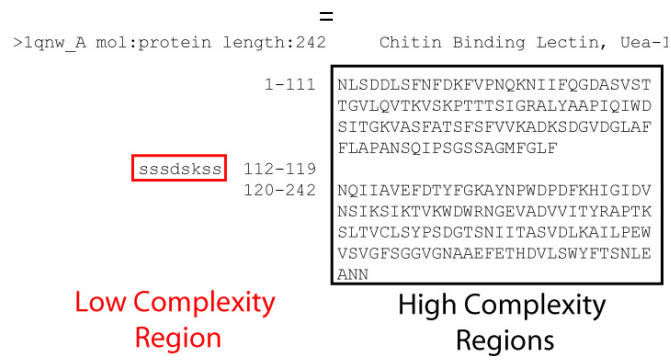
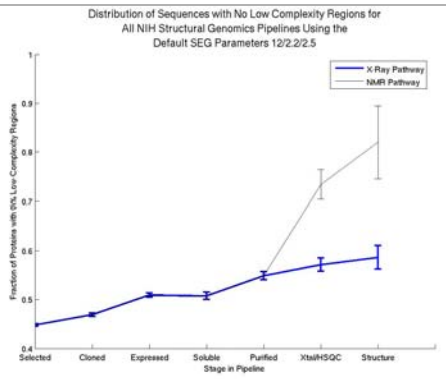
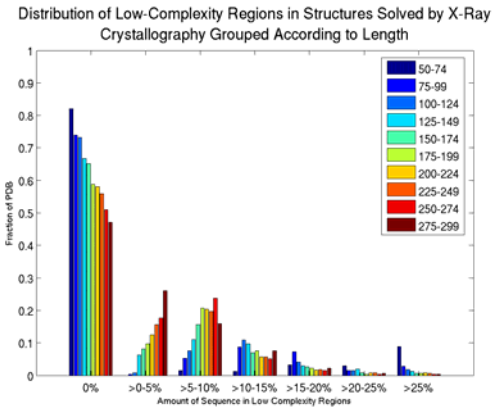
Center for Eukaryotic Structural Genomics

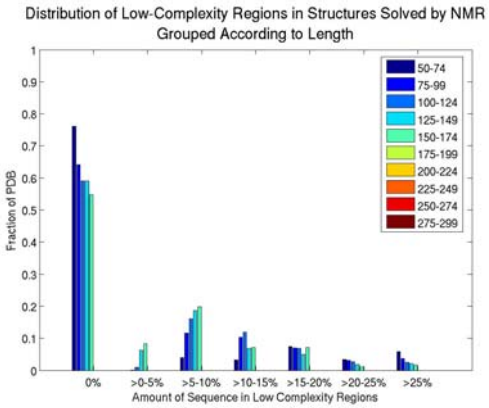
Technology Dissemination Report

CESG Tech Report No.	004
Title	Evaluation of Low-Complexity in Amino Acid Sequences for Target Selection in Structural Genomics
Research Unit	Crystallography
Authors	Bannen, R.M., Bingman, C.A., and Phillips, G.N., Jr.
Primary Contact	phillips@biochem.wisc.edu









Summary

It has been previously shown that protein sequences containing a quasi-repetitive assortment of amino acids are common in genomes and databases such as Swiss-Prot but are under-represented in the structure based Protein Data Bank (PDB). Structural genomics groups have been using the absence of these “low-complexity” sequences for several years as a way to select proteins that have a good chance of successful structure determination. In this study, we present a careful examination of the data deposited in the PDB as well as the available data from structural genomics groups in TargetDB and PepcDB to reveal interesting trends that could be taken into consideration when using low-complexity sequences as part of the target selection process [1]. In particular, while the presence of low-complexity regions appears to inhibit the structure determination of protein structures by both nuclear magnetic resonance (NMR) and X-ray crystallography, it appears that when the proteins are normalized by length, NMR shows a higher tolerance for proteins with low-complexity regions.

Publication:

[1] Bannen, R.M., Bingman, C.A., Phillips, G.N., Jr. (2007) Effect of low-complexity regions on protein structure determination. *JSFG* 8(4):217-26.

Acquiring the Technology	Low-complexity calculated by SEG (Wootton, J. and Federhen, S. *Computers Chem.* 17:149-163). Low-complexity scripts are available from Ryan Bannen.
Other Acknowledgements	Thanks to Gary Wesenberg and Xiaokang Pan for the ORF scoring software.

Center for Eukaryotic Structural Genomics (CESG), University of Wisconsin-Madison Biochemistry Department, 433 Babcock Drive, Madison, WI 53706-1549; phone: 608.263.2183; fax: 608.890.1942; email: cesginfo@biochem.wisc.edu; website: http://www.uwstructuralgenomics.org.

This research funded by NIH / NIGMS Protein Structure Initiative grants U54 GM074901 and P50 GM064598.