Guided Prediction with Sparse NMR Data

Gaetano T. Montelione, Natalia Dennisova, G.V.T. Swapna,
and Janet Y. Huang, Rutgers University

Antonio Rosato
CERM, University of Florance

Homay Valafar
Univ of South Carolina

The CASP Organizing Team

CASP: Data Guided Prediction Tutorial
April 23, 2018

CASP 13: Sparse Experimental Data Guided Prediction in CASP

SAXS Data Guided Prediction

Phase 1: Real SAXS data for CASP FM Targets
Phase 2: Real SAXS data for CASP Commons Targets

NMR Data Guided Prediction

Phase 1: Simulated Sparse NMR Data for CASP FM Targets
Phase 2: Real Sparse NMR Data for CASP Commons Targets

Cross–Link Data Guided Prediction

Real Sparse NMR Data for CASP Commons Targets

SANS Data Guided Prediction

CASP Commons Targets

CASP Commons Targets

Targets and data are being generated by CASP Organizers

Proposed by high-impact biomedical research labs.

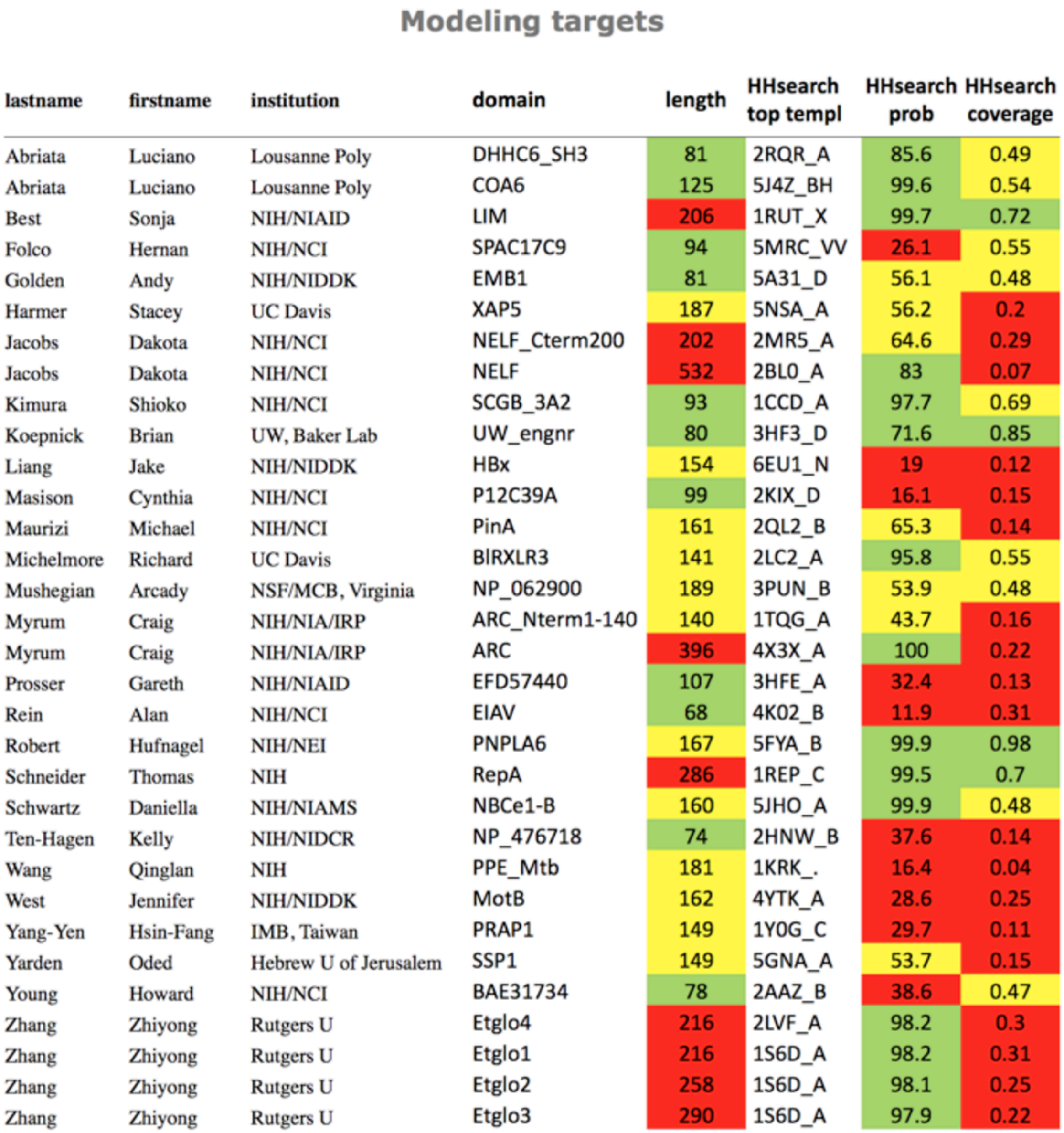
Range from 50 to 200 residues. May be monomers or oligomers.

No good templates can be identified for modeling.

Shallow multiple sequence alignments ($N_{eff} / L < 2$).

Structures to eventually be determined by CASP Organizers – may not have 3D structures by fall for assessment. Assessment will be an ongoing activity.

CASP Commons Targets



Modeling targets

lastname	firstname	institution	domain	length	HHsearch top templ	HHsearch prob	HHsearch coverage
Abriata	Luciano	Lousanne Poly	DHHC6_SH3	81	2RQR_A	85.6	0.49
Abriata	Luciano	Lousanne Poly	COA6	125	5J4Z_BH	99.6	0.54
Best	Sonja	NIH/NIAID	LIM	206	1RUT_X	99.7	0.72
Folco	Hernan	NIH/NCI	SPAC17C9	94	5MRC_VV	26.1	0.55
Golden	Andy	NIH/NIDDK	EMB1	81	5A31_D	56.1	0.48
Harmer	Stacey	UC Davis	XAP5	187	5NSA_A	56.2	0.2
Jacobs	Dakota	NIH/NCI	NELF_Cterm200	202	2MR5_A	64.6	0.29
Jacobs	Dakota	NIH/NCI	NELF	532	2BL0_A	83	0.07
Kimura	Shioko	NIH/NCI	SCGB_3A2	93	1CCD_A	97.7	0.69
Koepnick	Brian	UW, Baker Lab	UW_engnr	80	3HF3_D	71.6	0.85
Liang	Jake	NIH/NIDDK	HBx	154	6EU1_N	19	0.12
Masison	Cynthia	NIH/NCI	P12C39A	99	2KIX_D	16.1	0.15
Maurizi	Michael	NIH/NCI	PinA	161	2QL2_B	65.3	0.14
Michelmore	Richard	UC Davis	BIRXLR3	141	2LC2_A	95.8	0.55
Mushegian	Arcady	NSF/MCB, Virginia	NP_062900	189	3PUN_B	53.9	0.48
Myrum	Craig	NIH/NIA/IRP	ARC_Nterm1-140	140	1TQG_A	43.7	0.16
Myrum	Craig	NIH/NIA/IRP	ARC	396	4X3X_A	100	0.22
Prosser	Gareth	NIH/NIAID	EFD57440	107	3HFE_A	32.4	0.13
Rein	Alan	NIH/NCI	EIAV	68	4K02_B	11.9	0.31
Robert	Hufnagel	NIH/NEI	PNPLA6	167	5FYA_B	99.9	0.98
Schneider	Thomas	NIH	RepA	286	1REP_C	99.5	0.7
Schwartz	Daniella	NIH/NIAMS	NBCe1-B	160	5JHO_A	99.9	0.48
Ten-Hagen	Kelly	NIH/NIDCR	NP_476718	74	2HNW_B	37.6	0.14
Wang	Qinglan	NIH	PPE_Mtb	181	1KRK_.	16.4	0.04
West	Jennifer	NIH/NIDDK	MotB	162	4YTK_A	28.6	0.25
Yang-Yen	Hsin-Fang	IMB, Taiwan	PRAP1	149	1Y0G_C	29.7	0.11
Yarden	Oded	Hebrew U of Jerusalem	SSP1	149	5GNA_A	53.7	0.15
Young	Howard	NIH/NCI	BAE31734	78	2AAZ_B	38.6	0.47
Zhang	Zhiyong	Rutgers U	Etglo4	216	2LVF_A	98.2	0.3
Zhang	Zhiyong	Rutgers U	Etglo1	216	1S6D_A	98.2	0.31
Zhang	Zhiyong	Rutgers U	Etglo2	258	1S6D_A	98.1	0.25
Zhang	Zhiyong	Rutgers U	Etglo3	290	1S6D_A	97.9	0.22

Proposal for CASP11 Contact Assisted Prediction

Contacts could be experimentally accessible distances:

- chemical cross links (Mass Spec)
- backbone NH – NH and or ILV Me-Me contacts (< 6.5 Å, ^{2}H proteins)
- Paramagnetic Relaxation Enhancement (PRE) (15 – 30 Å)

Methods will be developed that use realistic types of contacts that can potentially be obtained on larger (20 – 80 kDa) proteins

CASP project will drive the experimental community to generate such contact data and to collaborate with CASP methods developers on specific projects

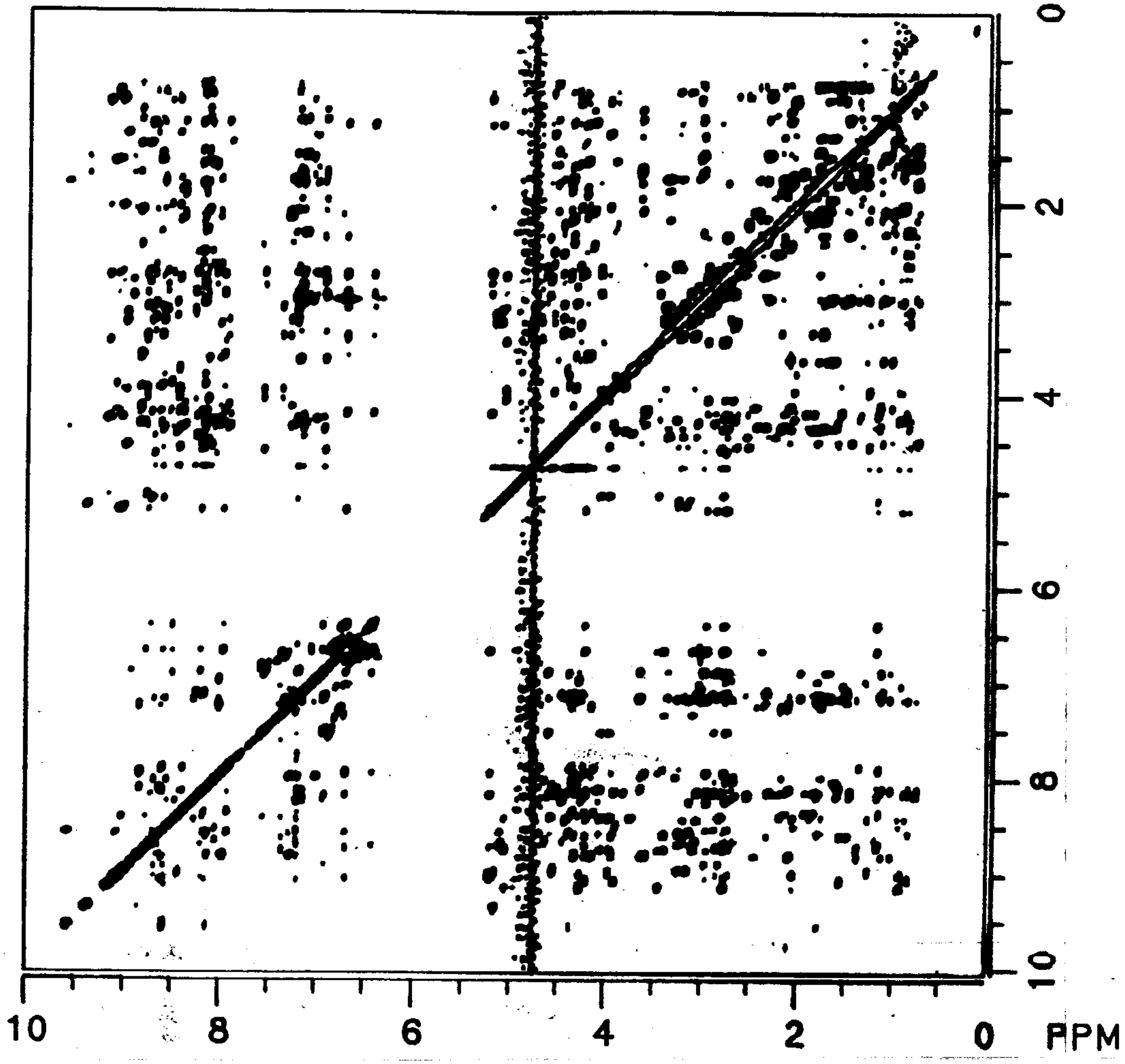
CASP 11: Experimentally Feasible Contacts

Scientific Premise: Structure prediction will be more accurate if some native contacts are known.

Rather than providing ”most missed contacts” we provided “experimentally accessible” simulated contacts.

- Contacts based on ambiguous contacts derived from simulated NOESY spectra for 19 FM targets (Montelione Lab).
- Real cross-link data for 4 CASP FM targets (Rappsilber Lab).

2D NOESY Spectrum of a Protein



Acc. Chem. Res., Vol. 22, No. 1, 1989 **Wüthrich**

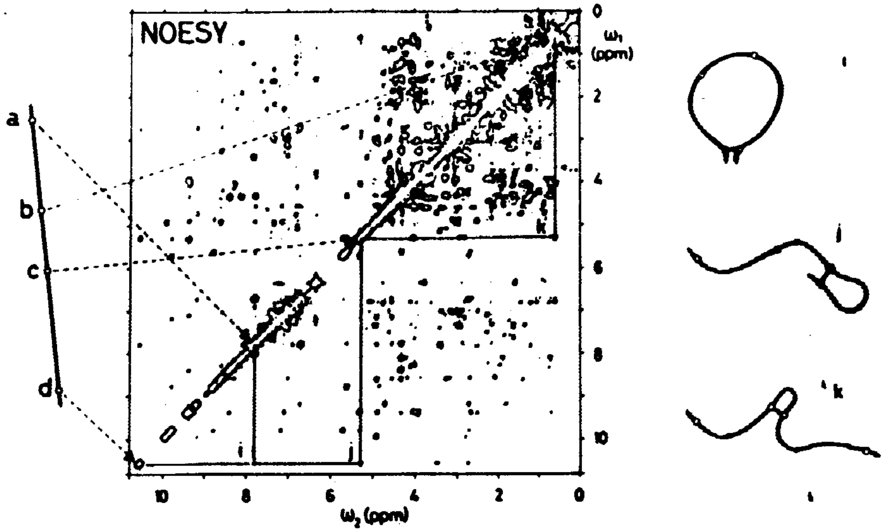


Figure 4. Illustration of the description of the NMR method for protein structure determination in solution. In the center, a contour plot of a 500-MHz ^{1}H NOESY spectrum of the protein basic pancreatic trypsin inhibitor (BPTI) is shown, with the two frequency axes ω_1 and ω_2. Three cross peaks are marked i–k and linked by horizontal and vertical lines with the diagonal positions of the protons connected by the corresponding NOEs. On the left, an extended polypeptide chain is represented by a straight line, and four protons in this chain are identified by circles and the letters a–d. The broken arrows connect these protons with their resonance positions on the diagonal of the NOESY spectrum. On the right, there is a schematic representation of three circular structures formed by the polypeptide chain, which are manifested by the cross peaks i–k.

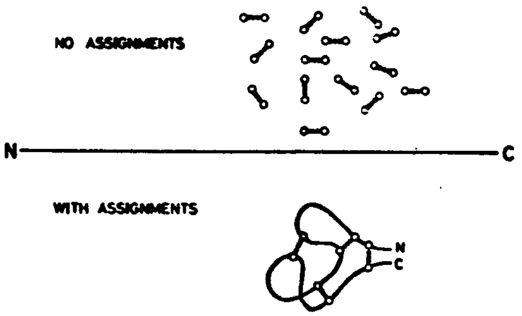
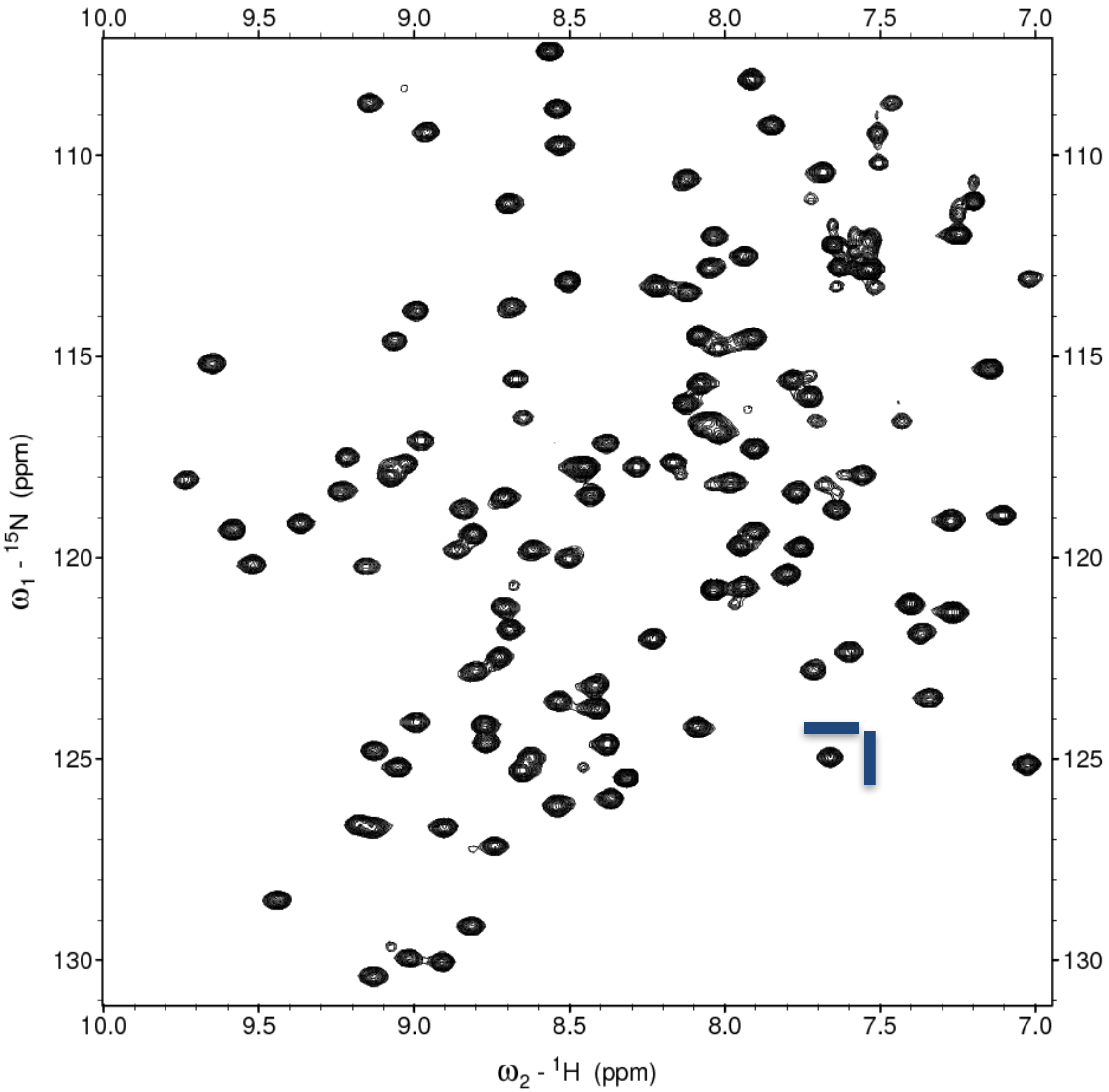


Figure 1.1. Information content of ^{1}H–^{1}H NOE's in a polypeptide chain with and without sequence-specific resonance assignments. Open circles represent hydrogen atoms of the polypeptide. The polypeptide chain is represented by the horizontal line in the center.

CASP: Data Guided Prediction Tutorial
April 23, 2018

The Ambiguity Problem in Analysis in Cross Peak Assignment



In NOESY

For a given cross peak, the Y-axis will, in general, match, within a “match tolerance”, to Y possible resonances assignments.

The X-axis will, in general, match, within a “match tolerance”, to X possible resonance assignments.

Hence – the NOESY cross peak may arise from any one (or more) of X * Y short (< 5 Å) distance interactions

CASP: Data Guided Prediction Tutorial
April 23, 2018

Ambiguous NOE-based Contact List for CASP11

(H^N-H^N, H^N-Me, Me-Me ^1H-^1H Contacts)

Residue 1	Residue 2	Peak No.	Upper-bound		Atom 1	Atom 2	
R1	**R2**	**P#**	**UPL**	**Confid**	**A1**	**A2**	
79	77	17	5.0	0.95	H	H	Peak 17
79	177	20	6.0	0.67	H	HD2	Peak 20
79	135	20	6.0	0.97	H	HD1	
79	249	20	6.0	0.96	H	HD1	
79	50	20	6.0	0.81	H	HD2	
79	217	23	5.0	0.68	H	H	Peak 23
79	230	23	5.0	0.75	H	H	
79	232	23	5.0	0.72	H	H	
79	106	23	5.0	0.76	H	H	
79	166	23	5.0	0.83	H	H	
79	100	23	5.0	0.83	H	H	
79	82	23	5.0	0.74	H	H	
79	246	23	5.0	0.71	H	H	
79	216	23	5.0	0.67	H	H	
45	37	28	7.5	0.84	HD2	HG1	Peak 28

19 CASP11-NMR Targets

CASP ID	PDB ID	Fold	#Residues with simulated CS	#ILVA	# Peaks	Avg Ambiguity Per NOESY Peak	Max Ambiguity
Ts761	4PW1	alpha+beta	214	51	3106*	9	70
Ts763	4Q0Y	alpha+beta	130	35	2029*	6	36
Ts767	4QPV	alpha+beta	274	58	1564	9	64
Ts777		alpha	345	101	2400	18	144
Ts785	4D0V	most beta	112	33	694	6	45
Ts794	4CYF	alpha+beta	462	124	3132	27	232
Ts800	4QRK	beta	212	60	1459	14	96
Ts802		beta	118	32	530	4	21
Ts804		beta	194	43	884	9	95
Ts806		alpha+beta	256	74	1791	15	136
Ts810		alpha	113	30	739	5	39
Ts812		alpha+beta	183	53	980	6	45
Ts814	4R7F	beta	397	90	2290	18	168
Ts818	4R1K	alpha+beta	134	23	516	4	21
Ts824		alpha+beta	108	27	522	3	23
Ts826		alpha	201	85	1666	14	145
Ts827		alpha	150	51	1091	8	61
Ts832	4RD8	alpha	209	56	1472	12	75
Ts835		most alpha	404	135	3517	22	223

* Distance cutoff of 6.5 ang were used for T0761 and T0763. Distance cutoff of 5 ang were used for all other targets

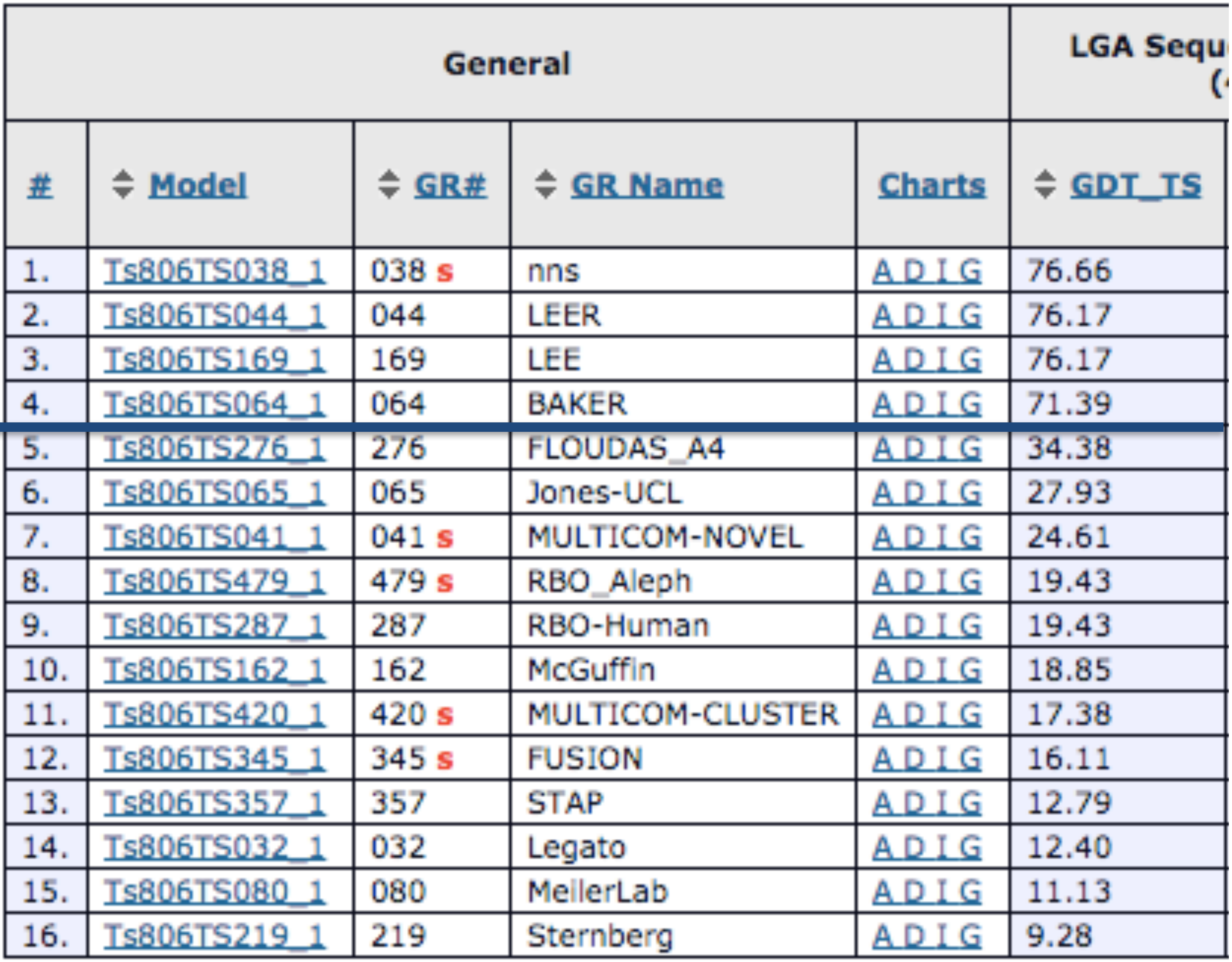
CASP 11: Experimentally Feasible Contacts

Assessment of CASP11 Contact-Assisted Predictions

L. N. Kinch, W. Li, B. Monastyrskyy, A. Kryshtafovycnh, and N. V. Grishin

Proteins (2016) 84: 164–180. doi:10.1002/prot.25020.

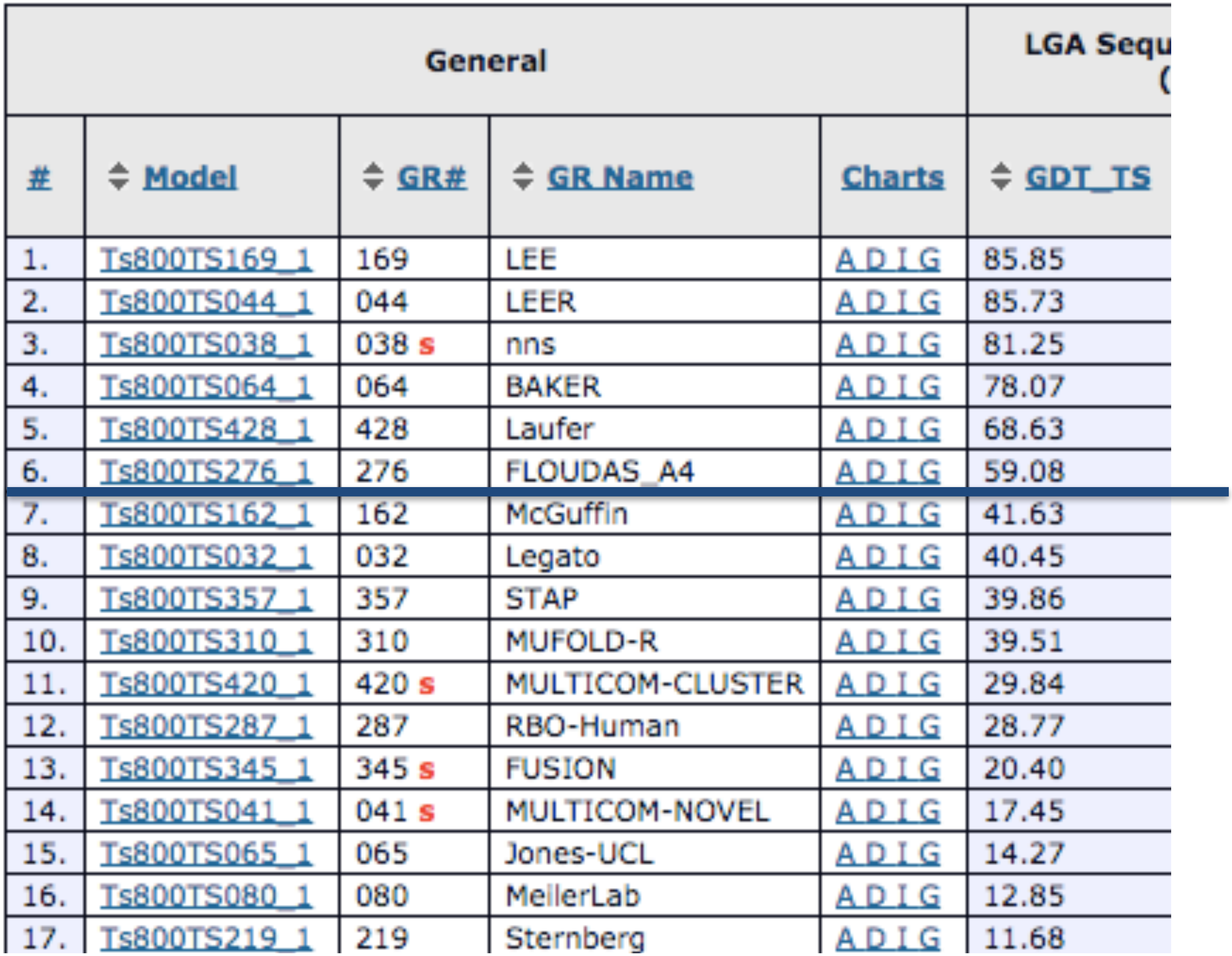
Some 'Predictors' did better than standard ASDP NMR Methods



General					LGA Sequ
#	Model	GR#	GR Name	Charts	GDT_TS
1.	Ts806TS038_1	038 s	nns	A D I G	76.66
2.	Ts806TS044_1	044	LEER	A D I G	76.17
3.	Ts806TS169_1	169	LEE	A D I G	76.17
4.	Ts806TS064_1	064	BAKER	A D I G	71.39
5.	Ts806TS276_1	276	FLOUDAS_A4	A D I G	34.38
6.	Ts806TS065_1	065	Jones-UCL	A D I G	27.93
7.	Ts806TS041_1	041 s	MULTICOM-NOVEL	A D I G	24.61
8.	Ts806TS479_1	479 s	RBO_Aleph	A D I G	19.43
9.	Ts806TS287_1	287	RBO-Human	A D I G	19.43
10.	Ts806TS162_1	162	McGuffin	A D I G	18.85
11.	Ts806TS420_1	420 s	MULTICOM-CLUSTER	A D I G	17.38
12.	Ts806TS345_1	345 s	FUSION	A D I G	16.11
13.	Ts806TS357_1	357	STAP	A D I G	12.79
14.	Ts806TS032_1	032	Legato	A D I G	12.40
15.	Ts806TS080_1	080	MeilerLab	A D I G	11.13
16.	Ts806TS219_1	219	Sternberg	A D I G	9.28

using
ASDP
GDT
0.61

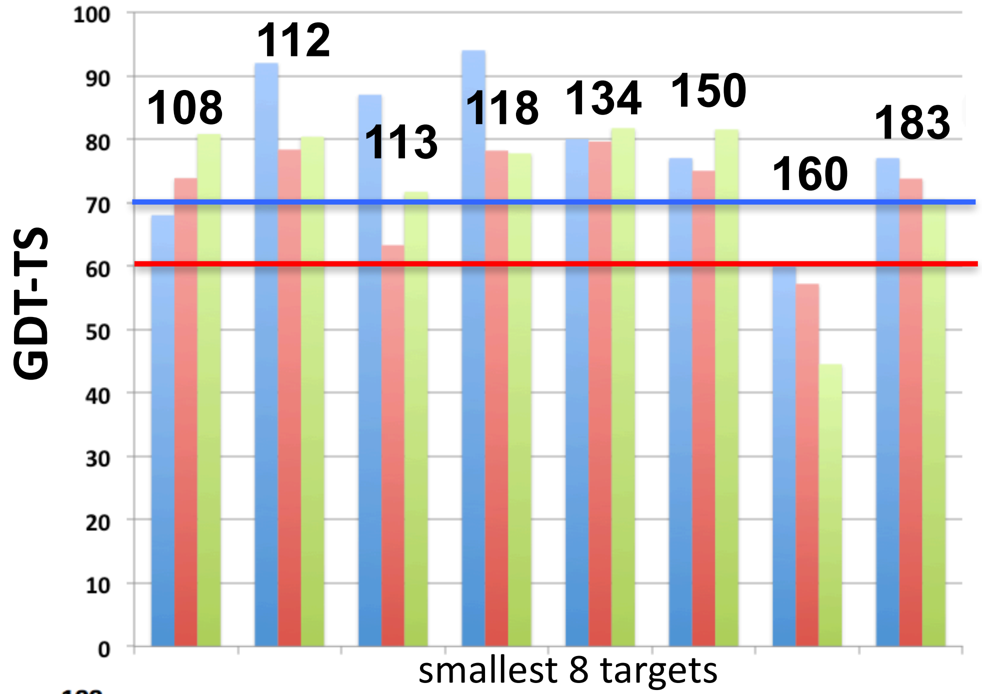
Some ‘Predictors’ did better than standard ASDP NMR Methods

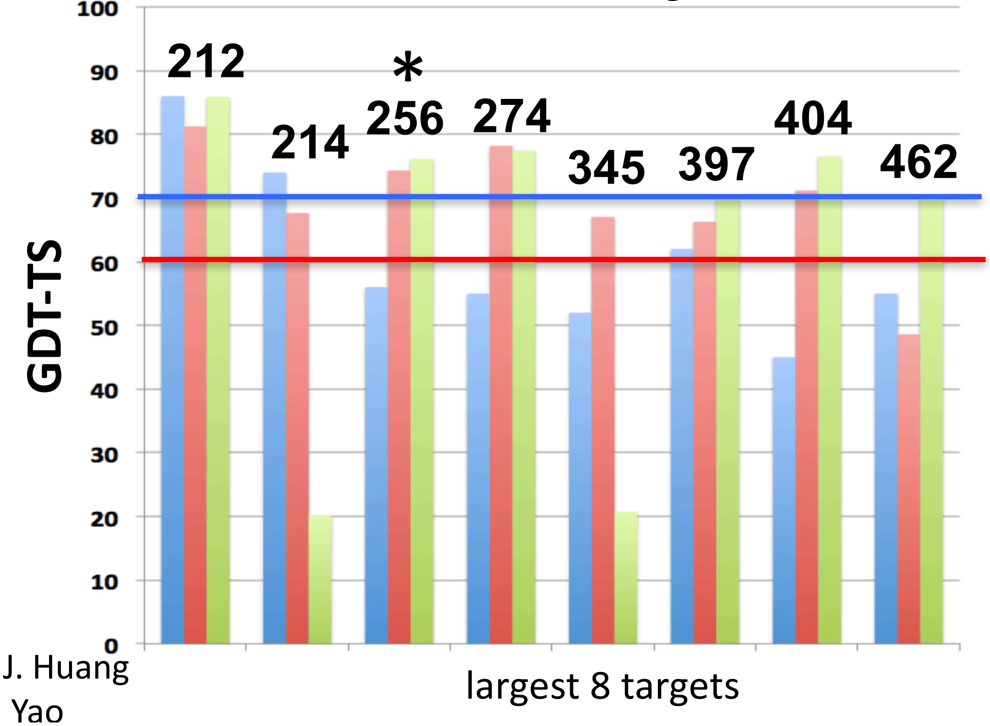


General					LGA Sequ
#	Model	GR#	GR Name	Charts	GDT_TS
1.	Ts800TS169_1	169	LEE	A D I G	85.85
2.	Ts800TS044_1	044	LEER	A D I G	85.73
3.	Ts800TS038_1	038 s	nns	A D I G	81.25
4.	Ts800TS064_1	064	BAKER	A D I G	78.07
5.	Ts800TS428_1	428	Laufer	A D I G	68.63
6.	Ts800TS276_1	276	FLOUDAS_A4	A D I G	59.08
7.	Ts800TS162_1	162	McGuffin	A D I G	41.63
8.	Ts800TS032_1	032	Legato	A D I G	40.45
9.	Ts800TS357_1	357	STAP	A D I G	39.86
10.	Ts800TS310_1	310	MUFOLD-R	A D I G	39.51
11.	Ts800TS420_1	420 s	MULTICOM-CLUSTER	A D I G	29.84
12.	Ts800TS287_1	287	RBO-Human	A D I G	28.77
13.	Ts800TS345_1	345 s	FUSION	A D I G	20.40
14.	Ts800TS041_1	041 s	MULTICOM-NOVEL	A D I G	17.45
15.	Ts800TS065_1	065	Jones-UCL	A D I G	14.27
16.	Ts800TS080_1	080	MeilerLab	A D I G	12.85
17.	Ts800TS219_1	219	Sternberg	A D I G	11.68

Using
ASDP
GDT
0.58

“Conventional“ ASDP Automated Analysis vs “CASP11 Contact-Assisted Prediction”





Blue – ASDP (conventional)
Brown – Baker iterative Rosetta-CM
Green – Lee – CSA-NMR

For smallest targets (108 – 183 residues), ASDP method generally performs as well or better than the best CASP11 “contact-assisted prediction” methods.

For largest targets (250 – 462 residues), the best CASP11 “contact-assisted prediction” methods generally provide correct structures even for cases where our ASDP method fails.

ASDP fails on largest proteins (> 250 residues) because the ambiguity in NOESY peak assignment becomes too high.

Y.J. Huang
Y. Yao

Shortcomings of CASP11 NMR-based Simulated Contacts

- How realistic are the simulated NMR data?
  - Missing resonance assignments? Data completeness?
  - Noise in NOESY peak list?
  - Can we have real data?

- If Assignments are available – why were not the Assignments provided to predictors.

BB ASSIGNMENTS -> BACKBONE DIHEDRAL ANGLES

- Sparse NMR data would also generally include Residual Dipolar Coupling (RDC) Data.

NMR Data for Structure Calculation

- Small proteins (< 20 kDa)
  - Chemical Shift data (Backbone dihedral angles)
  - NOE (long range distance restraints)
  - Residual dipolar coupling (RDC) data

- Large proteins (20 – 80 kDa)
  - Chemical Shift data (Backbone dihedral angles)
  - Sparse NOEs involving HN and ILVA Me protons on perdeuterated proteins (long range distance restraints ~ 6 Å)
  - RDC's (backbone orientation)
  - PRE (long range distance restraints ~30 Å)

Simulated and Real Sparse NMR Data To Be Provided for CASP13

Ambiguous NOE-based Atom-Atom Contacts
(100 to 800 residue proteins)

Backbone dihedral restraints (incomplete, uncertainty as derivable from chemical shift data)

Backbone N-H residual dipolar couplings (RDCs)
(incomplete, uncertainty based on typical error estimates)

CASP13 Sparse NMR Guided Prediction

Phase 1: Simulated Sparse NMR Data

- FM CASP targets will be used to simulate backbone chemical shift and ILVA Methyl assignments using ShiftX
- 30% of these assignments will be deleted
- Backbone phi-psi dihedral ranges for remaining assigned resonances will be computed using Talos software (however, chemical shifts will not be provided)
- Pseudo 4D NOESY spectra (3D spectra) will be simulated assuming realistic line widths and signal-to-noise; random noise peaks will be added to the spectra.
- ASDP software will be used to determine possible assignments for each cross peak -> ambiguous assignment list.
- N-H RDCs will be computed from X-ray crystal structure for assigned backbone N-H bonds.
- Oligomerization state will be provided.

CASP13 Sparse NMR Guided Prediction

Phase 2: Real NMR Data

- CASP Commons Protein Targets
  - 80 – 180 residues
  - solicited from biomedical research community
  - no good templates; shallow sequence alignments
- 14 targets selected for sample production
  - Gene synthesis, expression, purification
  - Oligomer state by Analytical Gel Filtration with Static Light Scattering
  - ^{15}N-HSQC spectrum
  - All expression plasmids will be provided to LBL for SAXS
- 2-4 samples will then be ^{13}C,^{15}N isotope-enriched for NMR studies
- Data to be provided to predictors
  - Backbone resonance assignments
  - Backbone dihedral angles from Talos
  - Ambiguous contacts from ^{15}N-edited NOESY (no RDC data).
- Reference structures will be completed by NMR and/or X-ray crystallography

Future CASP Challenges

Ongoing process of generating CASP Commons Targets, Data and Structure

Modeling Multiple Conformational States

Modeling Using Unassigned NOESY spectra

Modeling Using Unassigned RDC data

End of April 23 Presentation

The Following Slides are for More a Detailed Presentation to be Recorded and Posted on CASP Web Site

Some Background on How Current Data Guided Prediction Program Differs from Original Goals of CASP11 in Contact Assisted Prediction

Proposal for “Rational” Contacts for Future Contact Assisted CASP Experiments

James Aramini, Janet Y. Huang,
Gaetano Montelione

CASP10 Meeting
12/11/12

Proposal

Contacts could be experimentally accessible distances:

- chemical cross links (Mass Spec)
- backbone NH – NH and or ILV
  Me-Me contacts (< 6.5 Å, ^{2}H proteins)
- Paramagnetic Relaxation Enhancement (PRE) (15 – 30 Å)

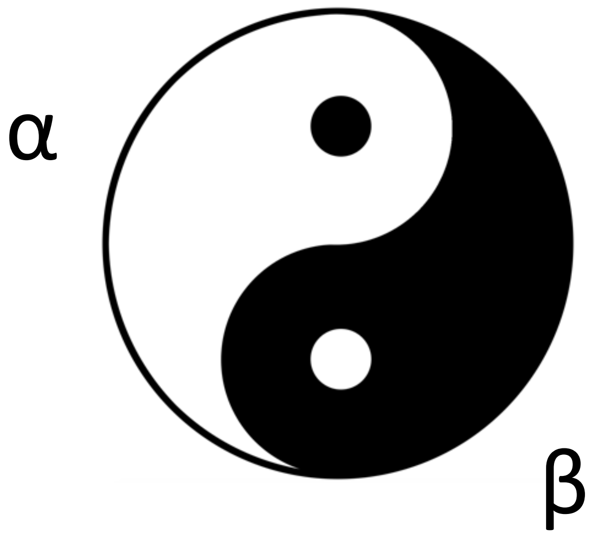
Methods will be developed that use realistic types of contacts that can potentially be obtained on larger (20 – 80 kDa) proteins

CASP project will drive the experimental community to generate such contact data and to collaborate with CASP methods developers on specific projects

Obvious Contact Restraints

- Mass Spec
  - Cross-linkable Lys amino - Lys amino distances
  - Other cross-linkable distances

- NMR
  - NH – NH short distances
  - NH – Me and Me – Me distances



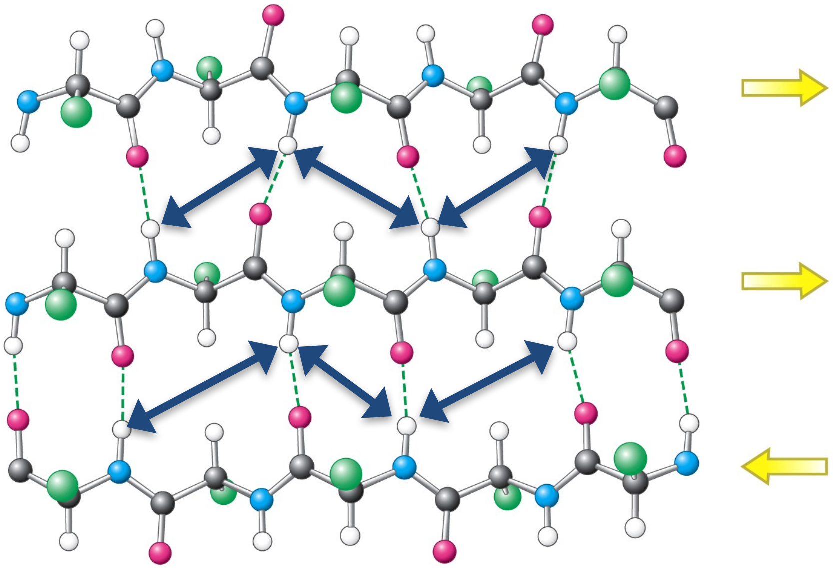


Figure 2.33

CASP 11: Experimentally Feasible Contacts

Scientific Premise: Structure prediction will be more accurate if some native contacts are known.

Rather than providing ”most missed contacts” we can provide “experimentally accessible” simulated contacts.

- Contacts based on ambiguous contacts derived from simulated NOESY spectra for 19 FM targets (Montelione Lab).
- Real cross-link data for 4 CASP FM targets (Rappsilber Lab).

The idea of “more realistic contacts based on what can be obtained by experiments” has been superseded by the advances since 2012 in contact prediction from sequence co-variance analysis.

These EC methods provide reliable contact predictions for deep sequence alignments with $N_{eff} / L > 5$.

No need to have a CASP category for how well modelers can do with “simulated contacts”.

CASP 12: Sparse Experimental Data Guided Prediction in CASP

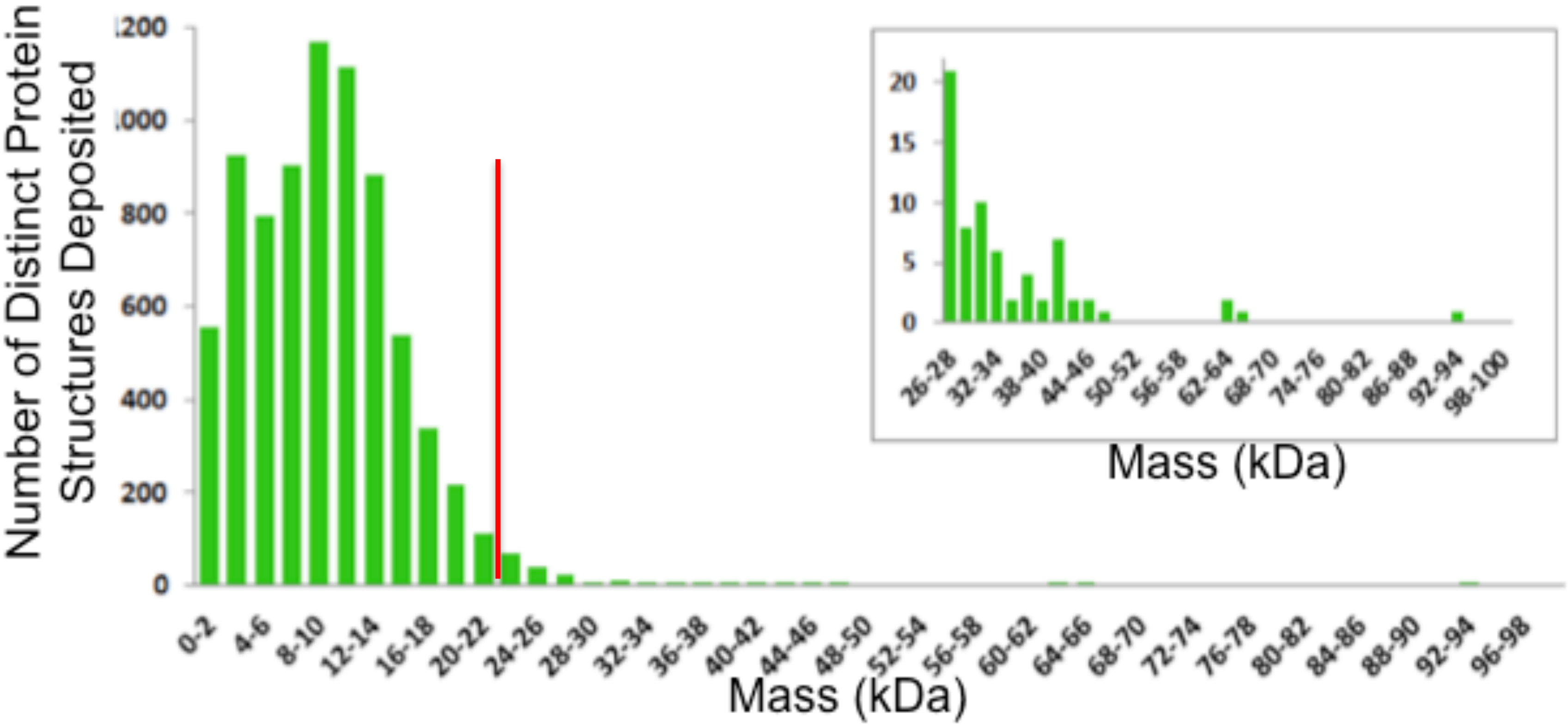
Hypothesis to Test: Structure prediction can be guided by real "sparse" experimental data.

Focus more on "real data"; connect with real issues in experimental data collection and analysis. Hybrid method of structure determination building on prediction methods.

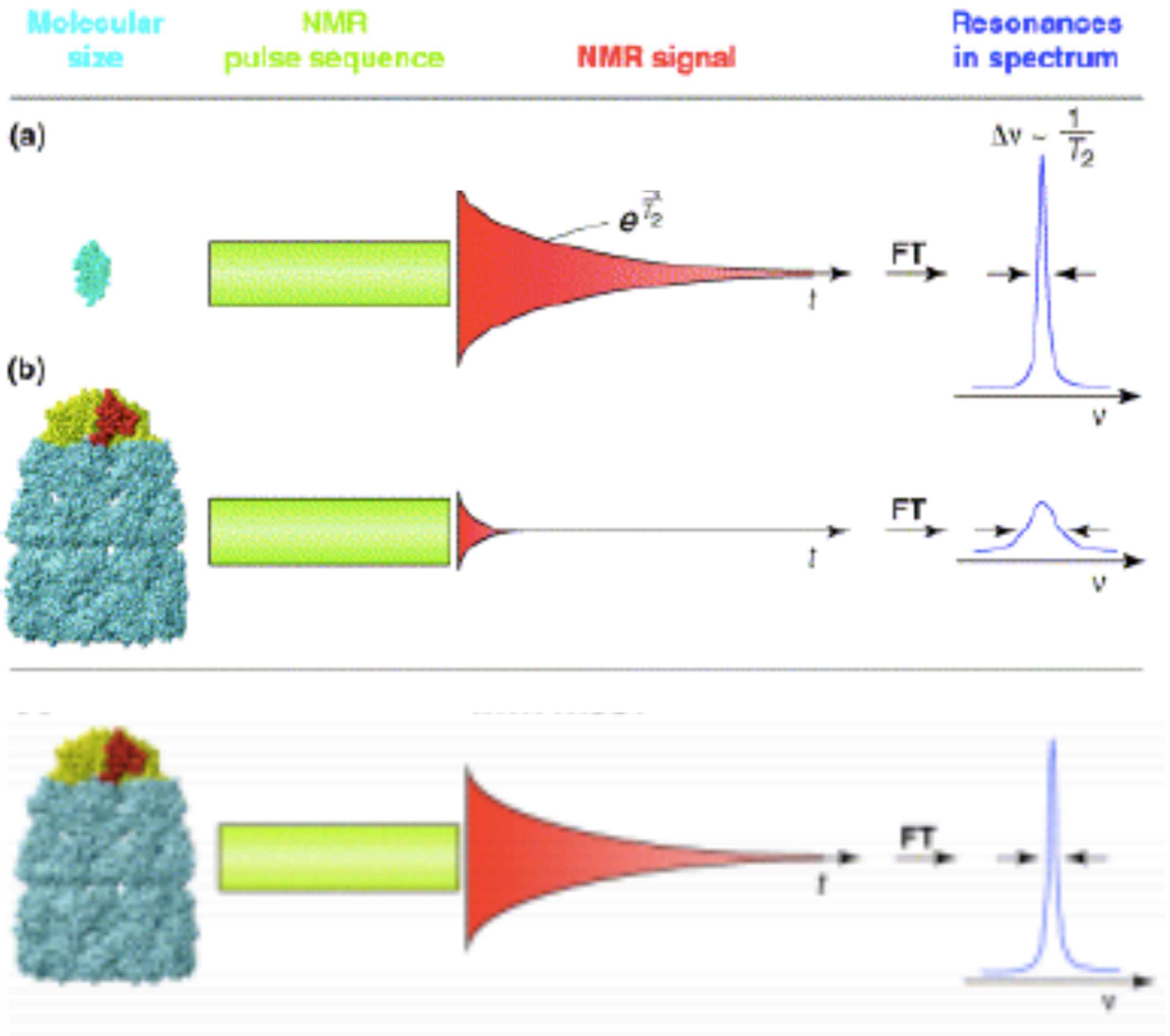
- Cross-link data for CASP FM structures
- SAXS curves for CASP FM structures
- No NMR data in CASP 12

Motivation and Challenges for Protein Structure Determination from Sparse NMR Data

Distribution of the Sizes of Protein Structures Determined by NMR Methods and Deposited in the Protein Data Bank



Perdeuteration Reduces Transverse Relaxation Rates (R_2) of Remaining ^{1}H, ^{13}C, and ^{15}N nuclei

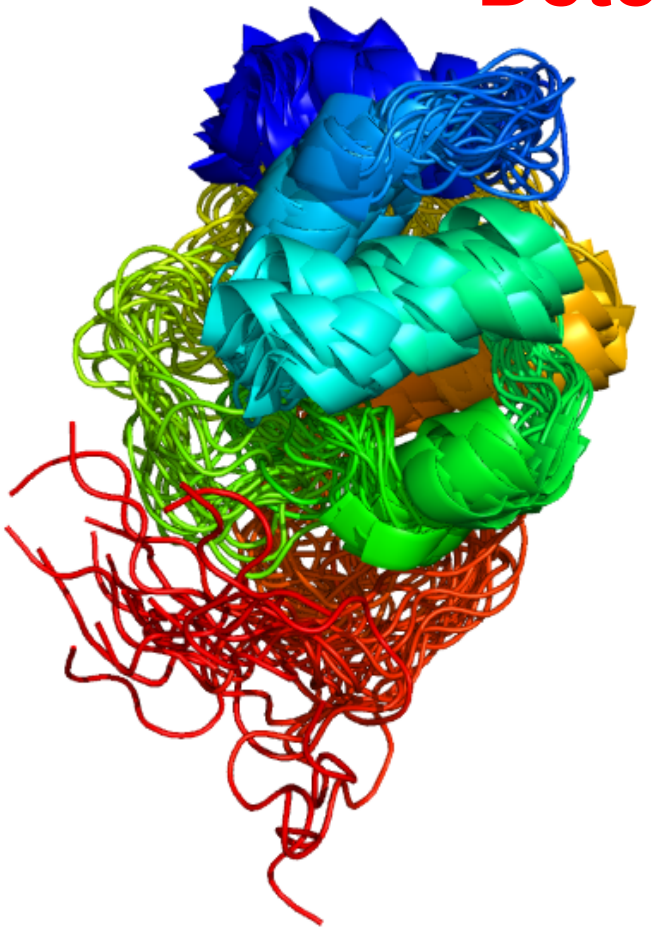


^{1}H -> ^{2}H **Back exchange HN, Label CH_3**

adopted from M. Girvin

CASP: Data Guided Prediction Tutorial
April 23, 2018

Solution NMR Structure of the N-terminal Cytosolic Domain of MEC-4 in Detergent Micelles



Xplor refined

MEC-4 N, 110 res
in detergent micelles

Can we combine "sparse NMR data" obtained on perdeuterated proteins with advanced molecular modeling methods to improve precision and accuracy?

J. Everett, M. Driscoll

Information on How the Sparse NMR Data are Generated

Sparse NMR Data

Ambiguous NOE-based atom-atom contacts

Chemical shift -> backbone dihedral restraints

Backbone N-H residual dipolar couplings (RDCs)

Pseudo 4D NOESY

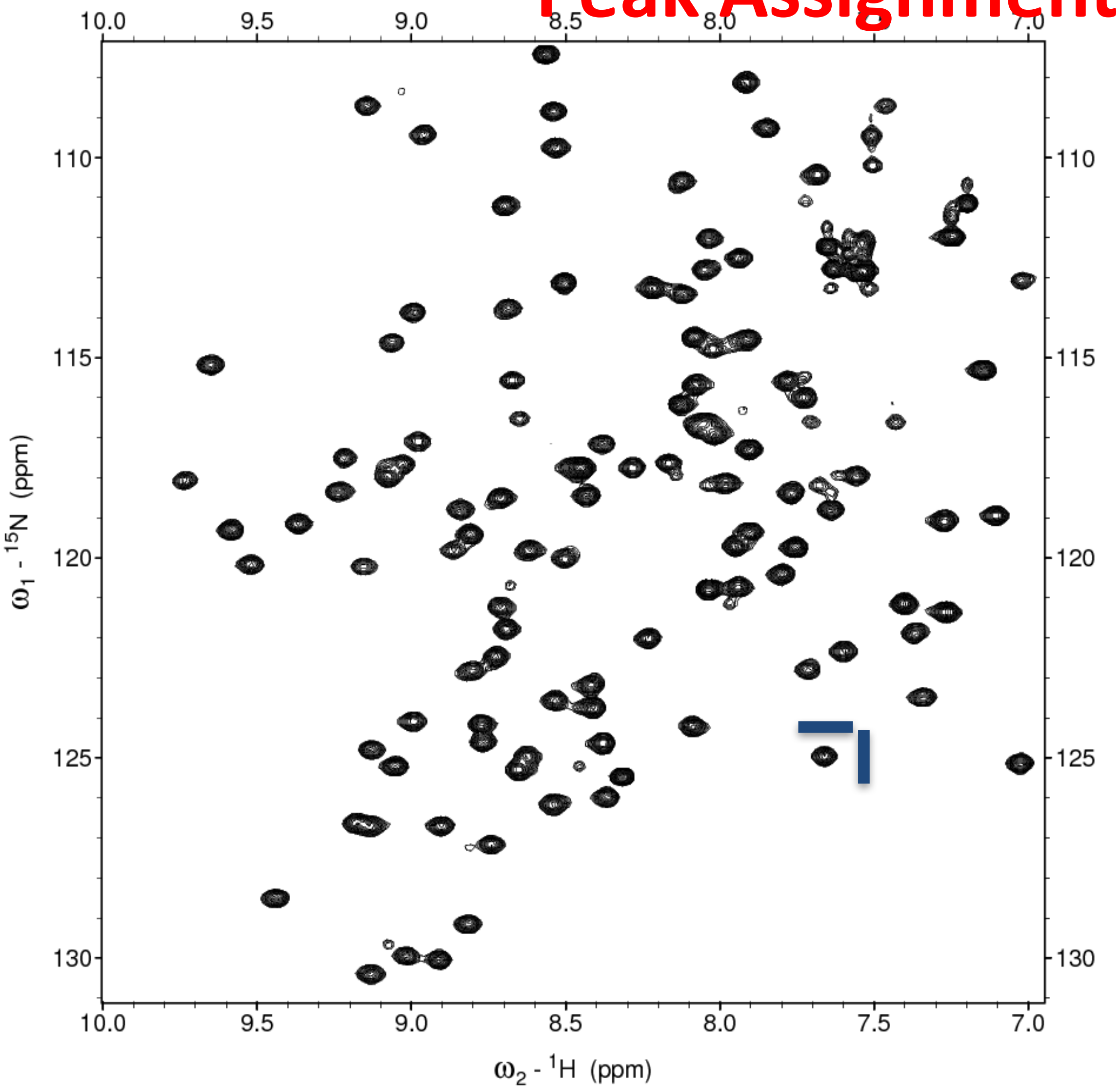
Consider 4D NOESY Experiments

- 4D HNNH-NOESY
- 4D HCCH-NOESY
- 4D HCNH-NOESY

For practical reasons, we use Pseudo 3D NOESY

- 3D (h)NNH-NOESY
- 3D (h)CCH-NOESY
- 3D (h)CNH-NOESY
- 3D (h)NCH-NOESY

The Ambiguity Problem in Analysis in Cross Peak Assignment



In NOESY

For a given cross peak, the Y-axis will, in general, match, within a “match tolerance”, to Y possible resonances assignments.

The X-axis will, in general, match, within a “match tolerance”, to X possible resonance assignments.

Hence – the NOESY cross peak may arise from any one (or more) of X * Y short (< 5 Å) distance interactions

Backbone Dihedral Restraints Can be Estimated from Backbone Chemical Shift Values

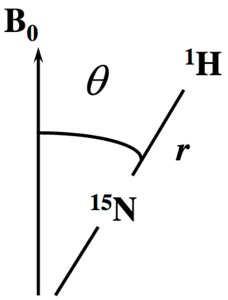
$^{13}C\alpha$ / $^{13}C\beta$ chemical shifts → backbone dihedral ranges (+/- 30 deg)

Y. Shen, A. Bax. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J. Biomol. NMR, 56, 227-241(2013)

https://spin.niddk.nih.gov/bax/software/TALOS-N/

Residual Dipolar Couplings

Provide Information about Bond Vector Orientations



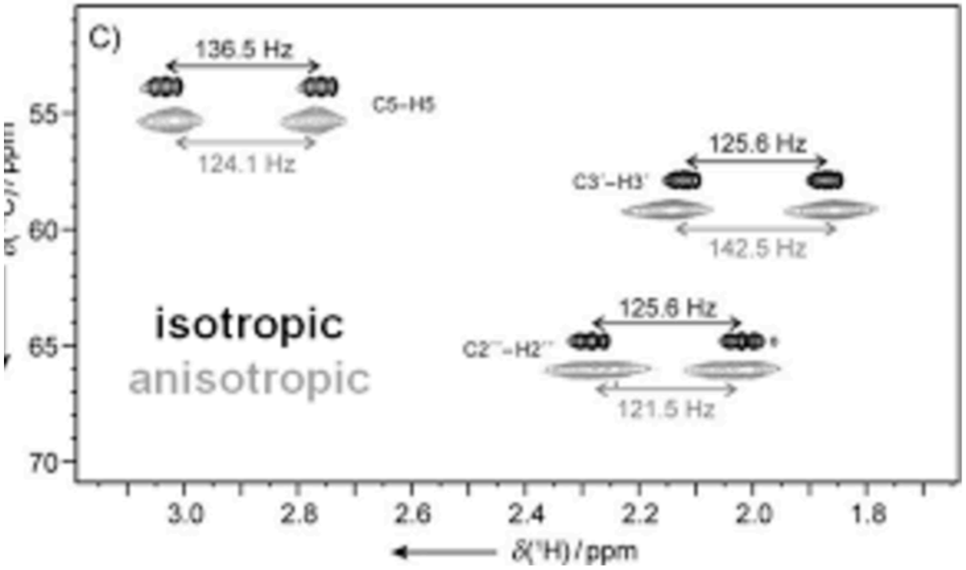
$$D = \frac{C}{r^3}\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle I_{NZ} I_{HZ}$$

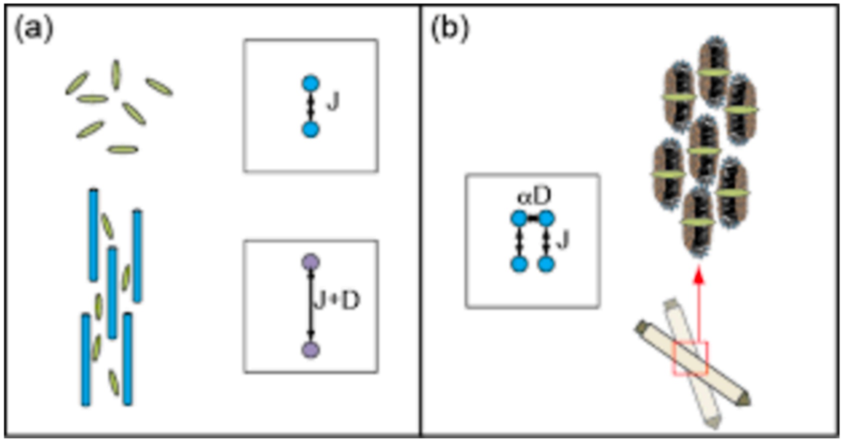
Brackets denote averaging –
goes to zero without partial orientation

Prestegard, A-Hashimi & Tolman, Quart. Reviews Biophys. 33, 371-424 (2000)

Bax, Kontaxis & Tjandra, Methods in Enzymology 339 , 127-174 (2001)

Prestegard, Bougault & Kishore, Chemical Reviews, 104 , 3519-3540 (2004)





Case Study: Restrained RASREC Rosetta for de novo Structure Generation of Larger (> 20 kDa) Proteins

H^N, ILV(A)-Methyl

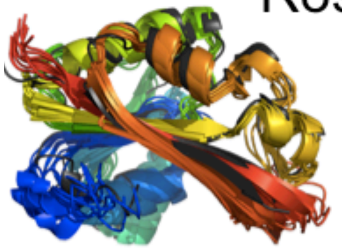
Pseudo 4D NOESY spectra

3D ^{13}C or ^{15}N edited HSQC-NOESY- HSQC spectra

run as
(h)NNH-NOESY
(h)CCH-NOESY
(h)CNH-NOESY
(h)NCH-NOESY

All raw data, input, and output files released on-line to encourage methods development

Restrained-RDC-CS-Rosetta



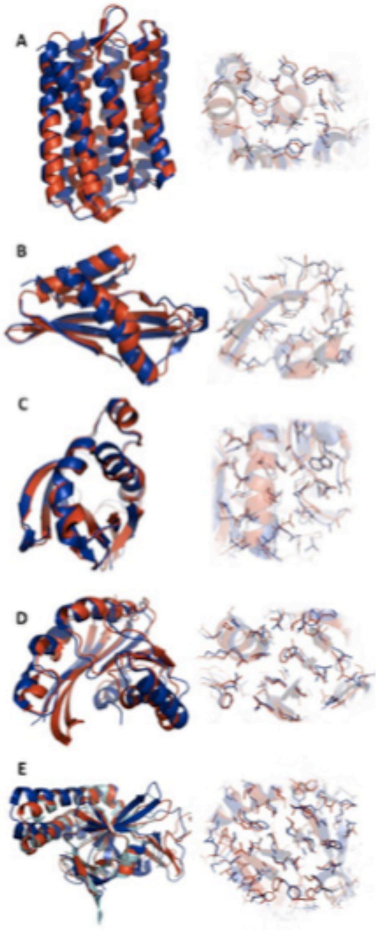
StR145 194-residue Monomer
Could not be solved by standard perdeuterated NMR methods

core 1.5 Å to Xtal structure

SgR145 P74712

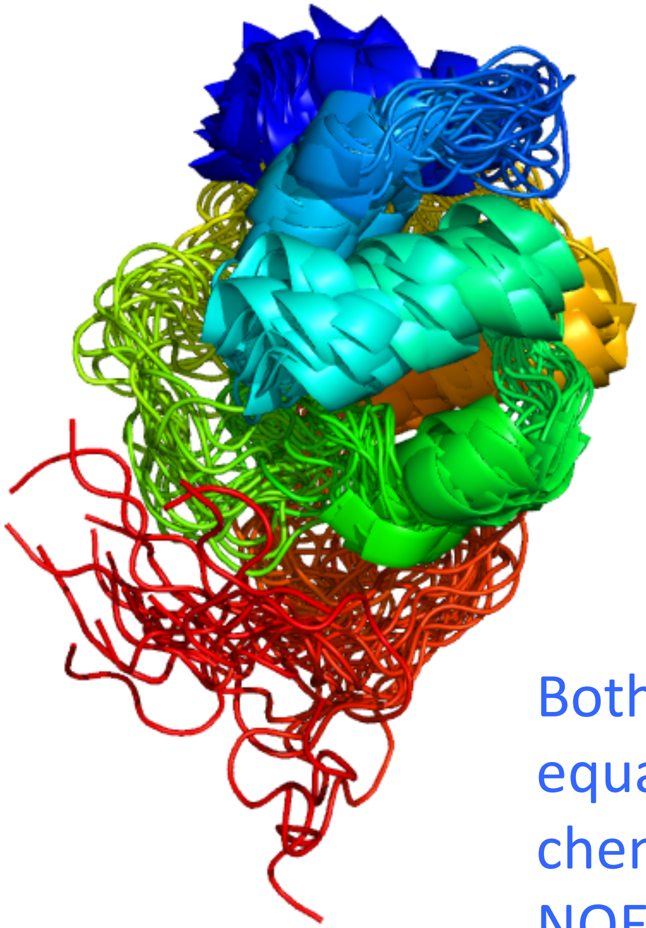
The determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples

Lange, et al **Proc. Natl. Acad. Sci. U.S.A** 2012 109: 10875



Raman et al **Science** 2010, 327 : 1014.

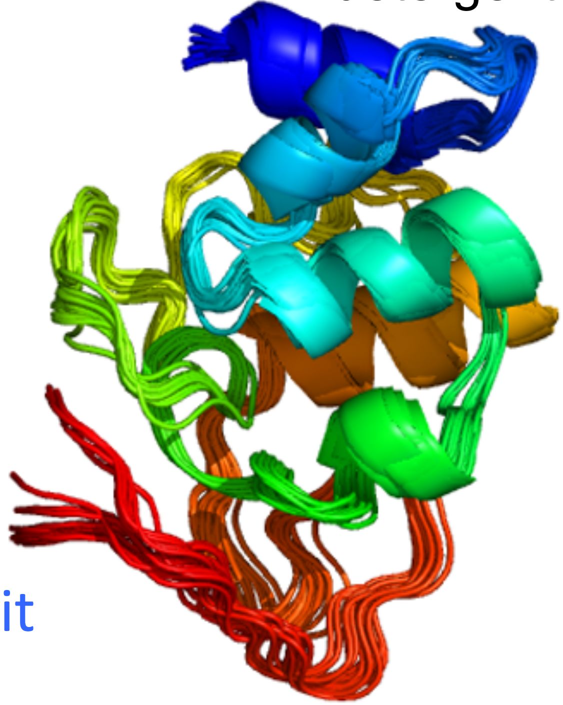
Solution NMR Structure of the N-terminal Cytosolic Domain of MEC-4 in Detergent Micelles



CNS refined

Both structures fit equally well to chemical shift, NOE, and RDC data

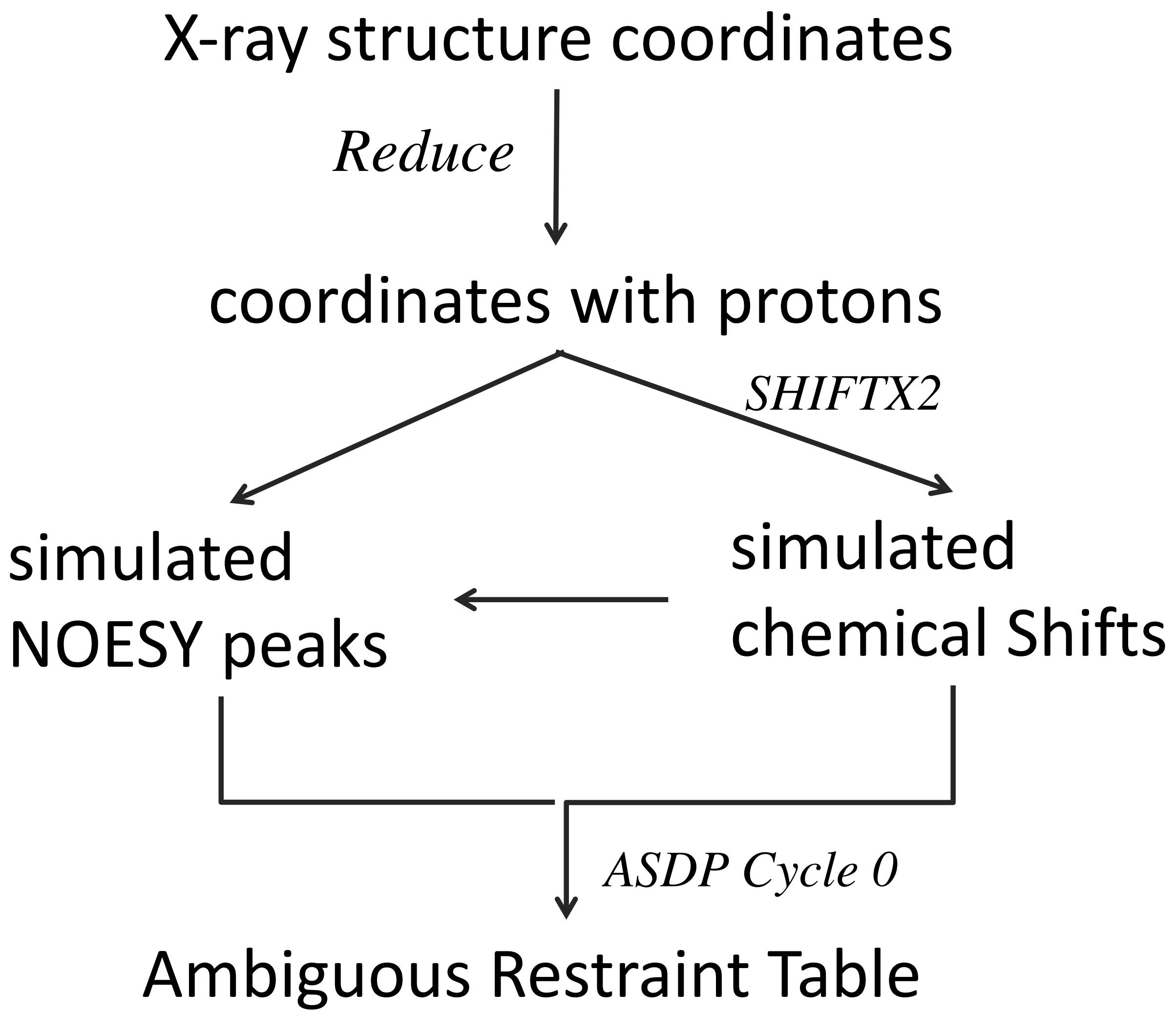
MEC-4 N, 110 res
in detergent micelles



Rosetta “refined”

J. Everett, M. Driscoll

Information on how Ambiguous NOE-Based Contacts and RDCs are Simulated from X-ray Crystal Structures of CASP Targets for CASP13



X-ray structure coordinates

Reduce

coordinates with protons

SHIFTX2

simulated
NOESY peaks

simulated
chemical Shifts

ASDP Cycle 0

Ambiguous Restraint Table

Predict of Chemical Shift

- X-ray structure of CASP target
  - Convert MSE to MET
  - Add H using Richardson *reduce* program.

    Word, et. al. (1999) Asparagine and Glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. 285, 1733-1747.

- Use Wishart *SHIFTX* program to predict chemical shifts from X-ray structure

    S. Neal, A. M. Nip, H. Zhang, D. S. Wishart (2003) "Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts" Journal of Biomolecular NMR, 26:215-240.

- Generate Chemical Shift Assignment table in BioMagResDB (bmrb format)
  - Backbone HN and N, methyl protons from ILVA residues
  - Carbons, including CA, CB, CO and ILVA methyl carbons.

Simulate 3D NOE Peak List

- Input: Completed Chemical Shift Assignment table and PDB file
- For All H-H pairs, if summation distance < 5 Å, simulate NOE cross peaks.
  - Intensity = 10000*distance^{-6} (min distance = 1.8)
- Add random noise to shift value
  - 0.01 ppm for HX
  - 0.20 ppm for X (C/N)
  - 0.02 ppm for H
- Merge overlapped peaks
  - 0.02 for HX ppm and 0.2 ppm for X, and 0.03 ppm for H

Assign NOESY Cross Peaks
Using ASDP Software

Huang, Y. J.; Tejero, R.; Powers, R.; Montelione, G.T. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *PROTEINS: Struct. Funct. Bioinformatics* 15, 587-603 (2006)

- Input
  - Simulated NMR data for each CASP FM targets
  - Tolerance: 0.03 ppm for H and 0.3 ppm for C/N
- Run one cycle of ASDP
  - Output: Ambiguous restraint table with upper limit of 5 to 7.5 Å
    - Exclude short range ($|i-j|<=4$) restraints
    - Confidence Score of matching to chemical shifts

$$\exp(-0.5*(HX/0.03)^2+(X/0.3)^2+(H/0.03)^2))$$

Validation: For each peak, at least one of the ambiguous restraints will match to the PDB structure.

Ambiguous NOE-based Contact List for CASP11

(H^N-H^N, H^N-Me, Me-Me ^{1}H-^{1}H Contacts)

Residue 1 R1	Residue 2 R2	Peak No. P#	Upper-bound UPL	Confid	Atom 1 A1	Atom 2 A2	
79	77	17	5.0	0.95	H	H	Peak 17
79	177	20	6.0	0.67	H	HD2	Peak 20
79	135	20	6.0	0.97	H	HD1	
79	249	20	6.0	0.96	H	HD1	
79	50	20	6.0	0.81	H	HD2	
79	217	23	5.0	0.68	H	H	Peak 23
79	230	23	5.0	0.75	H	H	
79	232	23	5.0	0.72	H	H	
79	106	23	5.0	0.76	H	H	
79	166	23	5.0	0.83	H	H	
79	100	23	5.0	0.83	H	H	
79	82	23	5.0	0.74	H	H	
79	246	23	5.0	0.71	H	H	
79	216	23	5.0	0.67	H	H	
45	37	28	7.5	0.84	HD2	HG1	Peak 28

19 CASP11-NMR Targets

CASP ID	PDB ID	Fold	#Residues with simulated CS	#ILVA	# Peaks	Avg Ambiguity Per NOESY Peak	Max Ambiguity
Ts761	4PW1	alpha+beta	214	51	3106*	9	70
Ts763	4Q0Y	alpha+beta	130	35	2029*	6	36
Ts767	4QPV	alpha+beta	274	58	1564	9	64
Ts777		alpha	345	101	2400	18	144
Ts785	4D0V	most beta	112	33	694	6	45
Ts794	4CYF	alpha+beta	462	124	3132	27	232
Ts800	4QRK	beta	212	60	1459	14	96
Ts802		beta	118	32	530	4	21
Ts804		beta	194	43	884	9	95
Ts806		alpha+beta	256	74	1791	15	136
Ts810		alpha	113	30	739	5	39
Ts812		alpha+beta	183	53	980	6	45
Ts814	4R7F	beta	397	90	2290	18	168
Ts818	4R1K	alpha+beta	134	23	516	4	21
Ts824		alpha+beta	108	27	522	3	23
Ts826		alpha	201	85	1666	14	145
Ts827		alpha	150	51	1091	8	61
Ts832	4RD8	alpha	209	56	1472	12	75
Ts835		most alpha	404	135	3517	22	223

* Distance cutoff of 6.5 ang were used for T0761 and T0763. Distance cutoff of 5 ang were used for all other targets

Backbone ^{15}N-^{1}H RDCs

For FM targets, these will be computed from the target X-ray crystal structure using the program REDCAT

H. Valafar and J. Prestegard. (2004) REDCAT: a residual dipolar coupling analysis tool. J Magn Reson. 167:228-41.

H^N Atoms will be added to the X-ray crystal structures using Richardson reduce program.

Word, et. al. (1999) Asparagine and Glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. 285: 1733-1747

We will also provide the overall molecular alignment tensor (D_a, R), together with RDC values for each assigned backbone NH.

Oligomerization state (monomer, dimer, etc) will be specified.

Future CASP Challenges

Ongoing process of generating CASP Commons Targets, Data and Structure

Modeling multiple conformational states

Modeling Using Unassigned NOESY spectra

Modeling Using Unassigned RDC data