SELECTION OF PROTEINS FOR STRUCTURE DETERMINATION USING NMR

Wojciech Augustyniak[1,2], Łukasz Słabiński[1,3], 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 systematic approach to the selection of structural genomics targets for structure determination using NMR spectroscopy is reported. It includes gene annotation for a set of 107 genomes, target selection based on the amino acid sequence and other criteria, protein preparation and screening using NMR spectroscopy. Target selection criteria were applied to a starting list of 430,440 open reading frames, of which 223 targets were cloned and expressed using the JCSG protein production pipeline, which yielded 52 of the proteins in soluble form. For NMR screening, typically, 10 µL of approximately 1 mM protein solution were used to record a 1D ^{1}H NMR spectrum. Automated equipment allowed for sample loading into capillary tubes, sample exchange and spectra acquisition in a high-throughput fashion. This screening approach based on the 1D ^{1}H NMR spectra allowed us to identify 20 promising globular target proteins. Eight of these proteins were selected for further studies based on 2D [^{15}N,^{1}H]-COSY spectra, and three monomeric proteins in this group are presently prioritized for structure determination using NMR spectroscopy.

NMR Amenability (fraction of targets selected): *
- molecular mass between 80 and 220 amino acid residues (32%)
- methyl-containing residues < 30% (26%)
- predicted disordered regions < 40% (96%)
- no predicted transmembrane helices (76%)
- instability index < 50 (89%)
- aromatic residues > 8% (47%)
- no triple tract of any amino acid (30%)
- no Pro-Pro sequence (64%)
- no Pro-Xxx-Pro sequences (57%)

Orthogonality to X-ray Crystallography:
- methionine residues < 2.7% (62%)
- pI above 6.5 (48%)

- low homology to proteins solved or in advanced stage at another SG center #
- available in the JCSG clone collection
- represented in human or mouse

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

NMR Target Pipeline

Annotation of Target Genes
Bioinformatics Core

430,440 targets

Target Filtering Based * on Amino Acid Sequence
Bioinformatics Core

1,814 targets

Target Selection Based # on Homology and Availability
Bioinformatics Core

223 targets

Target Cloning and ❑ Protein Preparation
Crystallomics Core

52 targets

Microscale 1D ^{1}H NMR Spectra Acquisition and Fold Assessment ❖
NMR Core

20 A and B grade targets

Additional Studies ➢
- gene re-cloning into pET vectors
- protein expression
- stability assessment
- oligomerization studies
- ^{15}N labeling
- evaluation of target quality with 2D ^{15}N,^{1}H-HSQC spectrum

NMR Core

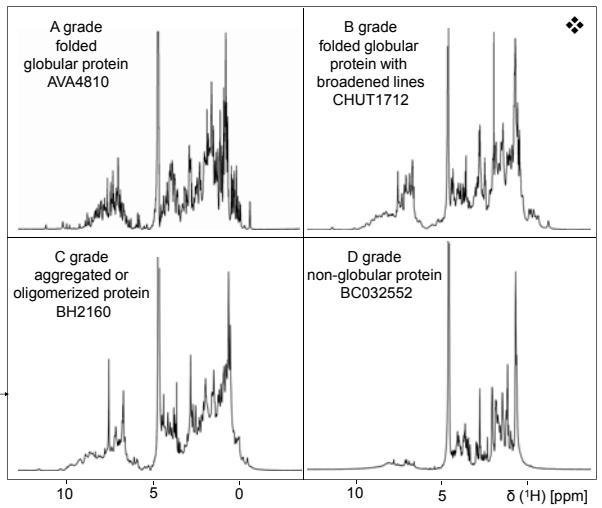
8 final targets

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

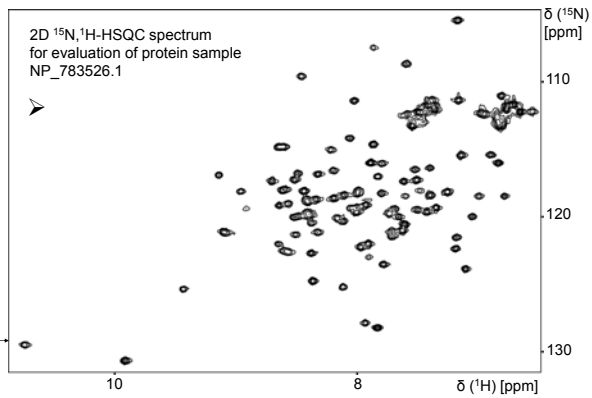
NMR Core

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
For NMR➢	**8**	**0**	**1**	**7**



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).



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.
The JCSG is supported by the NIH Protein Structure Initiative grant U54 GM074898 from the National Institute of General Medical Sciences (www.nigms.nih.gov).
W.A. is a Marie Curie Outgoing International Fellow sponsored by the European Community.