View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Harvard University - DASH

Order-Disorder Interface Characterization Reveals Critical Factors for Disease and Drug Targets

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters.

Citation	Kallenbach, Jonah, Wei-Lun Hsu, A. Keith Dunker, and Gil Alterovitz. 2013. "Order-Disorder Interface Characterization Reveals Critical Factors for Disease and Drug Targets." AMIA Summits on Translational Science Proceedings 2013 (1): 101.
Accessed	February 19, 2015 3:04:19 PM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:11879406
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

Order-Disorder Interface Characterization Reveals Critical Factors for Disease and Drug Targets

Jonah Kallenbach[1], Wei-Lun Hsu[5], A. Keith Dunker[5], Gil Alterovitz[1234]*

[1] Center for Biomedical Informatics, Harvard Medical School [Boston, MA 02115]
[2] Children's Hospital Informatics Program at Harvard-MIT Division of Health Science [Boston, MA02115]
[3] Partners Healthcare Center for Personalized Genetic Medicine [Boston, MA 02115]
[4] Department of Electrical Engineering and Computer Science at Massachusetts Institute of Technology [Cambridge, MA 02139]
[5] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, 410 W 10th Street, Suite 5000 Indianapolis, IN 46202
*corresponding author: gil@mit.edu

Abstract: Signal transduction pathways are of critical importance in disease and regulation of cellular functions. Proteins that do not fold to a state of stable tertiary structure, known as intrinsically disordered proteins, are highly represented in signaling pathways and protein interaction networks. Important examples of disordered signaling proteins include p53 and BRCA1, and approximately 40% of Eukaryotic proteins are estimated to have significant disordered regions. Certain regions within these disordered proteins, however, can take on an ordered structure upon binding to a partner. The nature of the resulting protein-protein interactions has not yet been established. Here we categorize and identify interactions between binding segments of disordered proteins and their ordered partners using a Bayesian network framework, constructed on a test set of 964 proteins mined for Molecular Recognition Feature (MoRF) characteristics from the PDB. This framework, more specifically Bayesian network learning, enables us to investigate the underlying biological processes involved, including the sequential and structural determinants of these interactions. After the construction of the training set (80% of data), features were successively eliminated to determine relative significances. The Bayesian network model was validated on the test set with excellent accuracy(>90% AUC). Examining features underlying the model provides a plethora of new and potentially useful biological information. The results also lend themselves to a strategy for rational drug design whereby disordered regions can be targeted with a high degree of specificity and small molecule peptide mimetics of their binding regions can be utilized as drugs.