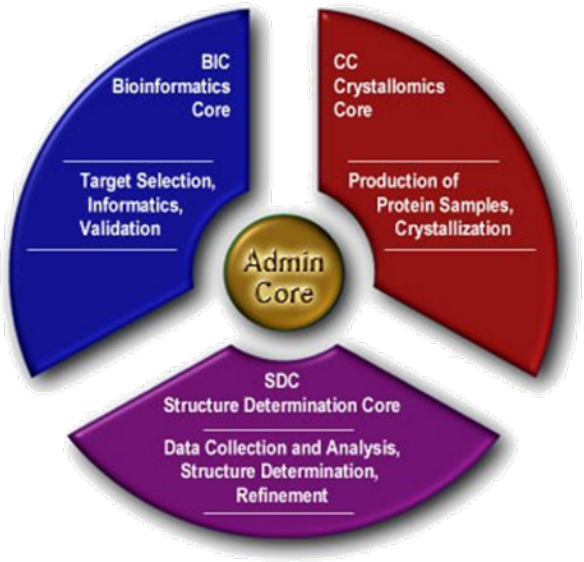
SELECTION OF PROTEINS FOR STRUCTURE DETERMINATION USING NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

Wojciech Augustyniak

Annual Meeting, April 26th-27th,2007



Biomedical theme: **The Central Machinery of Life** — proteins that are conserved in all kingdoms of life.

Goal: progress towards appropriate target selection and automated protein structure determination using NMR spectroscopy.

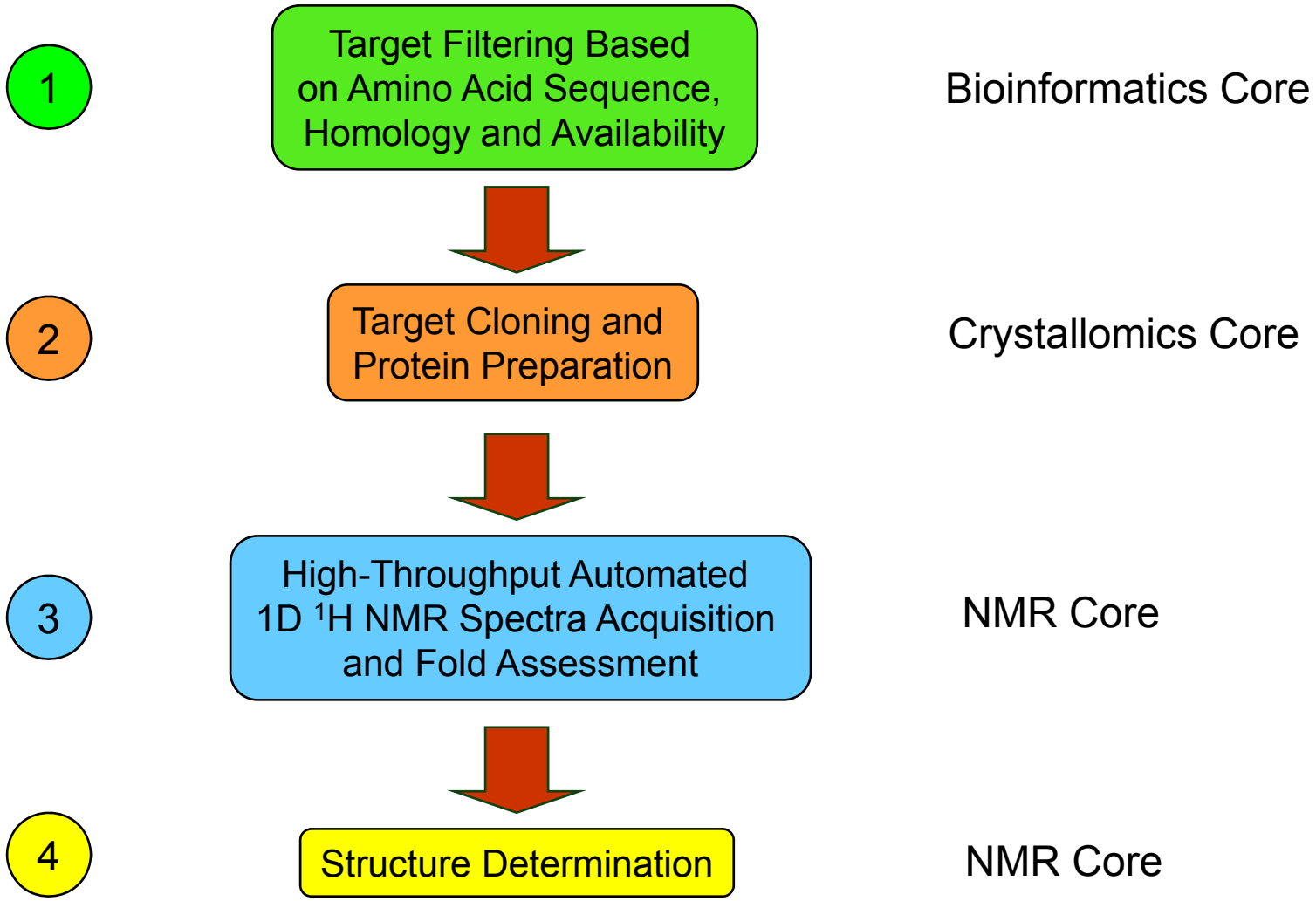
Starting point:

- High throughput fold screening protocol for protein samples was established.
- Microprobe technology was included into the screening protocol.
- Robots allowing automated high-throughput sample preparation and spectra. acquisition (sample loader / sample changer) were installed.

Next step: **setup of NMR Pipeline**

- Use bioinformatics to filter the proteins suitable for NMR structure determination.
- Select targets for structure determination in our Core.

NMR Pipeline Flowchart



1

Target Filtering Principles

Proteins easy to handle
- less than 220 residues
- high content of aromatic residues
- low instability index
- low content of disordered regions
- low content of cysteine residues

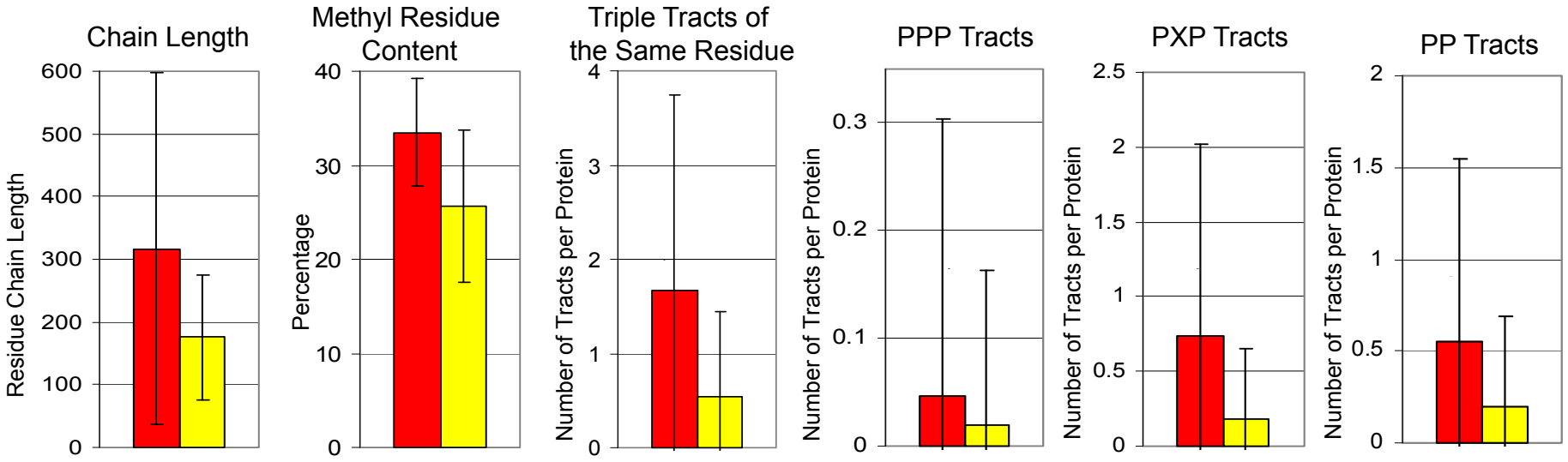
Amino acid composition suitable for resonance assignment
- no high excess of any amino acid
- low content of methyl residues
- no triple tracts of any amino acid
- no PP or PXP sequences

Orthogonality to the crystallography selection criteria
- high pI
- low content of methionine residues

1

Selection Criteria Adjustment

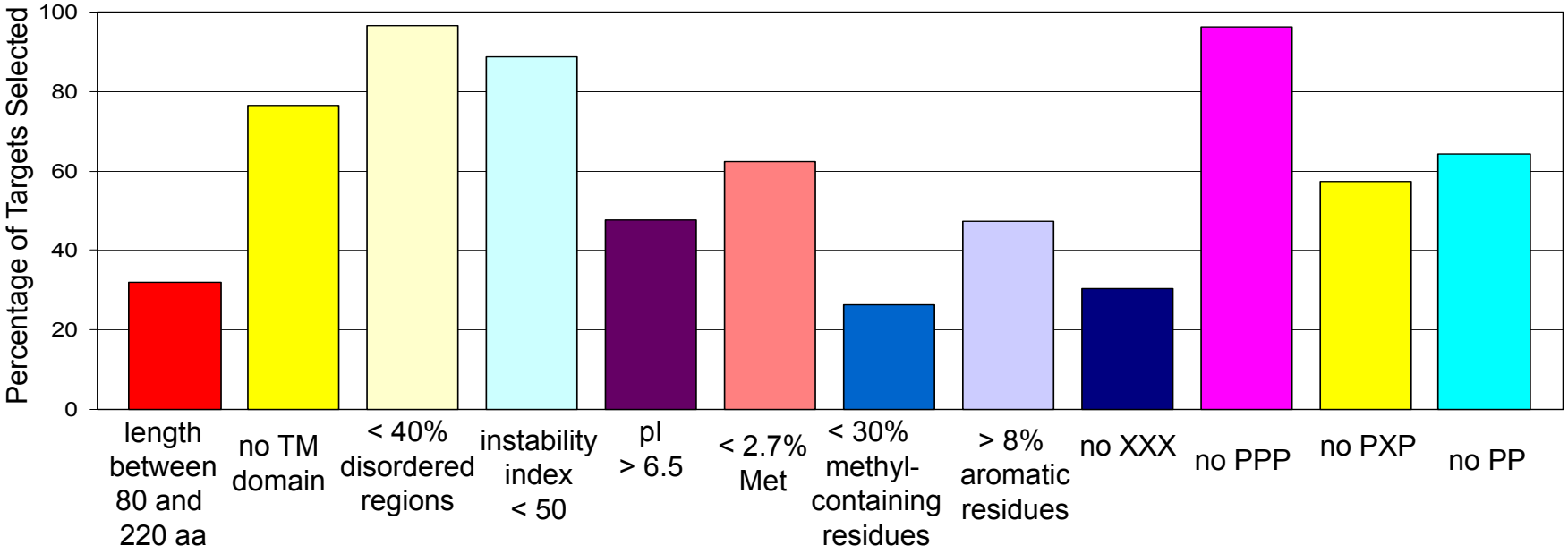
Comparison of properties of all proteins with proteins whose structures have been solved using NMR



- Proteins solved using NMR are smaller and contain less methyl-residues; they also contain lower number of sequences difficult for resonance assignment.
- Statistical analysis allowed to choose appropriate cut-off value for specific filtering criteria.

1

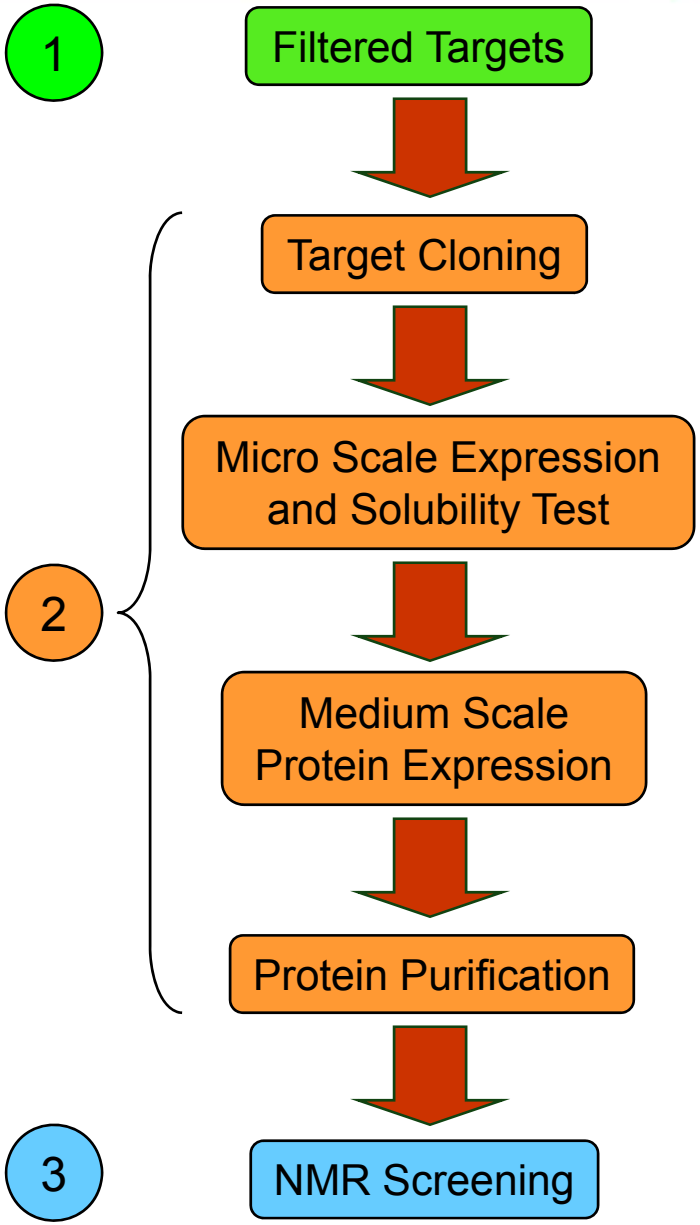
Selection Strength of the Criteria



Additional information:

- The NMR selection criteria selected 0.42% of all initial targets.
- Further selection of proteins:
  - ✓ Less than 30% homology to PDB-deposited proteins or to targets advanced in other SG centers: 35% selected.
  - ✓ Clones available at GNF (only mammalian targets): 19% selected.

223 (0.052%) of 430,440 targets were selected for cloning and expression.



Sample Preparation at JCSG Pipeline

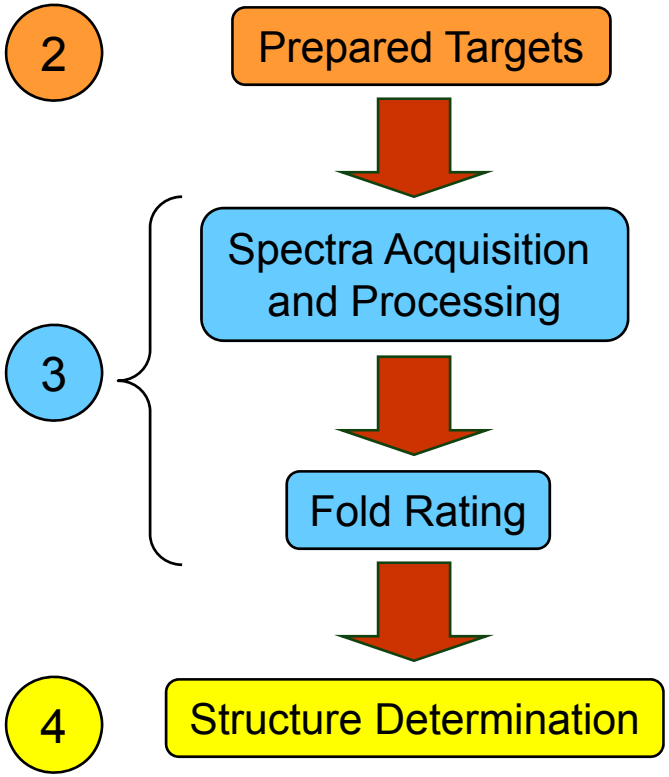
Expression vector: pSpeedET – a plasmid allowing expression under control of arabinose-inducible T7 RNA polymerase promoter.

Expression strain: *E. coli* GeneHogs (Invitrogen).

Purification: *N*-terminal TEV-cleavable polyhistidine purification tag was attached to the protein; purification on nickel-affinity chromatography.

52 (23%) of 223 cloned targets were solubly expressed and passed to NMR screening.

NMR Screening



High-throughput automated sample loading to 1 mm diameter capillary tubes: Gilson Liquid Sample Handler robot, ~ 5 µL of active volume in the tube.

1D ^{1}H NMR spectra acquisition: Bruker DRX-700 spectrometer, 1 mm probe, Bruker Automated Sample Changer used for high throughput automated spectra acquisition.

Fold rating according to R. Page *et al. PNAS*, 2005, *102*, 1901-1905.

Target Selection Summary

	All	Human	Mouse	Bacterial
Proteomes	107	1	1	105
All Targets	430,440	29,441	41,914	359,085
1 Filtered*	1,814	530	502	782
1 Selected**	223	57	37	129
2 Purified	52	13	14	25
3 A grade	**14**	**1**	**3**	**10**
3 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.

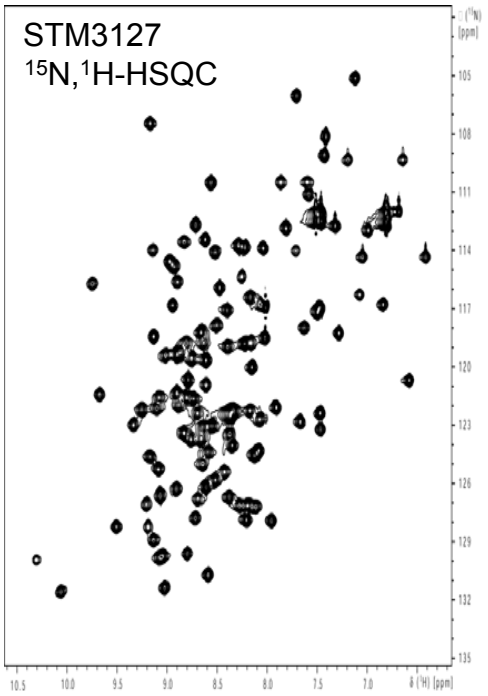
** Finally selected for cloning after i) removal of targets with high homology to PDB-deposited proteins, ii) removal of targets of advanced status at other SG centers, iii) matching with JCSG clone library.

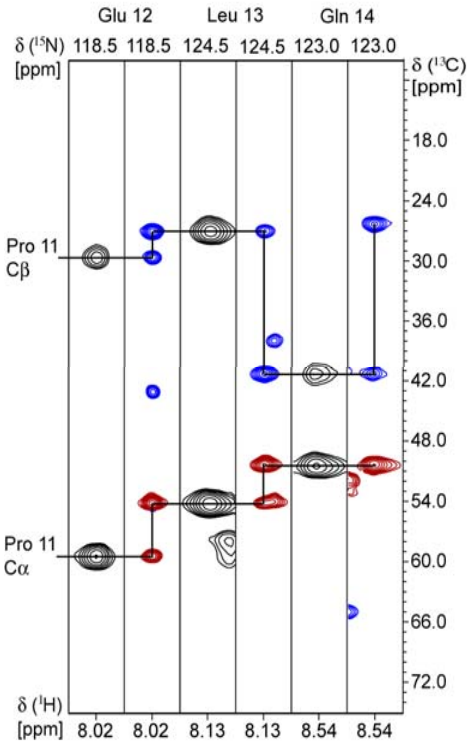
4

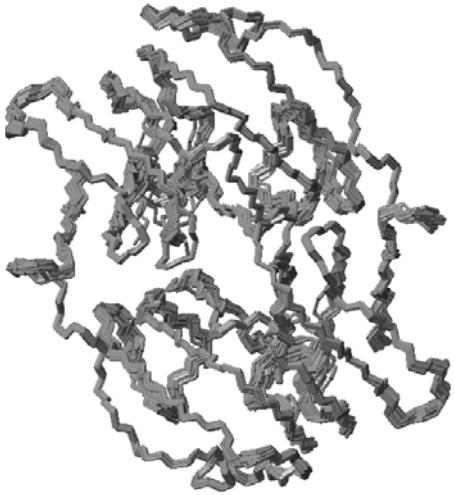
Structure Determination

sample preparation → spectra acquisition → resonance assignment → automated structure calculation based on NOESY spectra

- gene re-cloning
- protein expression and purification
- stability screening
- ^{15}N,^{13}C-labeling
- approximately 600 µL of 1 mM (or more) sample necessary







homodimer

Outlook

About half of the selected proteins turned out to be oligomers.

Structures of oligomeric proteins are difficult to calculate automatically. Time-consuming manual work is required to determine the structure.

Additional improvement: further screening for oligomerization state in various conditions and selection of monomeric proteins as the primary NMR targets.

Oligomerization state screening techniques:

- size exclusion chromatography
- light scattering
- Diffusion Ordered Spectroscopy
- chemical cross-linking

Acknowledgments

BURNHAM INSTITUTE for MEDICAL RESEARCH
From Research, the Power to Cure

Bioinformatics Core:
Łukasz Jaroszewski

THE SCRIPPS RESEARCH INSTITUTE

Genomics Institute of the Novartis Research Foundation

Crystallomics Core:
Scott A. Lesley
Mark W. Knuth
Daniel McMullan
Heath E. Klock
Julie Feuerhelm

NMR Core:
Kurt Wüthrich
Reto Horst
Bill Pedrini
Michael Geralt
Pedro Serrano
Will Placzek
Margaret Johnson
Amarnath Chatterjee