Intrinsically Disordered Proteins: The BMRB(data) Perspective

Pedro R. Romero

Director BMRB

University of Wisconsin - Madison

NMR and Intrinsic Disorder

- Best high-resolution experimental technique to study IDPs
  - Structural ensembles
  - Dynamics
  - Interactions
- Hybrid studies: SAXS, X-ray, MS, EM
- Since 2014, roughly 75% of intrinsic disorder publications mention NMR

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - Chemical shift assignments
    - Proton-detected NMR
      - Selective aa labelling
      - Prolines: long-range correlation methods
      - Sequential data acquisition over sub-types
      - Non-uniform sampling
      - Hi-D (4d and 5D) CON-CON spectra
    - Heteronuclear direct-detect NMR
      - CAS-NMR (carbon detected)
      - N-detected

Gibbs *et al. Archives of Biochem & Biophys* 628 (2017) 57-70

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - Structural constraints
    - α-helical structure (CS)
      - Use CS reference sets for random coil
    - Other secondary structure (or lack thereof)
      - Large scalar coupling datasets
    - Residual dipolar couplings (RDC)
    - Paramagnetic relaxation enhancement (PRE)

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - IDP Dynamics
    - Spin relaxation of C, N, and H
    - Spectral density mapping
      - Range of field strengths
        - Direct spectral density mapping
          - dynamic time scales

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - Post translational modifications (PTM)
    - Phosphorylation
      - HN-HSQC + NC-CON + 3D CCCON
      - Real-time NMR (RT-NMR) (multi-site phosphorylation)
    - Acetylation (e.g., histones)
      - Hi-Res NMR on cellular extract

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - IDP Interactions
    - Intermediate states and transient interactions (coupled folding and binding)
      - Relaxation dispersion NMR
        - Both kinetic and structural information
    - Membrane / lipid binding
      - Chemical exchange saturated transfer (CEST)
        - Characterize sparsely populated states

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - Liquid-liquid phase separation
    - Low-complexity domains
      - Chemical shift perturbations
    - Folding in droplet
      - Secondary chemical shifts
    - IDPs and LP viscosity
      - Pulse field gradient (PFG) diffusion
      - Nuclear spin relaxation
      - PRE
      - Relaxation dispersion

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - Hydrogel formation
    - Solid-state NMR
      - Cross-polarization (CP) and scalar coupling magnetization transfer (CP-HSQC and J-HSQC)

NMR and Intrinsic Disorder

- A summary of current studies and approaches
  - Amyloid proteins (aggregation)
    - Solution NMR
      - RT-NMR
      - F-direct detect NMR
      - PRE
      - Dark-state saturation transfer (DEST)
    - ssNMR
      - C-C dipolar assisted rotational resonance (DARR)
      - NC transferred-echo double resonance (TEDOR)

The Biological Magnetic Resonance data Bank

- Structural origins
  - Protein Structure Initiative (PSI)
  - Member of wwPDB consortium
- NMR-STAR data format
  - STAR-based (human and machine readable)
  - Large and continuously evolving ontology
  - Handles non-structural NMR experiments

NMR Data Currently Handled at BMRB

- NMR spectral parameters (chemical shifts, coupling constants, time-domain data, spectral peak lists, and RDCs),
- Relaxation data (R1/T1, R2/T2, R1rho/T1rho, heteronulear NOEs),
- Other kinetic data (H-exchange, chemical rates)
- Thermodynamic data (pKa, binding constants,order parameters).
- Other forms of data are accepted with sufficient description.

Some Stats

BMRB Query Grid Interface

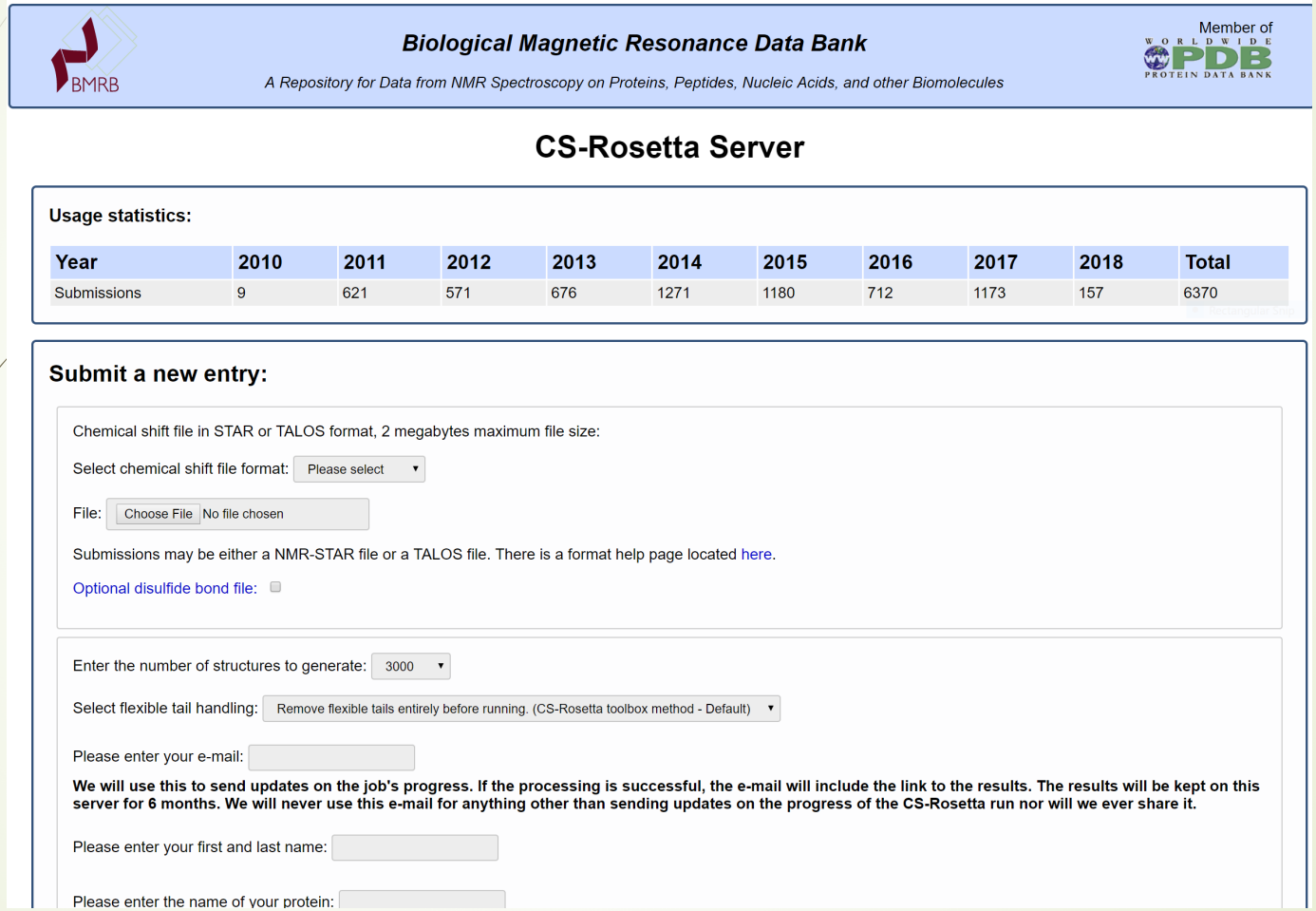
Current Content of the BMRB Archive

BMRB entry list (11932)

Clicking on a link in one of the boxes in the above table will take you to a BMRB entry listings for the type of biopolymer and type of data represented by the location of the box in the grid. Values in parentheses indicate the number of entries for that category.

Data Type	*Polymer Class*		
	Proteins/Peptides (11364)	DNA (376)	RNA (337)
All Chemical Shifts	7839154 (11064)	51929 (321)	75136 (289)
1H Chemical Shifts	4014667 (10740)	47513 (317)	47345 (288)
13C Chemical Shifts	2923696 (8046)	3232 (46)	23496 (180)
15N Chemical Shifts	904311 (8329)	121 (10)	3811 (118)
31P Chemical Shifts	-	1102 (70)	727 (55)
Other Chemical Shifts	-	-	-
Coupling Constants	28147 (363)	131 (5)	-
Dipolar Couplings	13972 (120)	-	-
T1 Values	37648 (246)	-	-
T2 Values	39226 (245)	-	-
Heteronuclear NOE Values	35789 (244)	-	-
S2 Values	15163 (93)	-	-
H-Exchange Rates	1561 (18)	-	-
H-Exchange Protection Factors	727 (10)	-	-

High Troughput Computing



Biological Magnetic Resonance Data Bank

A Repository for Data from NMR Spectroscopy on Proteins, Peptides, Nucleic Acids, and other Biomolecules

Member of WORLDWIDE PDB PROTEIN DATA BANK

CS-Rosetta Server

Usage statistics:

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Submissions	9	621	571	676	1271	1180	712	1173	157	6370

Submit a new entry:

Chemical shift file in STAR or TALOS format, 2 megabytes maximum file size:

Select chemical shift file format: Please select

File: Choose File No file chosen

Submissions may be either a NMR-STAR file or a TALOS file. There is a format help page located here.

Optional disulfide bond file:

Enter the number of structures to generate: 3000

Select flexible tail handling: Remove flexible tails entirely before running. (CS-Rosetta toolbox method - Default)

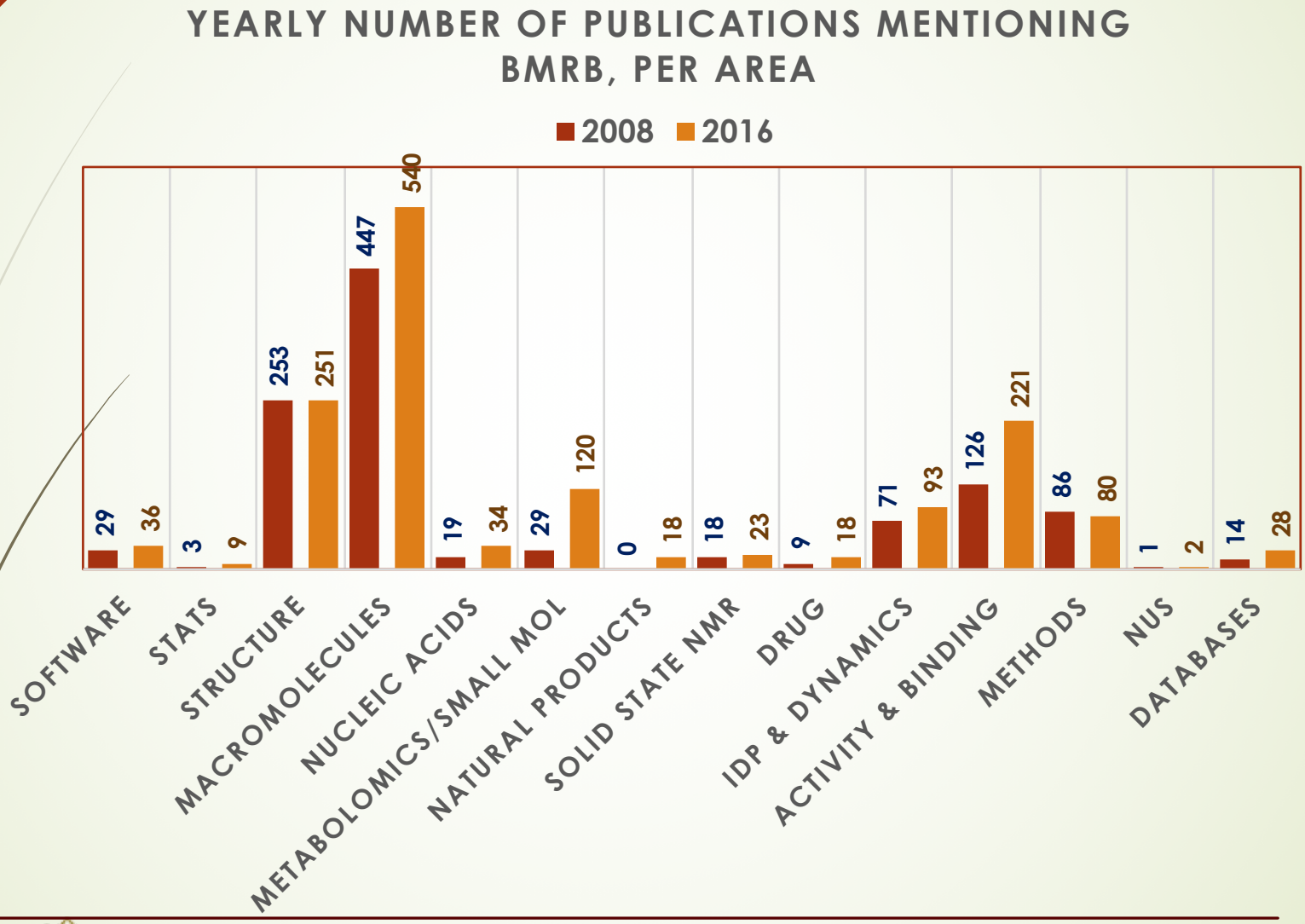
Please enter your e-mail:

We will use this to send updates on the job's progress. If the processing is successful, the e-mail will include the link to the results. The results will be kept on this server for 6 months. We will never use this e-mail for anything other than sending updates on the progress of the CS-Rosetta run nor will we ever share it.

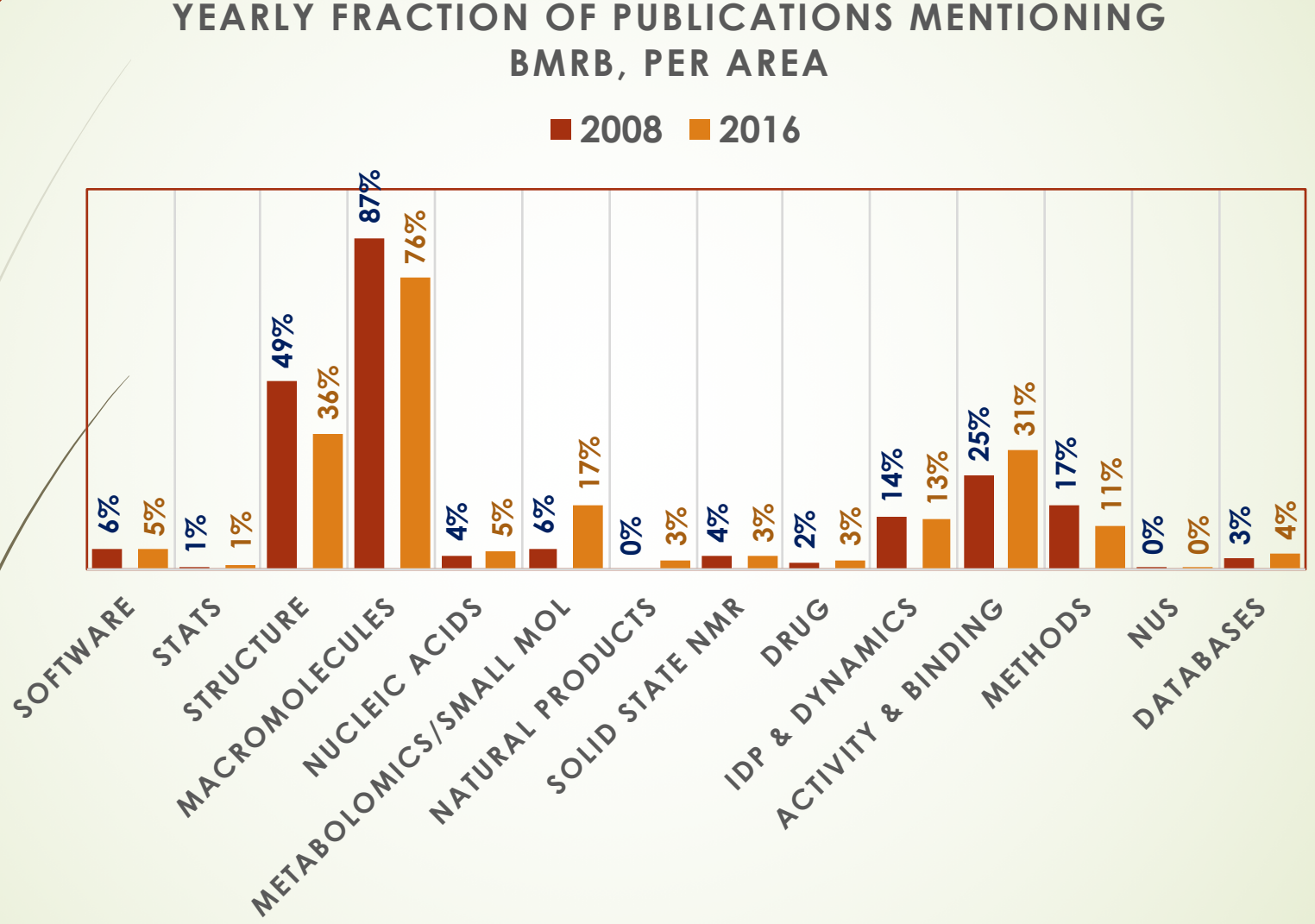
Please enter your first and last name:

Please enter the name of your protein:

BMRB in the NMR Literature

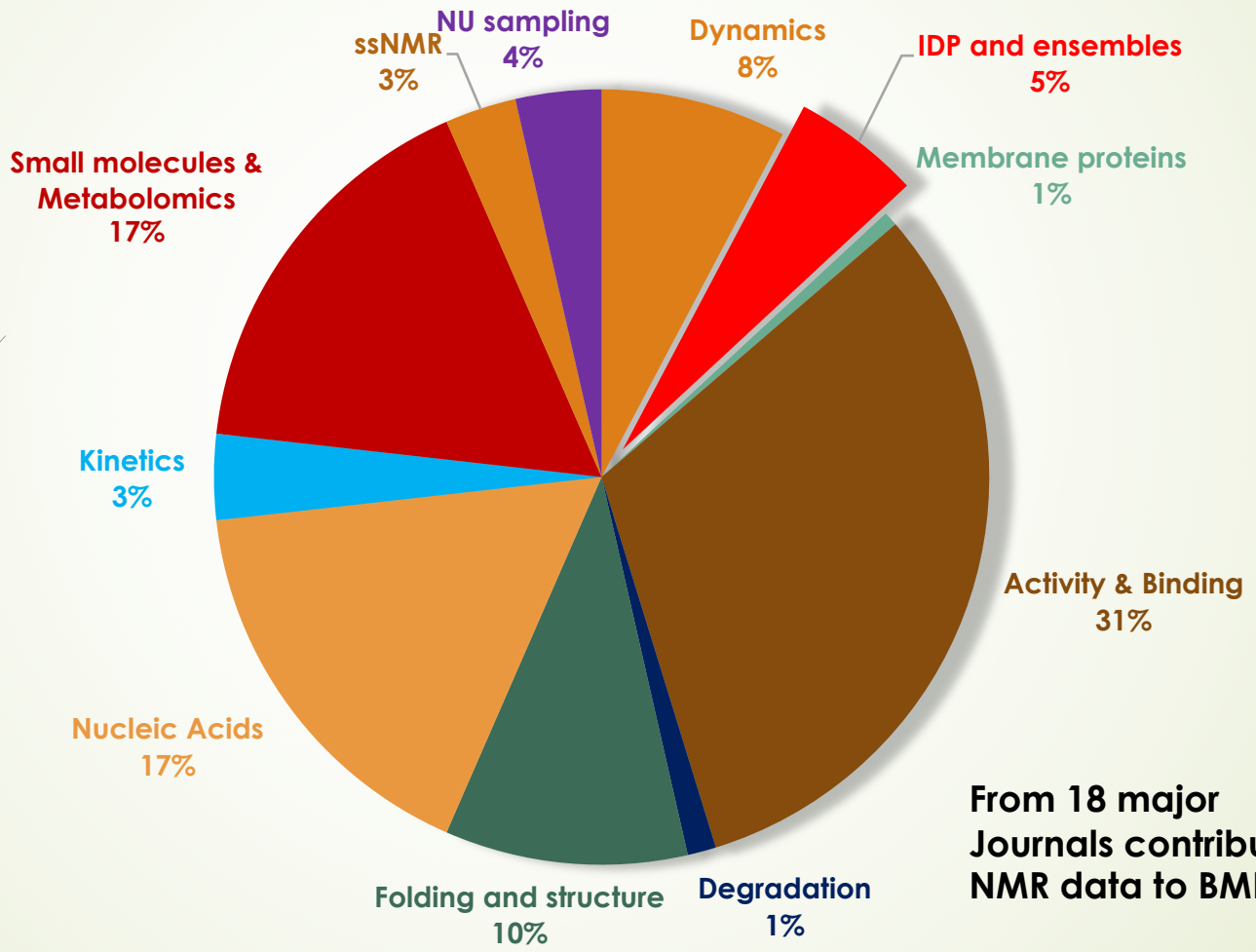


BMRB in the NMR Literature



Is There Missing Data in BMRB?

NON-DEPOSITED NMR PUBLICATIONS BY AREA (JULY 2017-JANUARY 2018)



From 18 major Journals contributing NMR data to BMRB

Finding disordered regions in NMR structures

J Struct Biol. Author manuscript; available in PMC 2014 Jan 1.
PMCID: PMC3529935
Published in final edited form as:
NIHMSID: NIHMS421345
J Struct Biol. 2013 Jan; 181(1): 29–36.
PMID: 23142703
Published online 2012 Nov 7. doi: 10.1016/j.jsb.2012.10.017

An assignment of intrinsically disordered regions of proteins based on NMR structures

Motonori Ota,[1,*] Ryotaro Koike,[1] Takayuki Amemiya,[1] Takeshi Tenno,[2] Pedro R. Romero,[3] Hidekazu Hiroaki,[2] A. Keith Dunker,[3] and Satoshi Fukuchi[4]

Author information · Copyright and License information Disclaimer

Finding disordered regions in NMR structures

- Used proteins with IDRs with both X-ray and NMR structures in the PDB
- Statistics analysis determined an optimal threshold of 3.2 Å RMSD between models
- Limited NOE data in BMRB

BMRB Internal Study

- Gathered NOE restraints and relaxation data for structural entries
- Used tools (Cyrange and RCI) to determine ill-defined regions ("disordered")
- Looked at correlations

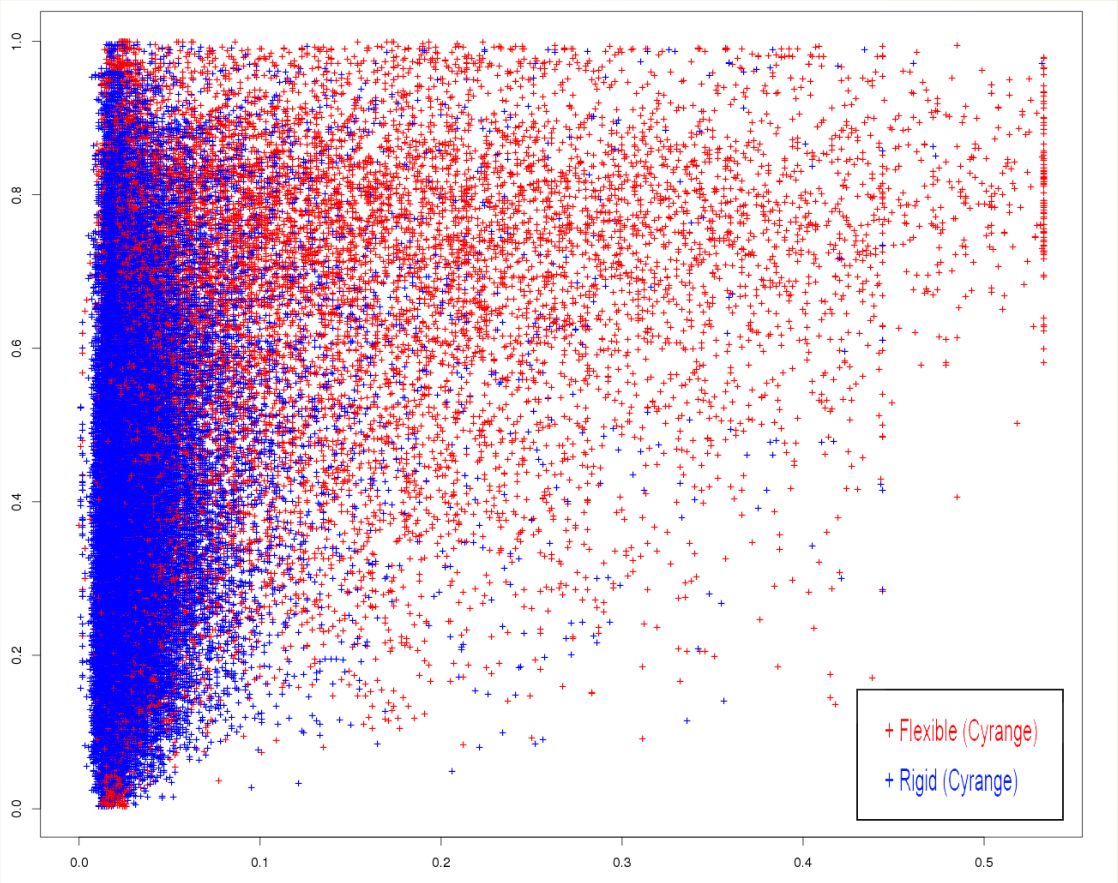
BMRB Internal Study

- Gathered NOE restraints and relaxation data for structural entries
- Used tools (Cyrange and RCI) to determine ill-defined regions ("disordered")
- Looked at correlations

Exploratory Analysis

Distribution of flexible and rigid residues (Cyrange) by RCI and disorder prediction score

PONDR VSL2b Score

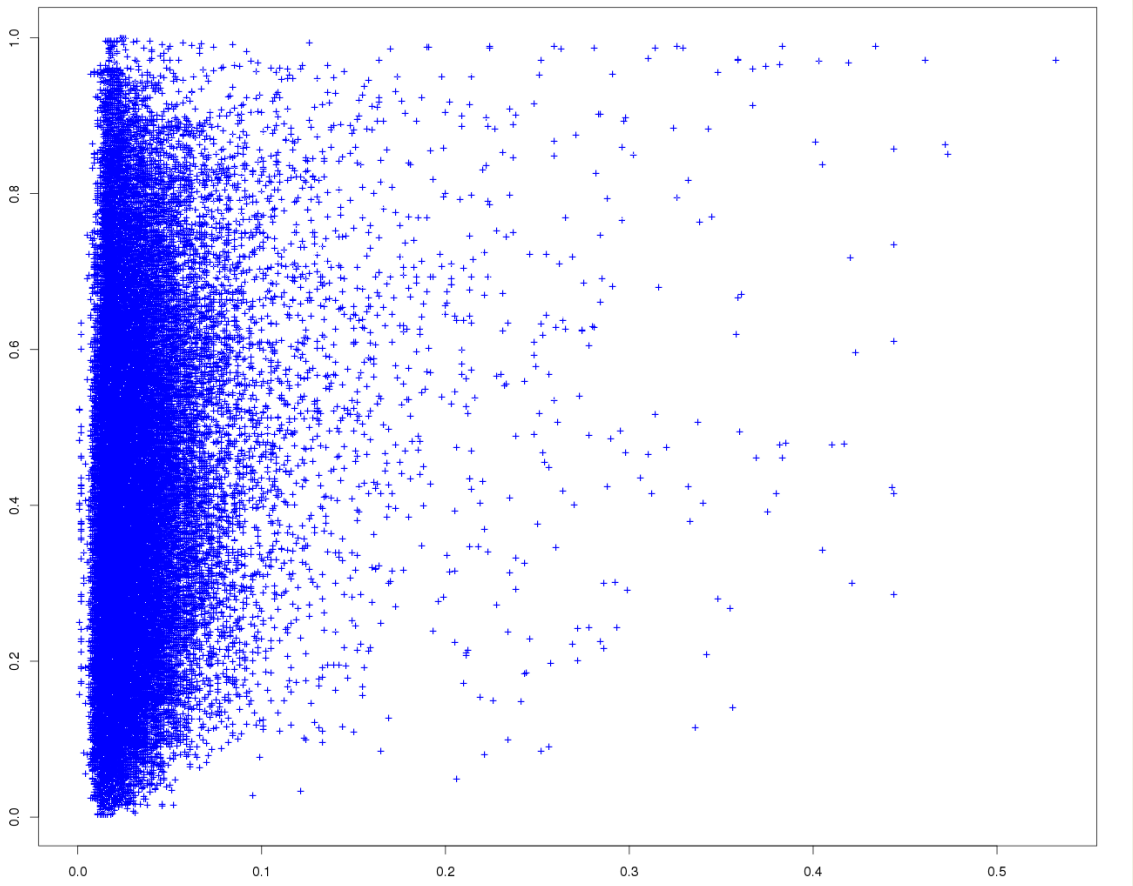


RCI

Exploratory Analysis

Distribution of rigid residues (Cyrange) by RCI and disorder prediction score

PONDR VSL2b Score

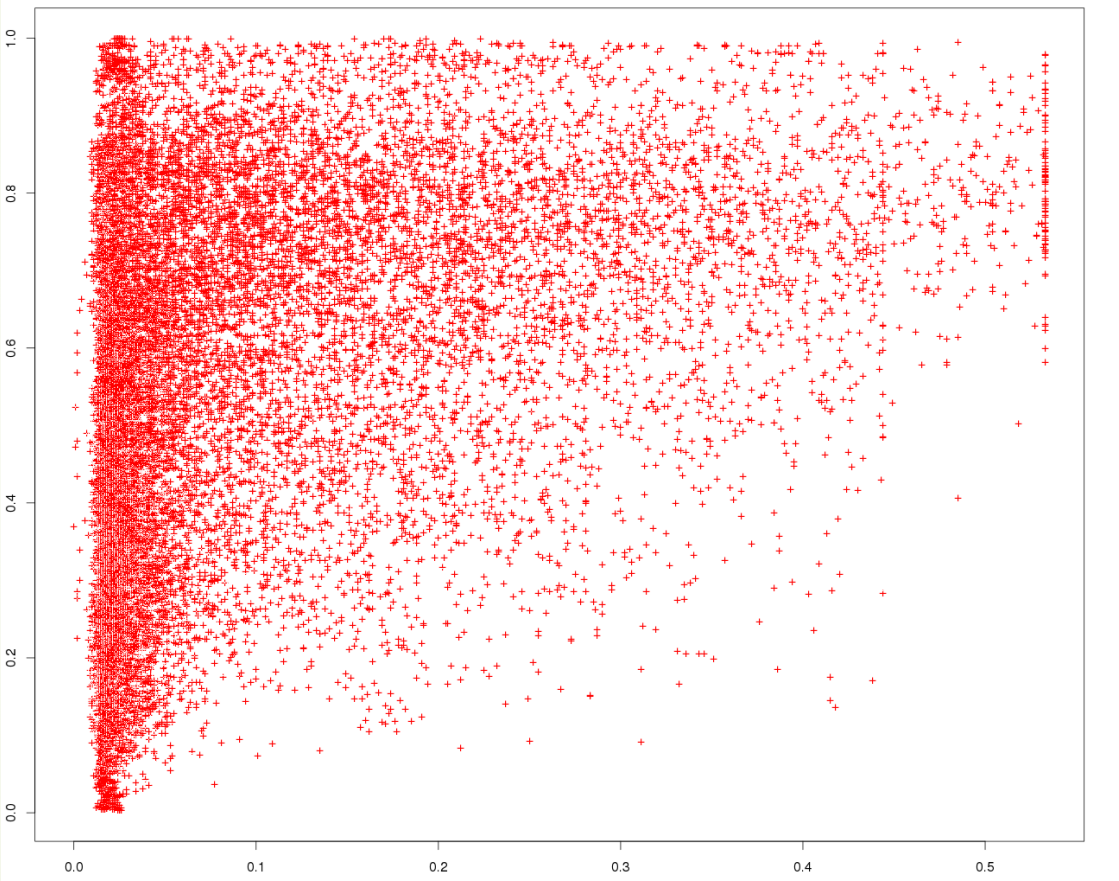


RCI

Exploratory Analysis

Distribution of flexible (Cyrange) by RCI and disorder prediction score

PONDR VSL2b Score



RCI

Exploratory Analysis

Classification order (rigid) / disorder (flexible) by RCI and PONDR VSL2b vs Cyrange

Confusion Matrices

RCI	Reference	(Cyrange)
Classification	Order	Disorder
ORDER	59,195	12,016
DISORDER	1,095	6,126

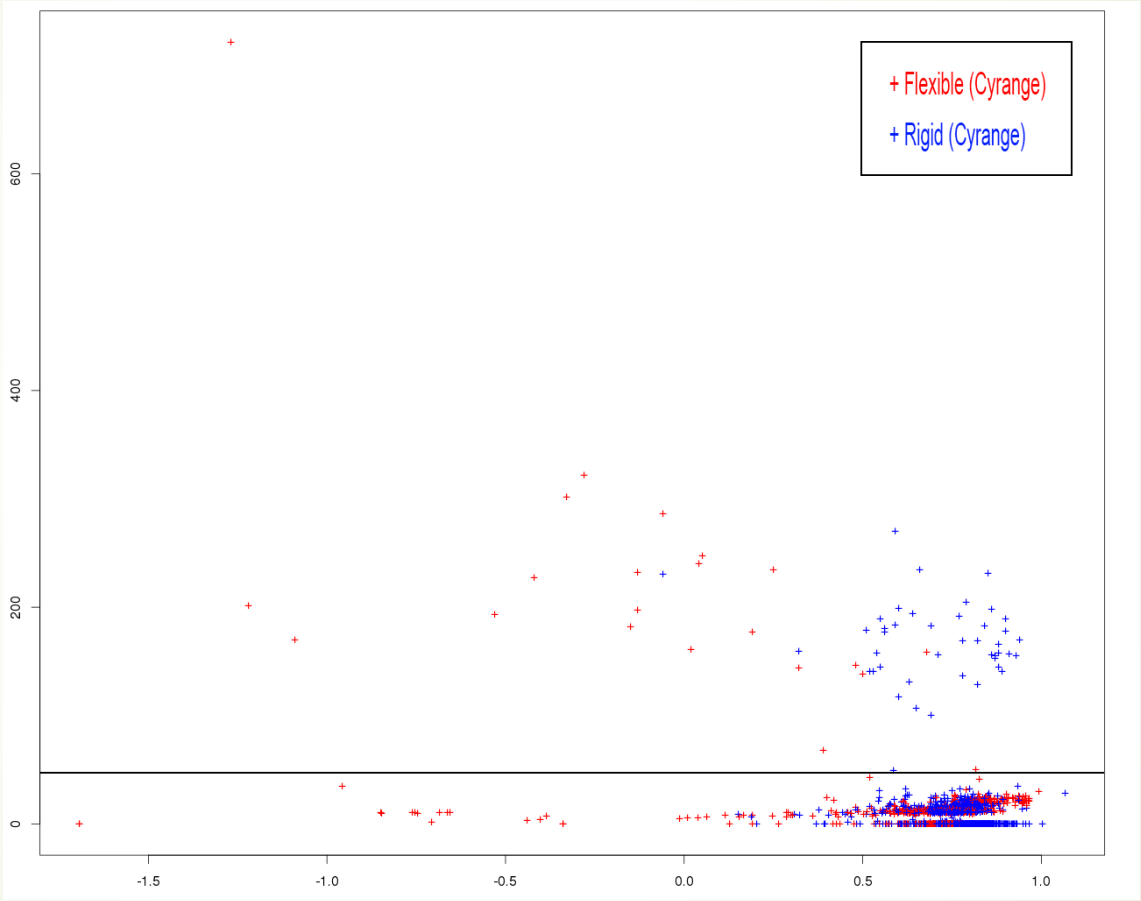
PONDR VSL2b	Reference	(Cyrange)
Classification	Order	Disorder
ORDER	42,183	7,108
DISORDER	18,107	11,034

Classification Stats

Measure	RCI (Disorder: > 0.1)	PONDR VSL2b (Disorder: > 0.5)
Accuracy	0.833	0.679
Sensitivity (order)	0.982	0.700
Specificity (disorder)	0.338	0.608
Balanced Accuracy	0.660	0.654

Exploratory Analysis

Distribution of flexible and rigid residues (Cyrange) by heteronuclear NOE



T2

~50

NOE

Exploratory Analysis

Classification order (rigid) / disorder (flexible) by (H)N NOE vs Cyrange

Confusion Matrices

(H)N NOE (global)	Reference	(Cyrange)
Classification	Order	Disorder
ORDER	752	501
DISORDER	22	68

(H)N NOE (T2 > 50)	Reference	(Cyrange)
Classification	Order	Disorder
ORDER	41	5
DISORDER	2	19

Classification Stats

Measure	(H)N NOE (global)	(H)N NOE (T2 > 50)
Accuracy	0.611	0.896
Sensitivity (order)	0.972	0.954
Specificity (disorder)	0.120	0.792
Balanced Accuracy	0.546	0.873

Exploratory Analysis

- Summary
  - RCI (Chemical shift based experimental measure)
    - Order (rigid): almost perfect discrimination (98%)
    - Disorder (flexible): extremely poor discrimination (34%)
  - PONDR VSL2b (Prediction)
    - Order: moderate discrimination (70%)
    - Disorder: moderate discrimination (61%)
  - Heteronuclear NOE (Experimental – ***few examples***)
    - Order: Excellent alone (98%) and for high T2 (95%)
    - Disorder: Dismal alone (12%), great for high T2 (80%)

BMRB as a Biomolecular NMR Database

- Structural origins and wwPDB membership has "fixed" BMRB as a structural DB for journals
  - Only require deposition of structural NMR data
- BMRB can handle most NMR experiments and techniques
- Data format keeps evolving according to the needs of the NMR community

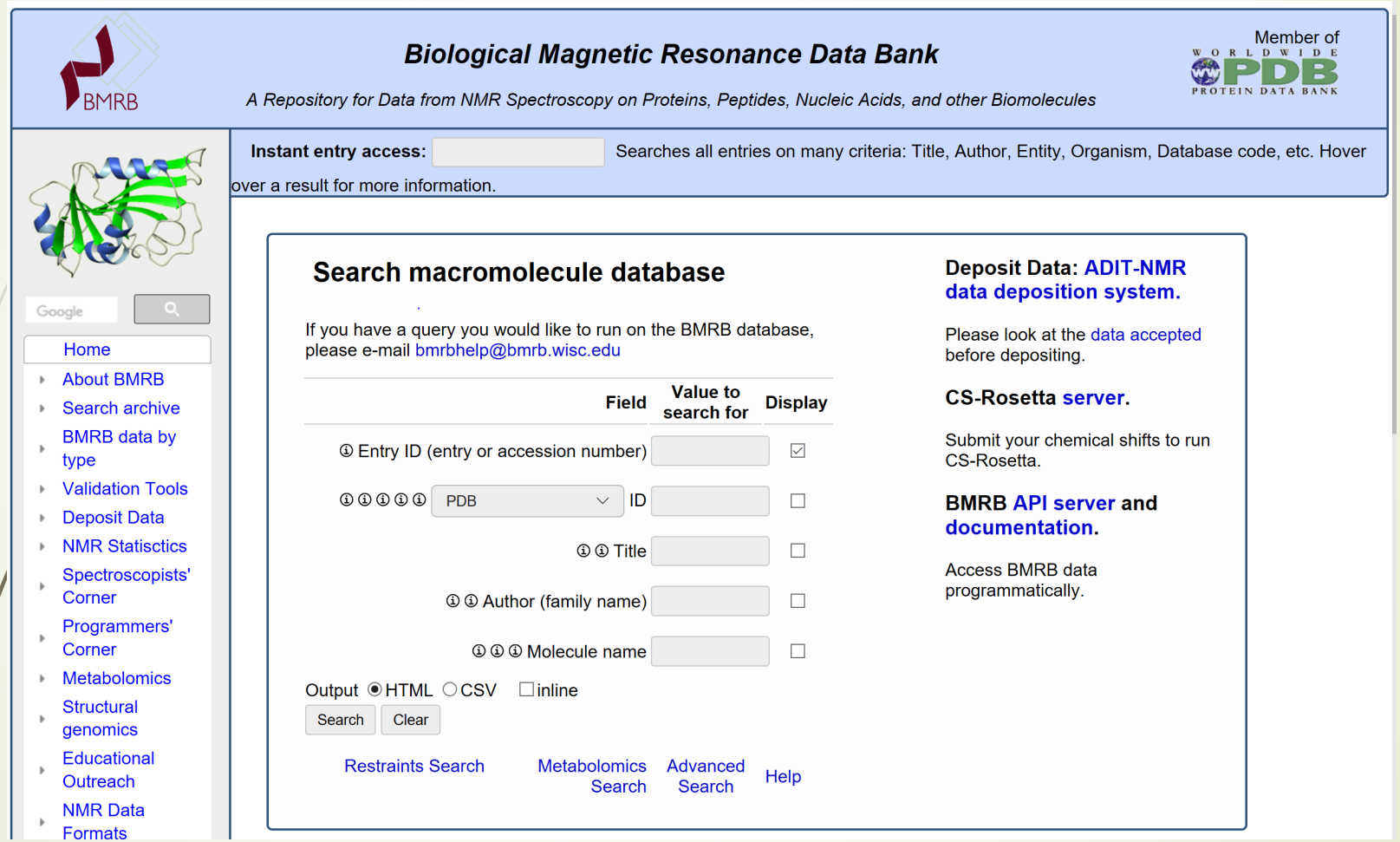
Securing Missing Depositions

- Missing data on important and emerging areas of NMR
- Work with journals and NMR scientific communities, like IDPs.
- Community involvement leads to improvements in data format and resources
- More and more complete data for data science research

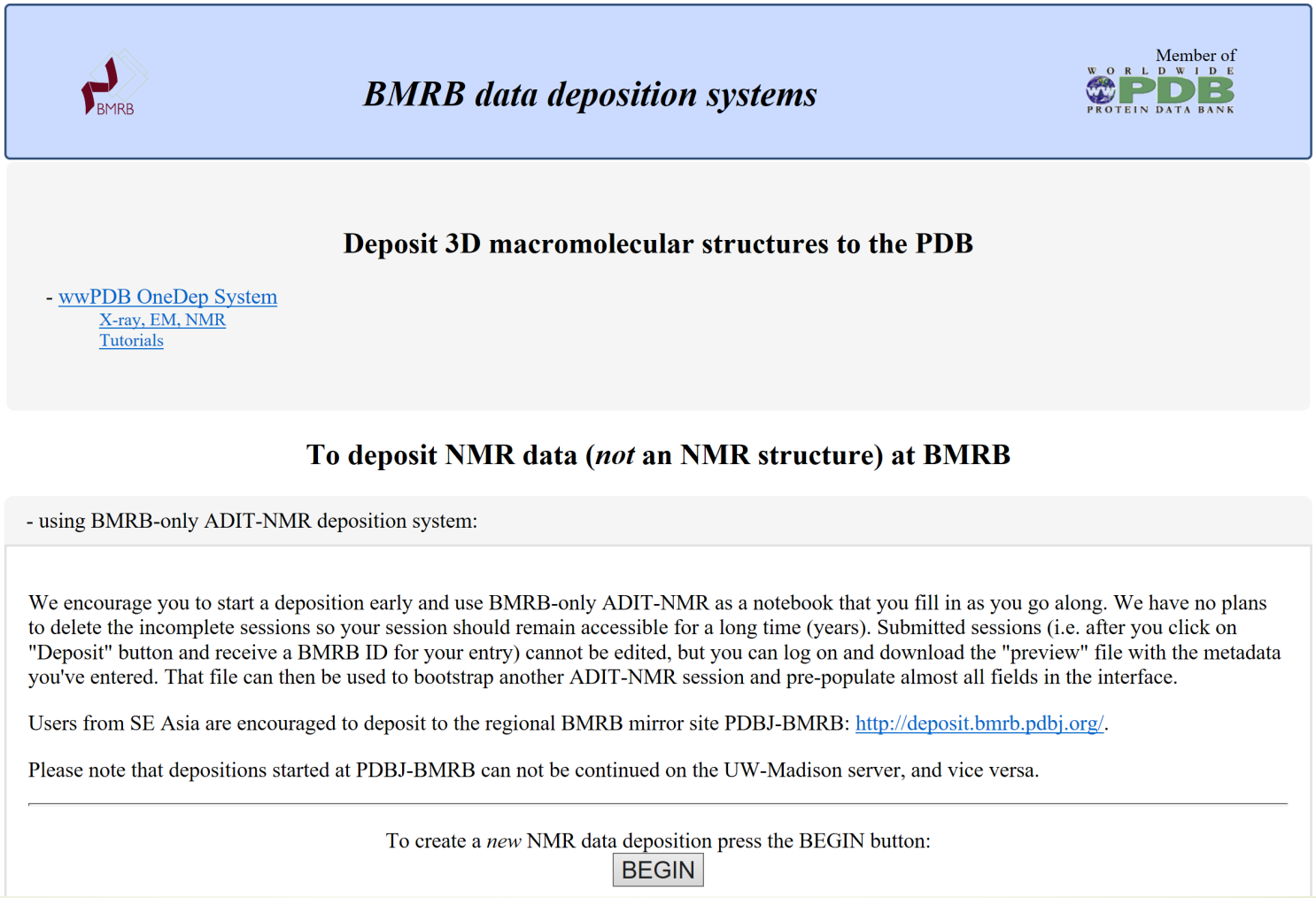
Our Request:

- Please deposit your NMR data into BMRB
- Help us make it a better resource for IDP scientists

BMRB portal (bmrb.wisc.edu)



ADIT-NMR Deposition System



Help is Always Available

- ADIT-NMR deposition tutorial
  - New easier, faster system, BMRBDep coming soon!
- Quick-response help list:
  - Send email to **bmrbhelp.bmrb.wisc.edu**
  - The entire team gets it!

BMRB Team

- John Markley – Head
- Pedro R. Romero – Director
- Eldon Ulrich – Director Emeritus
- Kumaran Baskaran – Scientist
- Hongyang Yao – Chief Curator
- Dimitri Maziuk – Senior IT and Systems Programmer
- Jon Wedell – Systems Programmer

THANKS!

We look forward to working with you!