Part 1: Protein Dynamics

Folded protein at physiologic or room temperature samples wide range of conformations
- most in general neighborhood of average structure
- however, at any given moment an individual protein molecule is likely to differ significantly from average structure
- folded protein is an ensemble of conformations
- fluctuations about the average can be functionally important
  - catalytic or binding competent state may be higher energy excited state
  - conformational changes may be required for catalytic or mechanical fn
  - dynamics leads to allostery (consequence of re-distribution of conformational equilibria)
- Ex: myoglobin, RMS fluctuation in energy ~30 kcal/mol of total kinetic energy of 10^4 kcal/mol, in volume ~50 Å^3 relative to total volume of 10^5 Å^3

Sampling of unfolded and folding intermediate states is one major source of conformational fluctuation
- Protein folding stability is defined by the population of the unfolded state samples under equilibrium conditions
  \[ \Delta G_{\text{folding}} = \Delta G(\text{folded state}) - \Delta G(\text{unfolded state}) \]

Aside 1: What disordered states are relevant to understand protein folding?

compact denatured states called D₀ (Shortle) or unfolded state (Forman-Kay) present under non-denaturing conditions
always present under physiological conditions due to the Boltzmann distribution (for a stably folded protein varies from about 1 / 10^4 - 1 / 10^15)
to study unfolded state, many use strong denaturing conditions but not the relevant D₀ so better to examine unfolded states significantly populated under less denaturing conditions since structures will differ
presence of small amount of persistent structure (non-random interactions, deviations from random coil) in unfolded states of proteins may be marginal for thermodynamics but could have dramatic effects on the kinetics of folding.
structure could serve as the principle starting point of folding and control the kinetically accessible pathways but may not be relevant to the final structure in the folded state
Aside 2: How compact (on average) are these disordered states?

depends on the chemical nature of the polymer and solvent
"good" solvents --> highly expanded chain, solvent-chain interactions
preferred over chain-chain while "poor" solvents --> compact chain, aggregate, precipitate, exclude solvent-chain contacts
for protein, a given solvent can be good for one part, poor for another
water is generally a poor solvent for polypeptide
better solvent above melting and below cold denaturation temperature
sidechain interactions bias ensemble toward compact conformation
strong denaturing conditions yield extended chain with a radius of gyration close to value expected for a fully solvated polymer
using thermodynamic analyses, only ~2/3 of protein expected to be solvent exposed in a fully extended chain is actually exposed, even in extreme denaturing conditions - often overlooked in calculations of the total free energy change on folding based on solvent accessible area
when protein transferred from high denaturant (good solvent) to water (poor solvent), it collapses to a more compact state
collapse may precede or coincide with formation of folded structure -- if it precedes it, then both expanded and compact denatured states exist

Dynamics vs Thermodynamics

dynamics: directionality and time scales important
displacements from avg structure essential element

thermodynamics: presence of flexibility (entropy) important
equilibrium behavior significant
transiently populated species can be functional

Protein Motions within Folded State Ensemble

high energy costs of deformations of bond lengths, angles or dihedrals with any double bond character (peptide $\omega$ angle, aromatic rings) -> restricted
high frequency vibrations occur, but important motions involve relative displacements of small ~rigid groups by torsional motion about single bonds
single bonds permit torsional rotations for $\phi$, $\psi$ and $\chi$ angles of sidechains
irregular elastic deformations of the entire protein driven by solvent collisions chaotic librations of interior groups driven by collisions with neighboring atoms
most groups are tightly "caged" by atoms of protein or solvent
for $t \leq 10^{-12}$ s, see small amplitude ($\leq 0.2$ Å) "rattling" in cage
for larger $t$, see collective motions to change cage structure and relative displacement of neighboring groups (local) or relative displacement of different regions of the protein with only small local changes (rigid body)
Folded State Dynamics (cont.)

Wide range of amplitudes (0.01 - 100 Å), energies (0.1 - 100 kcal/mol) and time scales (10^-15 - 10^3 sec)

Increase in either amplitude, energy or time scale often corresponds to increase in other quantities

Not always correlated, however (local events with high activation energy, i.e. ring flips - E_{act} ~ 16 kcal/mol, slow macroscopic rate ~ 1/sec, fast individual event ~ 10^{-12} s)

NMR was instrumental in changing view of “static” protein structure

- Single resonances observed for δ1/δ2 and ε1/ε2 protons in Phe and Tyr rings in cores of proteins
  - Implies rapid exchange between the two sites due to 180° ring flips
  - Tight packing around rings necessitates large scale “breathing” motions on fast timescale (faster than global unfolding)

Biological Role of Motion

May be correlated, directional character to fluctuations in active sites of enzymes contributing to catalysis

Collective motions involved in displacements required for transitions from inactive to active configurations

Changes in fluctuations induced by perturbations (i.e. binding) may make significant contribution to free energy

Ligand binding and substrate entrance and product exit from enzymes requires motion to overcome steric barriers

Domain hinge motions depend on smaller concerted motions can be critical for enzymatic activity and binding

Large scale motions such as folding/unfolding transitions or ordering of partly ordered protein:

- Certain ligand binding proteins are highly flexible or partly/fully disordered in the absence of ligand perhaps as a means of rapid response and control and often as a means of activation (calmodulin, TIM, RNA/DNA binding proteins)

- Activation by cleavage of propeptide sequence or pH can lead to ordering of partly disordered structures or global structural rearrangements (trypsin, hemoglobin)

Allostery: re-distribution of conformational equilibria
Allostery revisited

Proteins have the ability to be allosterically modulated.

Allostery: defined as binding of ligand at one site in order to shift substrate binding affinities, alter enzymatic activity or regulate protein–protein interaction at a distant site.

Need to understand the mechanisms by which information is transmitted across long distances in proteins.

Molecular basis of allostery: given the ensemble nature of folded protein states, ligand induces shift in the population of states.

Requires methods that can investigate higher energy states of proteins that are only transiently visited and dynamic aspects of protein structure that may be influenced by the binding of allosteric ligands.


Statistical coupling analysis of PDZ family showed that residues important for peptide recognition co-varied with residues at a site distal to the binding pocket, suggesting a pathway for allosteric communication.


Computational and experimental methods identified similar pathways. NMR comparison of protein dynamics of a PDZ domain with and without bound peptide revealed two distal sites with ligand-induced changes in sidechain dynamics; one of these sites coincides with the site identified by Ranganathan. The fact that the change in dynamics occurs in the absence of significant structural change argues that, in some cases, dynamics alone can convey allosteric information.


Allostery Ex 1: PDZ domain - binds C-terminal peptide motif

Statistical coupling analysis of PDZ family showed that residues important for peptide recognition co-varied with residues at a site distal to the binding pocket, suggesting a pathway for allosteric communication.


Computational and experimental methods identified similar pathways. NMR comparison of protein dynamics of a PDZ domain with and without bound peptide revealed two distal sites with ligand-induced changes in sidechain dynamics; one of these sites coincides with the site identified by Ranganathan. The fact that the change in dynamics occurs in the absence of significant structural change argues that, in some cases, dynamics alone can convey allosteric information.

Allostery Ex 2: via modulation of folding-unfolding equilibrium

Wiskott–Aldrich syndrome protein (WASP) integrates multiple cellular inputs to regulate actin polymerization by the Arp2/3 complex. Cdc42 activates WASP by altering the equilibrium between a folded autoinhibited conformation of WASP (T state) and a largely unfolded R state that binds Arp2/3. Cdc42 GTP binds R state better than Cdc42 GDP and causes a greater shift to the active WASP conformation. The pre-existing allosteric equilibrium of WASP corresponds to a folding-unfolding equilibrium. WASP activity inside the cell results from integration of all influences on this equilibrium.


Experimental tools to probe dynamics

X-ray Diffraction

Debye-Waller (temperature factor) from refinement
B factor is directly related to mean-square atomic fluctuations in the isotropic harmonic model
can get harmonic, anisotropic B factors with simplified models and/or very high resolution data

\[ \mathbf{W}_j(Q) = -\frac{8}{3} \pi^2 \Delta r_j^2 \mathbf{s}^2 = -B_j \mathbf{s}^2 \]

\[ \mathbf{F}(Q) = \sum_{j=1}^{N} \mathbf{f}_j(Q)e^{iQ \cdot \mathbf{r}_j} \mathbf{W}_j(Q) \]

Q is the scattering vector, \( <r_j> \) is the average position of atom j, \( f_j(Q) \) is the atomic scattering factor and \( W_j(Q) \) is the Debye-Waller factor

\( B_j \) is usually referred to as the temperature factor, which is directly related to the mean-square atomic fluctuations in the isotropic harmonic model

some studies show correlation of B factors with direct motional probes (i.e. NMR relaxation parameters) while other studies have not

problems:
- static (lattice) disorder in crystal contributes to B
- real motion highly anisotropic, even anharmonic
- large amplitude motion can be underestimated due to refinement
- choosing one of multiple conformations for regions with high mobility
- highly mobile regions invisible
- crystal contacts can damp motion
- no time scale discrimination

NMR methods to probe dynamics

1st - brief intro to NMR:

- 2 basic NMR parameters:
  - chemical shift (frequency) - measure of different electronic environments of different nuclei in molecule, can be used as structural information
  - lineshape (relaxation, rate of loss of signal) - measure of dynamic properties
- Changes in either parameter upon addition of ligand provide evidence for binding
- Chemical shift changes in the presence and absence of molecule A can be used to define its binding site on molecule B
- Broadening often occurs upon binding due to “intermediate” exchange between free and bound state or conformational exchange in dynamic complex
- If exchange between free and bound state is “fast”, can determine \( K_d \) values

Protein NMR Spectra

Figure 2. Multidimensional NMR. (A) $^1$H NMR spectrum of p-nitrophenyl phosphate in 50 mM HEPES, pH = 7.5. The two resonances at approximately 7 and 8 ppm correspond to the two aromatic protons. The water peak resonates at 4.703 ppm and the remaining peaks below 4 ppm correspond to HEPES. (B) $^1$H NMR spectrum of phosphatase of regenerating liver-1 (PRL-1), a protein target implicated in cancer metastasis. The peaks above 6.0 ppm are the protein’s backbone amides and side chains. The water peak resonates at 4.703 ppm and the remaining signals represent the protein’s aliphatic protons. The signals are too numerous and overlapped to make any specific assignments. (C) $^1$H-N HSQC of PRL-1. Each peak represents a single NH group in the protein, including backbone and side-chain amides. Using three-dimensional correlation experiments, each peak can be assigned to a given amino acid in the protein sequence to yield structural information that can guide structure-based inhibitor design to the protein target.

Figure 3. Protein backbone. Each amino acid is connected via a peptide bond between the carbonyl carbon of the first amino acid and the nitrogen of the next amino acid, which is highlighted by the dashed square box. The $^1$H-$^15$N HSQC experiment detects protons directly coupled to nitrogen (designated by the circle), and the resulting spectrum contains one peak for every amino acid in the protein. The HSQC experiment (see Isotope Labeling Strategies under the Advance in NMR section) correlates the carbon atom of amino acid (i) to the amide proton of the preceding amino acid (i-1).

Range of chemical shifts in $^1$HN amide proton dimension is diagnostic of folded/disordered state:

- **Folded:** 6-10 ppm
- **Disordered:** 8.1-8.7 ppm
- More limited due to averaged electronic environment

NMR

Extremely powerful tool for site-specific data on protein dynamics

NMR spin relaxation experiments provide a measure of the amplitudes and frequencies of fast timescale dynamics (ns-ps).

NMR chemical exchange approaches provide quantitation of motion on intermediate timescale dynamics (ms-us).

Residual dipolar couplings (RDCs), paramagnetic relaxation enhancements (PREs), scalar couplings, and nuclear Overhauser enhancements (NOEs) provide data that can be interpreted in terms of dynamics. These approaches are relatively insensitive to the timescale of dynamics, requiring only that exchange among conformations is rapid enough to ensure that experimental measurements are averages across all well-populated states.

Extremely infrequent and transient conformational excursions can be quantified by measurements of $^1$H/$^2$H exchange.

Thus, NMR can be used to measure macromolecular dynamics on timescales ranging from picoseconds to hours or days.
NMR Chemical Exchange: 1D spectra for nuclei exchanging between 2 environments in 3:1 ratio, with chemical shifts $\omega_A$ and $\omega_B$ ($\omega_B - \omega_A = 100$ Hz). Times on the right-hand side correspond to $(k_{AB} + k_{BA})^{-1}$.

(a) When rate of exchange is much slower than the difference in chemical shift (in frequency units), separate peaks are observed, and EXSY can be used to determine the kinetics. Variations in chemical shift produced by exchange lead to broader signals.

(b) This effect is most pronounced when the exchange rate is roughly equal to difference in chemical shifts between exchanging probes in the two states. Exchange broadening resulting from processes with $(k_{AB} + k_{BA}) < \sim 2$ kHz can be analyzed by CPMG, whereas processes up to $\sim 20$ kHz can be studied by $R_1p$ relaxation dispersion (dots).

(c) When exchange is far more rapid than the difference in chemical shift, a single sharp peak is obtained at the population-weighted average chemical shift.

NMR (cont.)

Nuclear relaxation processes are sensitive to both magnitude and time scale of motions of individual atoms
- $T_1$, $T_2$, NOE relaxation parameters related to relative motions of nuclei with respect to each other and the laboratory frame
- Internal motions on sub-nanosecond time scale
- Internal motions on microsecond-millisecond time scale
- Overall rotational diffusion of molecule (5-50 nanoseconds)
- Rates of magnetization transfer among protons with different chemical shifts and proton exchange
- Motions on slow timescale of milliseconds to days
- Motional information at virtually every position, backbone and sidechain typically $^{15}$NH (backbone, Arg, Trp, Asn, Gln), $^{13}$C-carbonyl, $^{13}$C- or $^2$H-methyl
- Large range of timescales from picoseconds through days

Good reviews:
Analysis of NMR relaxation data

model-free analysis (no specific model for internal motion)/Lipari-Szabo (L-S) yields order parameter, $S^2$

measure of magnitude of angular fluctuations of a bond vector (i.e. NH)

reflects flexibility when the bond is rigid in the molecular frame, $S^2 = 1$
when the internal motion is isotropic, $S^2 = 0$

**fast backbone motion:**

- average amplitude of backbone NH ns-ps motion: $S^2 \sim 0.8-0.9$
- no correlation with secondary structure type
- tight turns can have restricted mobility as well
- loops and flexible N- and C-termini often have $S^2 < 0.7$
- for unfolded/disordered proteins, see range from very low to 0.7 in center of chain

**fast sidechain motion:**

- methyl $S^2$(axis) removes effect of fast rotation about C-C bond
- analysis of NH $S^2$ and methyl $S^2$(axis) values for a group of proteins (Mittermaier et al, J Biomol NMR 13, 181-185, 1999)
  - Ala $S^2$(axis) relating to motion of Ca-Cb bond vector is highly correlated with NH backbone $S^2$
  - $S^2$(axis) ranges from 0.1 to 0.9
- on average $S^2$(axis) gets smaller the longer the sidechain
does not correlate well with structural features such as packing and surface accessibility --> relationship between structure and dynamics not well understood
Effect of temperature on fast motion
(Yang et al, J Mol Biol 272, 790-804, 1997)

Dynamics of folded and unfolded states of staphylococcal nuclease and the N-terminal SH3 domain from drk studied at two different temperatures

Backbone $^{15}$N S$^2$ show significantly larger changes with temperature in the unfolded states of both proteins relative to corresponding folded forms

differential temperature dependence interpreted in terms of differences in heat capacities of folded and unfolded polypeptide chains

Amplitudes of side-chain motions in the folded SH3 domain are more sensitive to changes in temperature than backbone dynamics

Side-chain ps to ns time-scale motions likely contribute more to the heat capacity than backbone motions

Slow motion

Motion alters the environment of a nuclear spin and changes the chemical shift

Motion on microsec-millisec timescale leads to broadening of signal

Can use approaches for probing chemical exchange to quantify motion

Measurement of conformations within few kcal of ground state

Dynamics and entropy (site-specific contributions to thermodynamics)

Binding sites often flexible and unfolded states highly flexible

Motion may stabilize free state and does stabilize unfolded state

May be important for function (ex. allow substrate entry, product exit)

Flexibility changes upon binding/folding --- change in config. entropy

Use order parameters to estimate contribution of configurational entropy changes to the energetics of binding/folding

Formulas relating S$^2$ to configurational entropy derived from physical models (Yang & Kay, JMB 63, 369-382, 1996; Li et al, Protein Sci 12, 2647-2650, 1996; Akke et al, JACS 115, 9832-9833, 1993)

Binding applications show that changes in backbone (NH) and/or methyl flexibility make significant contribution to free energy of protein binding variety of targets (small hydrophobic inhibitor, peptide and DNA)

Suggests significant enthalpy/entropy compensation effects

Folding application estimates that about a quarter of total conformational entropy associated with folding of staphylococcal nuclease is reflected in NH order parameter changes based on comparison with MD simulations of folded and denatured protein (Wrabl et al, Proteins 38, 123-133, 2000)
Relaxation dispersions provide: thermodynamic, kinetic, and structural data

**Relaxation Dispersion:** the dependence of $R_{2,\text{eff}}$ on the parameters of a modulating sequence

CPMG dispersions

Analyze data in terms of:

- $p_i$ populations of exchanging states (0.5 – 10%)
- $k_{ij}$ exchange rate constants (ms range)
- $\Delta \omega_{ij}$ chemical shift differences

$$R_{2,\text{eff}} = f(v_{\text{CPMG}}, k_{\text{fwd}}, k_{\text{rev}}, \Delta \omega)$$

Structures of dynamically sampled low-populated states by relaxation dispersion NMR


Demonstrates that different mutations can give rise to intermediates with very different amounts of residual native-like structure and that some intermediates can have little residual structure while others have significant amounts.

Figure 3: Structural analysis of the I state of mutants G48M and G48V of Fyn SH3 using chemical shifts. **A:** Ratio of the chemical shift differences, $\Delta \omega_{ij} = \omega_{ij}(\text{mut}) - \omega_{ij}(\text{wt})$, plotted on the x,y plane of the structure of the wild type Fyn SH3 (PDB 1J3A) with the color code indicated. Residues with values of $\Delta \omega_{ij} < -0.5$ or $\Delta \omega_{ij} > 2.5$ are grey. **B:** Comparison of $\Delta \omega_{ij}$ (shaded) with the $\Delta \omega_{ij}$ values obtained in the structure determination (shaded grey). Only residues with more than 50% native contacts are shown. A ribbon representation of the ensemble of 25 structures determined for each mutant is shown on the right. The structure with the lowest root mean squared deviation (RMSD) of the backbone is shown in gray. Geometrically, the energy content map of the native state (blue) and the ensemble is elevated by the state of each mutant (red). $\Delta \omega$ is the interaction energy (in kcal/mol) between the side chains of residues (Calpha) according to the DSSP (DSSP) force field. Residue numbers are indicated along each of the axes.
Relaxation Dispersion experiments to study conformational equilibria around DHFR reaction pathway. Millisecond dynamics in DHFR changing in response to substrates and products (Wright et al.). (a) Catalytic cycle of DHFR. (b–f) Locations of exchange broadening for DHFR in complex with (b) NADPH, (c) NADP+–folate, (d) NADP+–THF, (e) THF, and (f) NADPH–THF. Green, red, and blue spheres correspond to 1H nuclei in the NADPH binding site, active site loop, and DHF binding site, respectively, for which CPMG relaxation dispersion data were fit to obtain kinetic and chemical shift parameters. Yellow spheres indicate 1H nuclei with broadened signals for which CPMG traces were of insufficient quality to extract exchange parameters.

Order parameter comparisons of protein motions
ps-ns and µs-ms time-scale motions in mesophilic and thermophilic ribonuclease HI. \( S^2 \) NH is color-coded (\( S^2 < 0.75 \), red, to \( S^2 > 0.95 \), blue) onto the backbone of (A) E. coli RNaseHI (PDB 2RN2) and (B) T. thermophilus RNaseHI (PDB 1RIL) at 310 K. Residues with significant conformational exchange are indicated with yellow balls. The mesophilic E.coli enzyme tends to show increased ps-ns motions, but the T. thermophilus enzyme shows greater µs-ms motions. There are also regions in the thermophilic enzyme that show increased ps-ns motion over that of the mesophilic enzyme.
**Hydrogen Exchange**

NMR can give quantitative data on rates from ms to days. Coherent neutron diffraction on crystals gives qualitative data. Can also use IR for more global exchange information. Mass spectrometry method of choice currently for Hexch analysis. Timescale relevant to biological function can be difficult to interpret results structurally.

**Comparison of hydrogen exchange of thermophilic/mesophilic proteins**

IR study of isopropyl malate dehydrogenase showed thermophilic protein exchanges much more slowly than mesophilic protein at same temperature but approximately the same at their respective optimum temperatures (Zavodszky, PNAS 95, 7406-11, 1998).

NMR hydrogen exchange rate measurements on rubredoxin from extreme thermophile show rapid exchange (Hernandez, PNAS 97, 3166-70, 2000). Activation energy values for many residues within 2-3 kcal/mol of that observed for unstructured peptides. Conformational flexibility of protein sufficient for water and base catalyst access to the exchanging amide with quite limited structural disruption does not support hypothesis that enhanced conformational rigidity in folded state underlies increased thermal stability of hyperthermophile proteins.

**Fluorescence Depolarization**

Aromatic residues and other probes used. Provides amplitudes and time scales of motions in ns-ps range. Parameters related to time correlation function whose decay is determined by reorientation of transition moment vectors associated with the probe.

Problems:
- Limited to ns-ps time scale motions.
- Limited probes, not adequate coverage of protein.

**Single Molecule Fluorescence Approaches**

New methods are very powerful probes of folding and translational/rotational molecular dynamics (in vitro and in cells) but smaller amplitude internal motions more difficult to probe.

Other spectroscopies (not well developed or highly specialized techniques)

Vibrational Spectroscopy
- vibrations in proteins span range from 3-3000 cm⁻¹
- motions with periods of 10 ps - 10 fs (very fast motions)
- probe with infrared absorption, resonant & nonresonant Raman scattering and inelastic neutron scattering

Electron Spin Relaxation
- electron spin relaxation in paramagnetic systems dominated by vibrational modes modulating crystal field where electron spin localized
- analyze in terms of vibrational modes

Mössbauer Spectroscopy
- resonant absorption of nuclear gamma rays
- data on time scales and amplitudes (1 - 100 ns time scale)
- often limited to Mössbauer nucleus (such as Fe)

Laser Spectroscopy
- photochemical dissociation and rebinding kinetics and folding kinetics

Part 2: Link between Dynamics and Protein Structure Determination

Goals of Structural Biology
- Atomic resolution structures that accurately represent average atomic positions
- Structural changes as a function of time and/or range of structures in physiological ensemble
- Comparison between isolated molecules and in higher order complexes
- Link between structural/dynamic/interaction data and function to understand biology at atomic level (including energetics, thermodynamics, kinetics)

What is an atomic structure
- (1) Set of x,y,z coordinates for each atom in a macromolecule
- (2) Ensemble of x,y,z coordinates for each atom representing all relevant/populated conformations
- Difference between (1) and (2) related to dynamics and “ensemble view” of folded protein
- Very early X-ray structures of proteins led to wide-spread view that proteins are rigid since they were tightly packed like small molecule crystals
- Later spectroscopic data demonstrated significant mobility and functional relevance of mobility
- Dynamic properties still not fully appreciated …
Dynamic Ensemble Refinement

Use experimental NMR data on ubiquitin
- NOE data
- Order parameters ($S^2$) for NH and methyl axis

Perform molecular dynamics simulation requiring that a set of conformations is *simultaneously* consistent with NOE data and $S^2$ restraints (e.g. set of 128 structures)

Obtain ensemble representing both structure and dynamical variability (DER ensemble)

Other available data for validation
- Residual dipolar couplings (RDCs)
- Sidechain scalar couplings ($^3J$)

Calculation of residual dipolar couplings (RDCs) (a) and side-chain scalar couplings (b). The NMR data were back-calculated from an ensemble of conformations that was determined using DER. For clarity, the magnitudes of the residual dipolar couplings were normalized to those for an amide NH in the same orientation by scaling according to bond lengths and gyromagnetic ratios. The data point labels in a describe the atoms between which the RDCs were measured. In b, $^3J_{\text{CCgamma}}$ and $^3J_{\text{Cgamma}}$ are scalar couplings between the side-chain gamma carbon and the backbone amide nitrogen and carbonyl carbon, respectively.

Correlation between exp and calc RDC data ($q$):
- 26% (DER)
- 41% (mean of 128 DER strs), 33% (min of DER)
- 24% (1.8 Å X-ray)
- 14% (NMR ensemble, with RDCs)
[Note that lower $q$ is better correlation]

Correlation between exp and calc $^3J$ data ($r^2$):
- 0.96 (DER)
- 0.81 (mean of DER strs), 0.91 (max of DER)
- 0.84 (X-ray)
- 0.89 (NMR ensemble)
Lindemann $\Delta$ value

- Greater than 0.15 is "liquid-like"
- 0.14 for heavy atoms of backbone (bb)
- 0.12 for core bb
- 0.17 for surface bb
- 0.29 for heavy atoms of sidechains (sc)
- 0.25 for core sc
- 0.30 for surface sc

"Surface molten" and "sidechain molten" even in core

Multiple Sidechain Rotomeric States

- Joint distribution of the $\chi_1$ dihedral angles in Ile 13 and Leu 15 in our 128 conformer ensemble (black), the crystal structure (red), the X-ray rapper ensemble (orange) and the published NMR ensemble (green).
- Four structures chosen from our DER ensemble to represent the four groupings of dihedrals evident in a; the four structures are arranged to match the four regions. Heavy atoms in the side chains of Ile 13 and Leu 15 are shown as van der Waals spheres (Ile 13 is located to the right of Leu 15).
- Distribution of side-chain $\chi_1$ and $\chi_2$ dihedral angles of selected hydrophobic residues. Colouring scheme as in a. In some of the plots the histograms of dihedral angles for the X-ray, NMR and rapper structures overlap.
Comparison of experimental and calculated $S^2$ values for a number of different ensembles. a, 128 conformer DER ensemble from this work using both NOEs and $S^2$ values as restraints ($q = 26\%$). b, X-ray raptor ensemble ($q = 24\%$). c, The published NMR ensemble ($q = 14\%$). d, 64 conformer ensemble from this work using NOEs enforced as restraints on a single conformer ($q = 23\%$). e, 128 conformer ensemble from this work using NOEs enforced as restraints on an ensemble of molecules ($q = 24\%$). f, Ensemble obtained from molecular dynamics simulations ($q = 52\%$).

**Structural Variability:**

**1. Lack of precision of structure determination**

Lack of data -> lack of precision
NMR

- Amount/accuracy of NOE, RDC, $^3$J, chemical shift data depends on spectral quality (MW, tumbling properties, oligomers, concentration, **motional properties**, etc)
- Lots of data -> very precise "average" structures
- Lack of data -> large variability about average
- Either case not representative of true dynamics
- Important to experimentally probe actual dynamic properties

X-ray/EM

- Resolution limits amount of data …
Structural Variability:
(2) Actual Dynamic Properties

- Current structure determination approaches limit information regarding dynamic properties
  - goal of most precisely defined, accurate estimate of average structure (lowest energy state)
  - related to desire to understand enzyme catalysis requiring fraction of Å positioning
- Temperature
  - Lower than physiologic inhibits motion
  - Cryo-temperatures in X-ray/EM freeze motion
- State of sample
  - NMR: solution enables dynamic freedom
    - “typical” ensembles may portray some realistic dynamics
  - X-ray: crystal lattice inhibits motion

Representation of Variability

- NMR
  - Calculation and deposition (PDB) of multiple structures satisfying all experimental data
  - Superposition of structures -> RMSD describing variability (or difference distance matrices)
  - Graphical representation by thickness of backbone “ribbon” or “worm” diagram or isosurfaces
- X-ray (B-factor, limited # of conformations, …)
- Structural biology field requires better ways to represent ensembles and dynamics
Representations of Structures/Variability:

- Stereo representation of superposition of multiple structures from ensemble
- Isosurfaces of selected residues

Fig. 3. Solution structure of the IIAD\textsuperscript{mannose}-HPr complex. A. Best-fit superposition of the backbone (N, Cα, C atoms) of the 100 final simulated annealing structures with HPr in green, and subunits A and B of IIAD\textsuperscript{mannose} in red and blue, respectively. The sidechains of the active site histidines, His\textsubscript{120} of IIAD\textsuperscript{mannose} and His\textsubscript{13} of HPr, from the restrained regularized mean structure are shown in purple. B. Isosurface of the reweighted atomic density probability map (45) for selected sidechains drawn at a value of 25% maximum calculated from the 100 simulated annealing structures. The same color scheme as in panel A is used for the atomic probability density map grid. The backbone of the restrained regularized mean structure is shown as a tube with the same color scheme as in panel A, and the sidechain coordinates within the atomic density map are colored on the basis of atom type (carbon, cyan; nitrogen, blue; and oxygen, red). Residues of HPr are labeled in italics.

E. Coli IIAD\textsuperscript{mannose}-HPr complex (48 kDa)  

Measures of Structural Quality

<table>
<thead>
<tr>
<th>Table 1. Statistics for the 20 energy-minimized conformers of A12g4790J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distant constraints</td>
</tr>
<tr>
<td>Long</td>
</tr>
<tr>
<td>Medium</td>
</tr>
<tr>
<td>Sequential [2 (i-j) = 1]</td>
</tr>
<tr>
<td>Internal [i = j]</td>
</tr>
<tr>
<td>Dihedral angle constraints (Φ and ψ)</td>
</tr>
<tr>
<td>Hydrogen bond constraints</td>
</tr>
<tr>
<td>Target function (Å²)</td>
</tr>
<tr>
<td>Average pairwise RMSD, to the mean structure (Å)</td>
</tr>
<tr>
<td>Residue 9-161</td>
</tr>
<tr>
<td>Backbone (C\textalpha, C, N, O)</td>
</tr>
<tr>
<td>Heavy atoms</td>
</tr>
<tr>
<td>Regular secondary structure elements</td>
</tr>
<tr>
<td>Backbone (C\textalpha, C, N, O)</td>
</tr>
<tr>
<td>Heavy atoms</td>
</tr>
<tr>
<td>Deviations from idealized covalent geometry</td>
</tr>
<tr>
<td>Bond (Å)</td>
</tr>
<tr>
<td>Angles (°)</td>
</tr>
<tr>
<td>Improper (°)</td>
</tr>
<tr>
<td>RMSD from experimental distance restraints (Å)</td>
</tr>
<tr>
<td>R. M. S. D. from experimental dihedral restraints (Å)</td>
</tr>
<tr>
<td>Ramachandran statistics (% of all residues)</td>
</tr>
<tr>
<td>Most favored</td>
</tr>
<tr>
<td>additionally allowed</td>
</tr>
<tr>
<td>Generously allowed</td>
</tr>
<tr>
<td>Disallowed</td>
</tr>
</tbody>
</table>

- # of experimental restraints
- RMSDs from exp restraints
- RMSDs of str superposition
- “R-factors” for RDCs
- Calculated energies
- Covalent geometry deviation
- Torsion angle comparisons to values seen in PDB for:
  - backbone Ramachandran
  - sidechain chi values
Summary of Dynamics

• Structural data provides foundation for understanding of function, catalysis, binding interactions, etc.
  – Want very precise and accurate structure
  – Be cautious in interpreting published structures based on resolution/restraints and statistics
• Dynamic fluctuations from average structure and concerted motions are critical for function and energetics
  – Want to know motional properties at variety of timescales for all positions

Part 3: Intrinsically disordered proteins

• Lack stable globular tertiary folded structure under physiological or non-denaturing conditions
• Computational prediction/analysis (Dunker et al, Tompa et al, others) suggests that:
  – Large % of eukaryotic proteins have long contiguous stretches of disordered residues
  – More disorder in
    • more complex organisms
eukaryotes > prokaryotes; multi-cellular > single cell; mammals > insects
    • more complex regulatory proteins
      cell cycle/neuronal/cancer > metabolic/housekeeping
  – Roughly 2/3 of signaling and 3/4 of human cancer-associated proteins predicted to have significant disordered regions
  – Disorder is enriched in hub proteins (>= 10 interaction partners)
• Points to role for disorder in mediating regulatory protein interactions in complex biological processes
Disordered Protein Interactions

- Flexibility/Plasticity → bind to various partners in different conformations
- Large accessible binding surfaces
  - Also accessible to kinases, phosphatases, etc.
- Kinetic benefits due to larger “capture radius” for disordered protein (“fly-casting” P Wolynes hypothesis)
- Thermodynamic control of binding to targets: optimize specificity and affinity
  → Enthalpy gain: extended interaction surface
  → Entropy loss: when folding accompanies binding leads to lower affinity and more rapid off-rates
    - strong interactions not necessarily desirable
- Additional mechanisms ….

Intrinsically Disordered Protein Properties
Disordered Protein Features (Babu et al, Science 337, 2012)
-Molecular rheostat to support continuum of conformational states and transitions
-Mediate highly specific interactions with multiple binding partners
-Conformational fluctuations control exposure of short linear motifs that interact with modular binding domains, regulating protein interactions
-Post-translational modification within or near linear motifs modulated conformation and affinities
-Multiple motifs giving rise to multivalent interactions
-Switch-like binding due to poly-electrostatics
-Rheostat-like binding
-Drive formation of higher order assemblies
- Liquid phase transitions, micrometer-sized “organelles”
- Actin regulatory proteins, RNA processing, etc
- Alternative splicing to re-wire signaling networks
- Loss of regulation/abundance of disordered proteins -> disease

Binding of Disordered Regions
Disorder-to-order transitions
Partial disorder-to-order transitions
“fuzzy complexes”/dynamic complexes
Order-to-disorder transitions
Regulated unfolding (“cryptic” disorder)
Disordered regions can induce local unfolding in adjacent domains and facilitate allosteric communication between folded domains

Determinants of protein function. Structured domains and disordered regions are fundamental units of protein function. Most proteins in eukaryotic proteomes contain both types of regions. The bar thickness shown for each type of region indicates a continuum and extent of conformational heterogeneity.
Flexibility and disorder in protein complexes

Beyond “Disorder-to-Order” Transitions: Highly Dynamic Complexes

**Sic1 and Cdc4 interact via a dynamic complex**

- each phosphorylation site interacts transiently in dynamic equilibrium based on TCS data (binding & release)
- suboptimal CPD sites gives significant off rates to stimulate exchange
- local ordering of CPDs while rest of Sic1 remains disordered
- transient structure leads to compact rapidly interconverting conformers
  -> mean electrostatic field
- phosphorylation changes electrostatics of dynamic ensemble (net charge)
  -> phosphorylation threshold for binding and “switch”-like control of cell cycle

---

**Why is Sic1:Cdc4 a dynamic complex?**

Replacement of multiple suboptimal CPD sites for a single high-affinity CPD site leads to premature cell cycle transitions, chromosome loss and genome instability.

Recognition of WT Sic1 multiple phosphorylation sites by Cdc4 leads to switch-like response to kinase levels.

---


Ubiquitination within Sic1-Cdc4 complex

- Dynamic exchange facilitates ubiquitination on multiple lysines of Sic1 in catalytic site on E2, far (64Å) from Cdc4 binding site
  - disordered conformations bridge gap
- Sic1 lacking all natural CPD phosphorylation sites with high-affinity CPD at the extreme N-terminus -> lysine residues 84/88 preferentially ubiquitinated with ubiquitination of more N-terminal lysines disfavored (Tang et al., 2007)

“Polypelectrostatic” model of interaction of IDPs
Borg et al, PNAS, 2007
Schematic of an IDP (ribbon) interacting with folded receptor (grey shape) through multiple distinct binding motifs and an ensemble of conformations (4 shown). The IDP has positive and negative charges (blue and red circles) giving rise to net charge $q_i$, while the binding site in the receptor (teal) has a charge $q_r$. The effective distance $<r>$ is between the binding site and the center of mass of the IDP.
Dynamic complex of Regulatory (R) region of CFTR

~200 residue disordered region
contains 9 PKA as well as PKC/AMPK phosphorylation sites
acts as protein interaction “hub” and integrator of regulatory inputs

![Diagram of CFTR in non-phosphorylated and phosphorylated states](image)

CFTR in non-phosphorylated and phosphorylated states

R region binds CFTR NBD1, NBD2, C-terminus, N-terminus, …
R region binds STAS domain of SLC26A, 14-3-3, kinases, phosphatases, …
with many of these interactions phosphorylation dependent

Ensemble model of full length CFTR in closed channel state

NMR data for R region transient helical structure and R region:NBD interactions in non-phosphorylated state

P-glycoprotein X-ray structure-based homology model for core of CFTR

allowed disordered elements to sample multiple conformations

variety of states with up to three helical elements formed and with R region interactions to NBD1, NBD2 or both

→ block NBD heterodimerization (released by phosphorylation)
Why disordered proteins & dynamic complexes?

Phosphorylation/PTMs enables changes in binding affinities

- Kinase/phosphatase activity requires extended target

Conformational plasticity to bind multiple partners in different conformations (helical/extended)

Potentially faster off-rates due to entropic penalty to binding to enable dynamic exchange

Overlapping binding sites leads to competitive binding and a variable functional output that is responsive to “input” of large number of available partners

Dynamic averaging of electrostatic charges potentially enabling affinity to be sensitive to net charge that may contribute to rheostat effect/switch-like binding

-> Together intrinsically disordered regions can function as integrators of various intra- and inter-molecular regulatory inputs to control biology

Multivalent interactions of disordered proteins facilitate higher order assemblies - granules, organelles, liquid phase separated droplets, etc.

Disordered proteins also are able to form amyloid beta-sheet structures

Disordered proteins are the “polymer” state of proteins with many possible conformations accessible


Figure 1. Intracellular Biomolecules Undergo Phase Transitions

Diagrams show that biomolecules can exist in different phases (gas, liquid, solid) with many transient intermolecular interactions and phase transitions between these phases are driven by external parameters such as temperature. These can be induced by small external perturbations (e.g., ATP) and other small molecules can induce or inhibit these transitions. Multivalent interactions involving low-complexity sequences drive soluble and amorphous structures to transiently associate with each other, forming liquid-likemolecules in the cell. ATP-dependent biological activity could enable the cells to transiently control the assembly and disassembly of these phases. Intriguingly, structural characterization suggests that the same intermolecular interactions underlying granule assembly can lead to the formation of stable amyloid aggregates seen in diseases such as Alzheimer's.
Multivalent interactions between proline-rich motifs within a disordered protein and multiple SH3 domains give rise to macroscopic association

This mimics interactions thought to be required for actin polymerization, leading to a “granule” or organelle in which substrates for actin are concentrated within the cell


---

**Continuums**
many/most aspects of biology/biochemistry are not discrete (quantized) but are continuums, distributions, ranges
– requiring thinking in more “ensemble” and “dynamic” ways

- folded protein to partially disordered to fully disordered
- static to mobile to highly dynamic
- static interface to highly dynamic interface
- complex between folded domains (domain-domain) to complex between folded domains and disordered regions (domain-peptide) to complex between disordered regions
- interactions involving 1:1 complex (dimer) to discrete oligomers to large scale assemblies
- folding upon binding to complex retaining significant dynamic properties to unfolding upon binding
- allostery via conformational transition between 2 discrete conformations to allostery involving energetic changes leading to conformational shift within more complex ensemble to allostery involving energetic shift leading primarily to change in dynamic properties