracies
  - Forcefields are (typically) a computationally tractable implementation of a simplified version of our understanding of physical chemistry
    - Three layers of inaccuracies
    - Most ”standard” forcefields get a lot wrong (some exceptions: AMOEBA/HIPPO being beacons of rigor in an empirically parameterized world)
  - Most modern forcefield were developed with folded proteins in mind

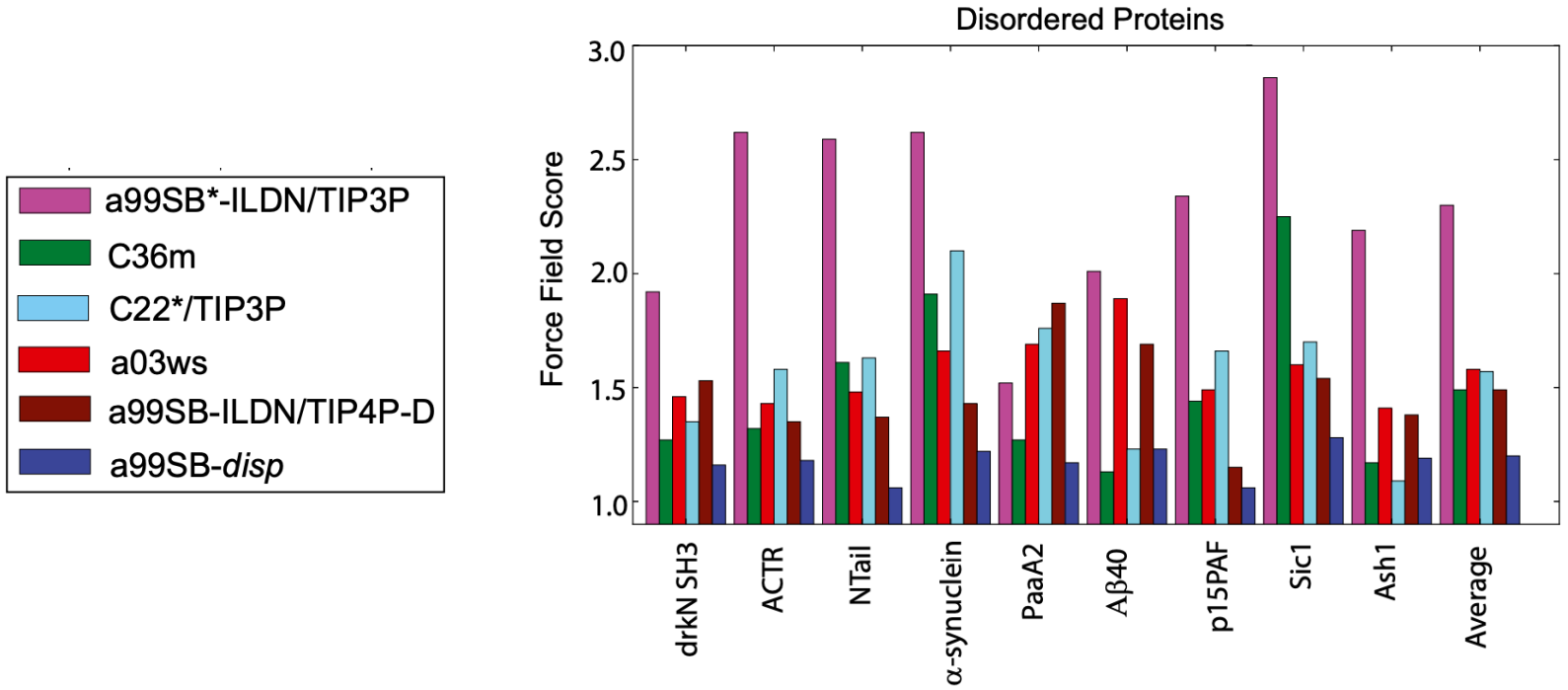
Simulations

Drawbacks of using simulations for studying IDPs

- Forcefields suffers from limitations and inaccuracies
  - Forcefields are (typically) a computationally tractable implementation of a simplified version of our understanding of physical chemistry
    - Three layers of inaccuracies
    - Most ”standard” forcefields get a lot wrong (some exceptions: AMOEBA/HIPPO being beacons of rigor in an empirically parameterized world)
  - Most modern forcefield were developed with folded proteins in mind
  - For IDPs, especially, forcefields have historically been error-prone

Simulations

Drawbacks of using simulations for studying IDPs



Robustelli et al. ***PNAS*** (2018)

Simulations

Drawbacks of using simulations for studying IDPs

- Sampling is hard!

Simulations

Drawbacks of using simulations for studying IDPs

- Sampling is hard!
  - This has historically been a major issue for MD
    - For a 100-residue IDP, re-arrangement takes ~60-100 ns
  - MC can sometimes circumvent this (but still issues)

Simulations

Next lecture (Friday)

Sequence-function relationships for IDRs