Northeast Structural Genomics Consortium
Consensus approach to predicting protein disorder

Natively disordered or unstructured regions in proteins are both common and biologically important, particularly in modulating intermolecular recognition processes. From a practical point of view, however, such disordered regions often can pose significant challenges for crystallization. Disordered regions are also detrimental to NMR spectral quality, complicating the analysis of resonance assignments and three-dimensional protein structures by NMR methods. Identification of such disordered regions, by either experimental or computational methods, is a fundamental step in the NESG structure production pipeline, allowing the rational design of protein constructs that have improved expression, better solubility, improved crystallization, and which provide better quality NMR spectra.

The **DisMeta Server** (www-nmr.cabm.rutgers.edu/bioinformatics/disorder) runs a wide range of disorder prediction software, including DISEMBL (Linding *et al*., 2003 Structure *11*, 1453), DISOPRED2 (Ward *et al*., 2004 J Mol Biol *337*, 635), DISPro (Cheng J, 2005 Data Mining and Knowledge Discovery *11*, 213), DRIP-PRED (MacCallum, CASP 6 meeting; Online paper), FoldIndex (Prilusky *et al*., 2005 Bioinformatics *21*, 3435), FoldUnfold (Galzitskaya *et al*., 2006 Bioinformatics *22*, 2948), GlobPlot2 (Linding *et al*., 2003 Nucleic Acids Res *31*, 3701), IUPred (Dosztanyi *et al*., 2005 J Mol Biol *347*, 827), Prelink (Coeytaux and Poupon, 2005 Bioinformatics *21*, 1891), RONN (Yang *et al*., 2005 Bioinformatics *21*, 3369), VL2 (Vucetic *et al*., 2003 Proteins *52*, 573), VL3 (Obradovic *et al*., 2003 Proteins *53 Suppl 6*, 566), VL3H (Obradovic *et al*., 2003), and VSL2 (Peng *et al*., 2006 BMC Bioinformatics *7*, 208). Representative output are shown in Fig. 1.

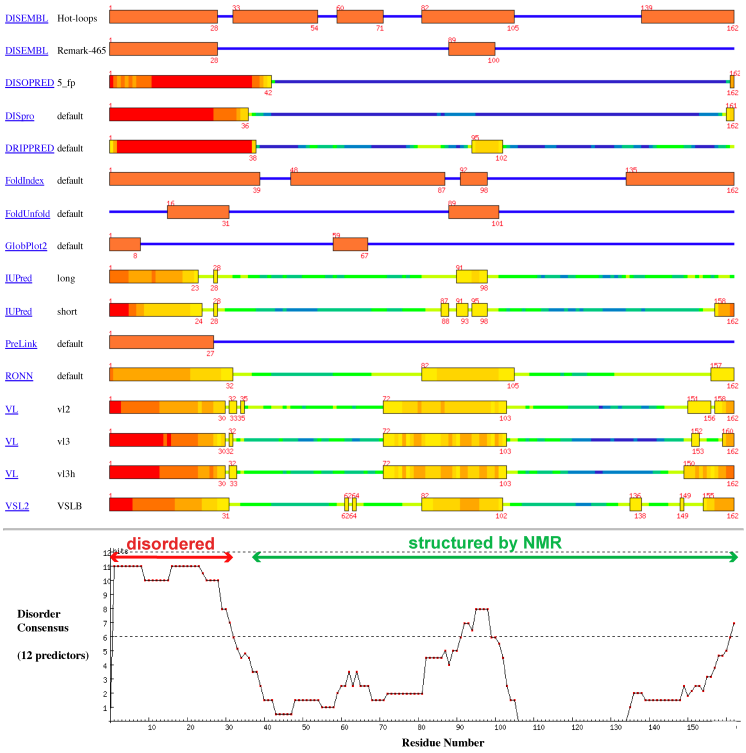

Fig. 1. Disorder consensus report for the *Escherichia coli* Spr lipoprotein, NESG target ER541, which originally provided NMR data of marginal quality, and no crystals in HPT crystal screens. In this case, the prediction programs provide a clear consensus result, namely strong evidence for disorder in the N-terminal region of the protein (red double-head arrow). On the basis of these results, several truncated constructs lacking residues from this region were generated, ultimately leading to the production of Spr(37-162) (green double-headed arrow) whose solution NMR structure was solved in our consortium (PDB ID, 2K1G) (Aramini *et al*., 2008, Biochemistry *47*, 9715).

The DisMeta Server also provides sequence-based structural prediction results from other bioinformatics software, including PROF (Rost *et al*., 2004 Nucleic Acids Res *32*, W321), PSIPred (Jones, 1999 J Mol Biol *292*, 195), SignalP (Emanuelsson *et al*., 2007 Nat Protocols *2*, 953), TMHMM (Krogh *et al*., 2001 J Mol Biol *305*, 567), Coils (Lupas *et al*., 1991 Science *252*, 1162), and SEG (Wootton and Federhen, 1996 Methods Enzymol *266*, 554). These graphical reports provide information allowing successful construct optimization for both NMR and crystallization studies.