An introduction to the study of intrinsically disordered proteins with NMR:
Are they ever completely disordered?

Geoff Moore
Centre for Molecular & Structural Biochemistry
UEA Norwich

- Characteristics of intrinsically disordered proteins
- Colicins & Tol proteins
- Studying disordered proteins with NMR
- Order within disordered states
Intrinsically disordered proteins aka Intrinsically unstructured, Natively disordered and Natively unfolded proteins

What does it mean to be natively unfolded?

- extreme flexibility
- extended
- little or no 2° structure

“Amino acid sequence characterised by low mean hydrophobicity and relatively high net charge”
Intrinsically disordered proteins

Multi-domain proteins with intrinsically disordered regions interspersed with structured regions are common (‘Beads-on-a-string’).

Example:
CBP – 2442 residue scaffold protein for assembly of the transcriptional machinery as well as key enzyme activity. 7 globular domains that fold independently (4 requiring Zn\(^{2+}\)) separated by disordered regions.

Origin of disorder: Complex IDPs

Amino acid sequence resembles those of folded proteins, but

\[ U \leftrightarrow N \]

Often (always?) fold on binding partner.

Many examples.

Origin of disorder: Low complexity IDPs

High content of small amino acids (e.g. G,N,S,D)

Often (always?) retains considerable dynamic disorder in bound state.

Often linear intermolecular binding epitope(s) within the disordered sequence.

Examples: Translocation domains of colicins, disordered regions of scaffold proteins.
Origin of disorder: Unbalanced charge IDPs

Amino acid sequence has high uncompensated charge density leading to high repulsive electrostatic energies.

Example:
Structure of Hen Phosvitin: A $^{31}$P NMR, $^1$H NMR, and Laser Photochemically Induced Dynamic Nuclear Polarization $^1$H NMR study. Vogel (1983) *Biochemistry* 22, 668-674 (more than half of the 220 residues are phosphoserine).
Common experimental indicators of disorder

- Invisible in X-ray electron density maps – though invisibility may also occur for a structured domain tethered to another structured domain by a flexible linker
- Anomalous passage through gel-filtration columns – behaves as though mass is greater than it really is
- Heightened tendency for peptide cleavage
- Little or no compactness indicated by hydrodynamic radius
- Narrow chemical shift dispersion in NMR spectra
- Sharp resonances in NMR spectra
Bioinformatics and IDPs

Many programs predict IDPs from amino acid sequences.

PONDR prediction for N-terminal domain of colicin N

http://www.pondr.com
DisProt
Database of Protein Disorder

http://www.disprot.org/
Accessed Sept 5 2011

Number of proteins: 645
Number of disordered regions: 1388
How common are intrinsically disordered proteins?

- >50% of human proteins may have disordered regions of 50 amino acids or longer [PONDR analysis]

- 33% of eukaryotic proteins and 4.2% of eubacterial proteins have disordered regions of 30 residues or more [DISOPRED analysis]

But, prediction of post-translational modifications, including metal binding and oligomerisation, is not as advanced as prediction of structure from amino acid sequences so these may be over-estimates.
Are all reported IDPs really IDPs?

NMR shows *Salmonella typhimurium* FlgM is unstructured *in vitro* in absence of crowding agents but structured inside living *Escherichia coli* cells and *in vitro* in crowded solutions. 


Many NMR groups - α-Synuclein is an IDP

α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. 

Disordered proteins and intermolecular interactions

- Disordered proteins allow large intermolecular interfaces with smaller protein, genome and cell sizes
  Gunasekaran et al. (2003) *TIBS* 28, 81-85

- Disorder enhances rates of intermolecular interactions by a ‘flycasting’ mechanism
  Shoemaker et al. (2000) *PNAS* 97, 8868-8873

- Many natively unfolded proteins fold into an ordered structure on binding a partner

- Entropic cost of folding a disordered protein on binding compensated by enthalpic gain leading to high specificity and low affinity
A cartoon of how fly casting increases folding speed.

Shoemaker B A et al. PNAS 2000;97:8868-8873
Binding mechanisms of disordered proteins

**Conformational selection** - the conformational ensemble of the unbound IDP has a population of molecules in the 'bound' conformation which the binding partner preferentially selects

\[ \sum U \leftrightarrow U_1 + U_2 + \ldots + U_n + F \]

**Induced fit** - the binding partner acts as a template for the IDP to fold

\[ U + \text{partner} \leftrightarrow F\text{-partner} \]
Binding mechanisms of disordered proteins

For many IDPs probably not simply 'binding before folding or folding before binding' but more complex.


NMR determination of binding mechanisms of disordered proteins

Mechanism of coupled folding and binding of an intrinsically disordered protein. (i.e. induced fit)

Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP.
Colicins & Tol proteins

- Secreted by producing bacteria to kill competing bacteria
- 3-domain structure:
  - Translocation – Receptor – Killing Domain
  - Domain – Domain – Domain

Colicin N – forms a pore in the inner membrane
Colicins E2, E7, E8 & E9 – DNase action
Tol A & B – in periplasm of target cell; helps colicin enter cells
Colicins have globular domains

X-ray structure of the 551 residue colicin E3 (with bound Im3) (Soelaiman et al., 2001, Molecular cell 8, 1053-1062)

X-ray structure of the 351 residue colicin N (Vetter et al., 1998, Structure 6, 863 – 874)
And also disordered domains that are functional

TolB binds at residues 32-47 of colicin E9

TolA binds at residues 40-67 of colicin N

Residues 1-83 of colicin E3 and 1-90 of colicin N are not visible in X-ray electron density maps
1H-15N NMR of colE9 T_{61}-DNase fusion protein

<table>
<thead>
<tr>
<th>MSGGDGRGHN</th>
<th>TGAHSTSGNI</th>
<th>NGGPTGIGVS</th>
<th>GGA</th>
<th>ASDGSGWS</th>
<th>SENNPWGGGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td></td>
<td>50</td>
</tr>
</tbody>
</table>

Spectrum of 15N labelled protein at a high threshold of display showing only sharp peaks of T-domain (left) and a level with additional broad signals from the DNase domain (right)

MacDonald et al. (2004) J Biomol NMR 30, 81-96
Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and $^{15}$N sequential correlations in ($^{13}$C, $^{15}$N) labeled proteins: Application to unfolded proteins


HNN is a modified HNCA sequence

It correlates $^1$HN and $^{15}$N of (i) with $^{15}$N of both (i+1) and (i-1)
Gly resonances and HNN spectrum

Two $^{15}$N dimensions enhance resolution; the phase pattern of the diagonal and sequential peaks obtained for (i) is determined by the amino acid type of both (i-1) and (i+1) residues.
Assigning Gly resonances of ColE9 T-domain

HNCACB (left) and HNN (right) strips; red (-ve), black (+ve). Potential (i+1) strips for HNCACB shown. HNN gave assignment.

MacDonald et al. (2004) J Biomol NMR 30, 81–96
$^{1}H-^{15}N$ NMR of 203 amino acid $T_{61}$-DNase colicin E9 fusion protein

$^{15}N$ $T_1$ ($1/R_1$)

$^{15}N$ $T_2$ ($1/R_2$)

$^{1}H-^{15}N$ NOE

MacDonald et al. (2004) J. Biomol. NMR 30, 81–96
Modelling polypeptide $^{15}$N $R_2$ rates

In a fully disordered polypeptide the motions of one residue are unaffected by other residues.

Inter-residue side chain – side chain interactions may modulate backbone motions where clusters of side chains form (e.g. as a consequence of their hydrophobicity).
Modelling polypeptide $^{15}$N $R_2$ rates


$$R_{2i} = k \cdot \sum_{j=1}^{N} \tau_j \cdot e^{-(i-j)/\lambda_j} + \sum_{\text{cluster}} R_{\text{cluster}} \cdot e^{-(i-n_c)^2/\lambda_{\text{cluster}}^2}$$

$\eta_c$ is the cluster centre

$\lambda_j$ and $\lambda_{\text{cluster}}$ are the persistence lengths

$\tau_j$ proportional to radius of gyration

$R_2 = 1/T_2$

$\tau_j = 1^{\text{iii}}$
Clusters in residues 1-61 of colicin E9

- Major clusters formed around Trp39, Trp46, Trp56
- Minor clusters may be based around Asn10 and His14

Colicin E9 clusters interact with each other

1/\(T_2\) (s\(^{-1}\))

residue number

\(\Delta \delta^{15}N\)

Wild-type - Asp35Ala

Wild-type - Ser37Ala
Clusters of colicin E9 predicted by the Random Coil Index

wishart.biology.ualberta.ca/rci/cgi-bin/rci_cgi_1_e.py

Berjanskii & Wishart (2005) JACS 127, 14970–14971. And
AABUF mirrors colE9 clusters detected by NMR


AABUF = Average area buried upon folding; correlated with residue hydrophobicity. Rose et al (1985) Science 229, 834-838
$^{15}\text{N} \ T_{61}$-DNase fusion protein without (blue) and with (red) unlabelled TolB

Residues affected by TolB extend beyond the TolB box

Only the TolB binding residues are significantly affected, the remaining T61 region is dynamically disordered.
TolA recognition site of colicin N is not completely extended

MGSNGADNAH NNAFGGGKNP GIGNTSGAGS NGSASSNRGN SNGWSWSNKP
10 20 30 40 50
HKNDGFHSDG SYHITFHGDN-------------------
60 70 387

Hydrodynamic radius of his-tagged N-terminal 90 residue polypeptide (R_h)
- Experimental R_h (NMR): 19.2 ± 0.2 Å
- Predicted R_h for fully unfolded chain: ~30 Å
- Predicted R_h for globular protein: ~18 Å

Colicin N translocation domain is flexible and lacks secondary structure but is compact

Colicin N T-domain$^{1-89}$
TolA binding site of colicin N binds intramolecularly to colicin N

G.S.N.G.A.D.N.A.H.N.N.A.F.
G.G.G.K.N.P.G.I.G.N.T.S.G.
A.G.S.N.G.S.A.S.S.N.R.G.N.
S.N.G.W.S.W.S.N.K.P.H.K.N.
D.G.F.H.S.D.G.S.Y.H.I.T.F.
H.G.D.N.N.S.K.P.K.P.G.G.N.
S.G.N.R.G.N.N.G.D.G.A.

TolA binding site

Residues unassigned in isolated T-domain and intact colicin

Underlined residues not detected in intact colicin

TolA binding site of colicin N is a **sticky patch** binding to OmpF, TolA & colicin N

**OmpF** – outer membrane receptor for colicin N

**TolA**₃ – helps colicin N translocate to inner membrane

**Colicin N R-domain** binds T-domain and provides protection against proteolysis

Compare IDPs with unfolded/partially folded states of globular proteins

i.e. go from $U \leftrightarrow N$ to $U \leftrightarrow N$

Addition of urea
High temperature
Low pH
Removal of cofactor
Mutation
Combination
Non-interacting clusters in urea-unfolded Im7

\[ R_2 = \frac{1}{T_2}^{15N} \]

Le Duff et al. (2006) JMB 364 824.
Order in unfolded states of globular proteins

Figueiredo et al. (2011) Biochem Soc Trans. in press

NMR shows globular proteins unfolded with urea often have non-random and non-native structure with few (if any) long-range interactions (like Im7). But globular proteins unfolded without urea (like colicin T-domains) generally do have non-random structure that may be native-like and with long-range interactions. Examples:


Acknowledgements

UEA Norwich
Ruth Boetzel
Nick Cull
Olli Hecht
Colin Macdonald
Kaeko Tozawa

York
Colin Kleanthous

Birmingham
Sara Whittaker

Nottingham
Chris Penfold
Richard James

Newcastle
Jeremy Lakey
Helen Ridley

HWB•NMR

UEA
University of East Anglia

BBSRC

Wellcome Trust

Wolfson Foundation