

# Knowledge-Based Throughput Enhancement for NMR-Based Structural Proteomics

Jikui Song<sup>1</sup>, Robert C. Tyler<sup>2</sup>, Claudia C. Cornilescu<sup>1</sup>, Francis Peterson<sup>2</sup>, Brian F. Volkman<sup>2</sup>,  
Dmitriy Vinarov<sup>1</sup>, Carrie Loushin Newman<sup>1</sup>, Brian G. Fox<sup>1</sup>, and John L. Markley<sup>1</sup>

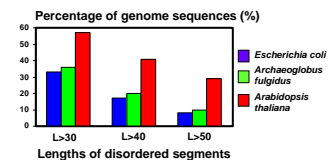
<sup>1</sup>University of Wisconsin-Madison, Department of Biochemistry, 433 Babcock Drive, Madison, WI 53706-1549, <http://www.uwstructuralgenomics.org>

<sup>2</sup>Medical College of Wisconsin, Department of Biochemistry, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA

## Abstract

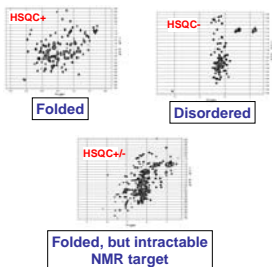
A major bottleneck in high-throughput eukaryotic structural proteomics arises from the fact that the majority of proteins that can be produced in soluble form fail to qualify as good candidates for NMR structure determination. Reasons for this include dynamic disorder of a large fraction of the sequence, partial or full aggregation, and limited stability over time. Proteins from eukaryotes appear to be more dynamically disordered than proteins from prokaryotes. In recognition of these challenges, the Center for Eukaryotic Structural Genomics (CESG) has explored ways of converting unpromising proteins to successful structural targets. We have established an integrated approach in which bioinformatics, literature knowledge, small-scale screening, and NMR spectroscopy are combined to guide the process of optimizing the protein sequence, the buffer and pH, salt concentration, or additives such as DTT or detergents. We present examples this integrated approach and show how it has significantly enlarged the size of target pool for NMR structure determination at CESG

## Increased Protein Disorder in Eukaryotic Genomes

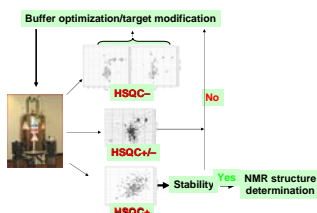


Dunker AK et al. (2000) *Genome Informatics*, 11, 161-170.

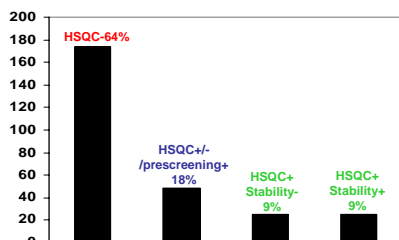
## Characterization of Protein Folded-ness by [<sup>1</sup>H, <sup>15</sup>N]-HSQC NMR Spectra



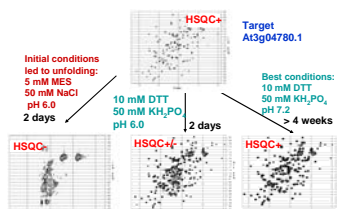
## NMR Target Screening



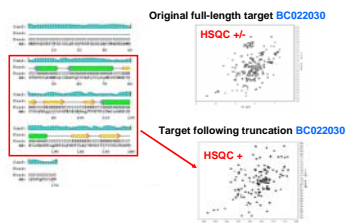
## NMR Target Screening (July, 2003-March, 2006)



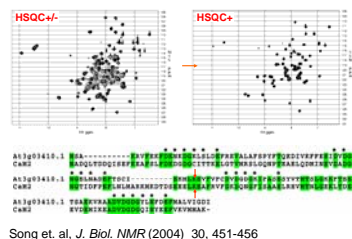
## Screening of Solution Condition for Minimal Protein Aggregation



## Sequence Modification for Target Optimization

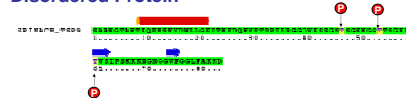


## Domain Separation for Target Optimization



Song et al. *J. Biol. NMR* (2004) 30, 451-456

## Spinach Thylakoid Soluble Phosphoprotein of 9 kD (TSP9), an Intrinsically Disordered Protein

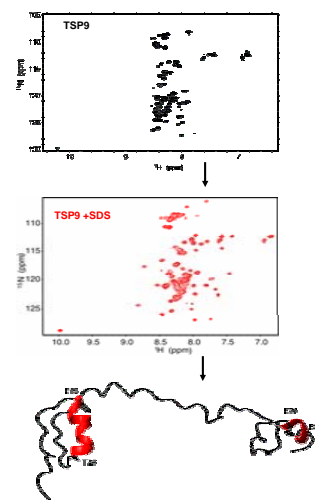


### Biochemical features:

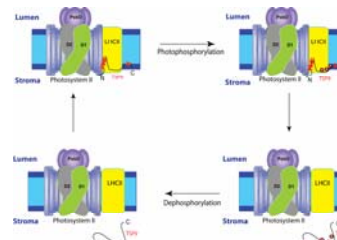
- Highly basic, pI = 9.8
- Low sequence complexity: Gly (18%), Lys (18%), Ser /Thr (20%), three phosphorylation sites, highlighted in yellow in sequence

Carlberg et al. (2003) *PNAS*, 100, 757-762

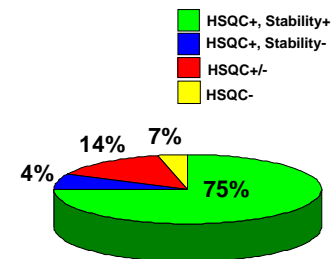
## SDS Micelle-Induced Folding of Spinach TSP9



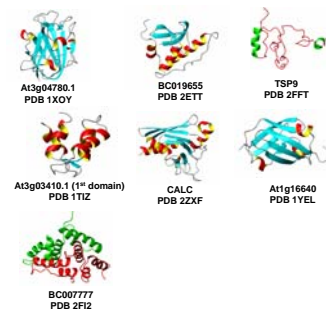
## A Disorder-to-Order Model of Spinach TSP9 Upon Membrane Association



## Original Screening Categories for NMR Structures



## Examples of NMR Structures which Were Determined from Originally Intractable NMR Targets



## Conclusion

Increased protein disorder in eukaryotic genomes has limited the suitability of a large number of targets for structure determination. To enhance the throughput in the structure genomics pipeline, significant efforts are required to convert the intractable targets to appropriate ones. NMR spectroscopy provides direct information to assess the protein folding and stability. The knowledge-based target salvage approach, where the information from NMR spectroscopy is combined with bioinformatics and literature knowledge, plays a powerful role in guiding the process of buffer optimization and target modification

## Acknowledgements

All CESG staff members and collaborators