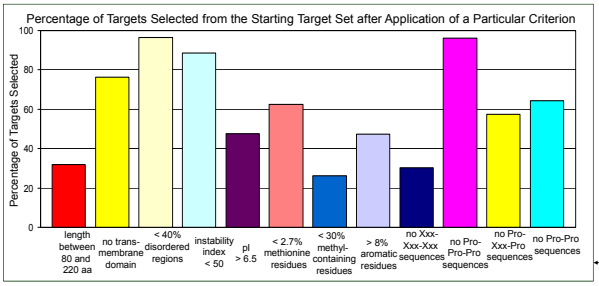
SELECTION OF PROTEINS FOR STRUCTURE DETERMINATION USING NMR

Wojciech Augustyniak[1,2], Łukasz Jaroszewski[1,3], Heath E. Klock[1,4], Daniel McMullan[1,4], Reto Horst[1,2] and Kurt Wüthrich[1,2]

[1] *The Joint Center for Structural Genomics*
[2] *The Scripps Research Institute, La Jolla, CA*
[3] *The Burnham Institute for Medical Research, La Jolla, CA*
[4] *Genomics Institute of the Novartis Research Foundation, San Diego, CA*

A new systematic approach to selection of structural genomics targets for structure determination by NMR spectroscopy is reported. A selection based on amino acid sequence was performed (selection criteria were established towards NMR amenability), followed by protein preparation and screening using NMR spectroscopy. Target selection criteria were applied to a starting target list of 430,440 open reading frames from 107 proteomes, of which 223 targets were selected based on amino acid sequence analysis and other factors. These targets were cloned and expressed using the JCSG pipeline, which yielded 23% of the proteins in soluble form. The soluble targets were subsequently prepared on a larger scale for NMR screening. Typically, 10 µL of protein solution were used to record 1D ^{1}H NMR spectra. Automated equipment allowed for sample loading into capillary tubes, sample exchange and spectra acquisition in a high-throughput fashion. Based on the 1D ^{1}H NMR spectra, targets were rated with 4 grades: A (folded globular proteins), B (folded globular proteins displaying broadened lines), C (aggregated proteins) and D (non-globular proteins). This approach allowed us to identify 20 globular proteins of A or B grade (38% among the soluble targets). The oligomeric state of these proteins will be screened and monomeric proteins will be targets for structure determination using NMR spectroscopy.



NMR Target Pipeline

Selection Criteria

430,440 targets

Target Filtering Based on Amino Acid Sequence, Homology and Availability

223 targets

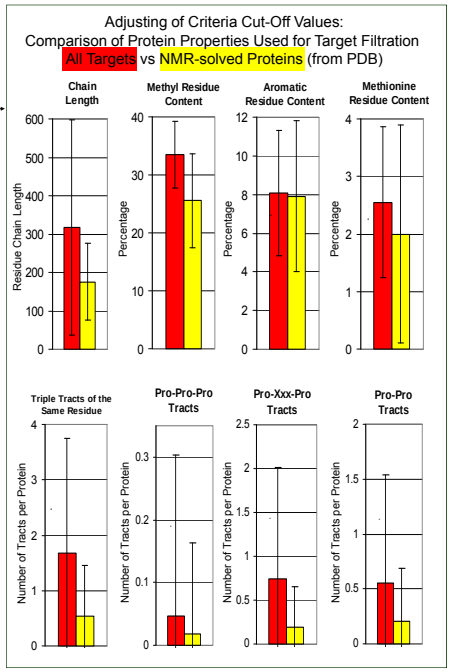
Target Cloning and Protein Preparation
JCSG Pipeline

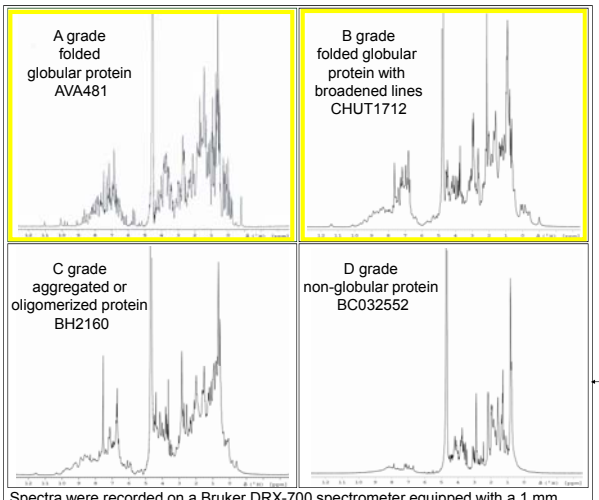
52 targets

High-Throughput Automated 1D ^{1}H NMR Spectra Acquisition and Fold Assessment

20 A and B grade targets

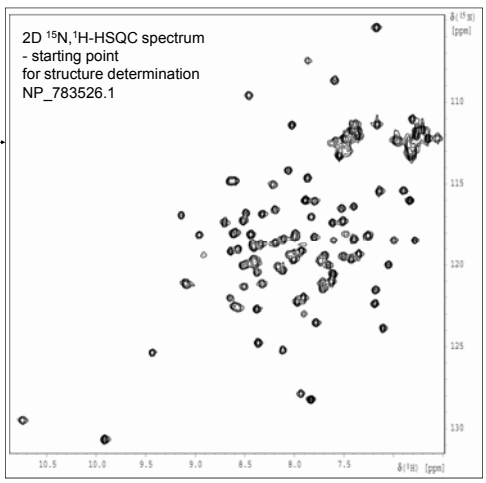
Structure Determination
- ^{15}N,^{13}C-labeled protein
- spectra acquisition
- resonance assignment
- automated structure calculation





Spectra were recorded on a Bruker DRX-700 spectrometer equipped with a 1 mm TXI probe. ~ 10 µl of protein sample were used. Fold rating was performed according to R. Page, W. Peti, I. A. Wilson, R. C. Stevens and K. Wüthrich (*Proc. Nat. Acad. Sci. USA*, **2005**, *102*, 1901-1905).

- expression strain: *Escherichia coli* GeneHogs (Invitrogen)
- vector: pSpeedET, which encodes a cleavable N-terminal expression and purification tag (MGSDKIHHHHHH)



Target Selection Summary

	All	Human	Mouse	Bacterial
Proteomes	107	1	1	105
All Targets	430,440	29,441	41,914	359,085
Filtered*	1,814	530	502	782
Selected**	223	57	37	129
Purified	52	13	14	25
A grade	**14**	**1**	**3**	**10**
B grade	**6**	**4**	**0**	**2**
C grade	18	4	4	10
D grade	8	2	3	3
Not graded	6	2	4	0

* Filtered with NMR selection criteria exclusively; ** Finally selected for cloning after i) removal of targets with high homology to PDB-deposited proteins, ii) removal of targets of advanced status at other SG centers, iii) matching with JCSG clone library

Acknowledgments

Dr. Scott A. Lesley, Dr. Mark W. Knuth and Julie Feuerhelm from the Genomics Institute of the Novartis Research Foundation are acknowledged for their generous help during target cloning and expression.

W.A. is a Marie Curie Outgoing International Fellow sponsored by the European Community.