LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-410910

Computational biology for target discovery and characterization: a feasibility study in protein-protein interaction detection

C. Zhou, A. Zemla

February 27, 2009

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

R&R LLNL-____

Title: Computational biology for target discovery and characterization: a feasibility study in protein-protein interaction detection

Authors: Carol L. Ecale Zhou and Adam T. Zemla

Auspices Statement

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This work was funded by the Laboratory Directed Research and Development Program at LLNL under project tracking code 08-FS-009.

Abstract

In this work we developed new code for detecting putative multi-domain protein-protein interactions for a small network of bacterial pathogen proteins, and determined how structure-driven domain-fusion (DF) methods should be scaled up for whole-proteome analysis.

Introduction/Background

Protein-protein interactions are of great interest in structural biology and are important for understanding the biology of pathogens. The ability to predict protein-protein interactions provides a means for development of anti-microbials that may interfer with key processes in pathogenicity. The function of a protein-protein complex can be elucidated through knowledge of its structure. The overall goal of this project was to determine the feasibility of extending current LLNL capabilities to produce a high-throughput systems bio-informatics capability for identification and characterization of putative interacting protein partners within known or suspected small protein networks. We extended an existing LLNL methodology for identification of putative protein-protein interacting partners (Chakicherla et al (in review)) by writing a new code to identify multi-domain-fusion linkages (3 or more per complex). We applied these codes to the proteins in the Yersinia pestis quorum sensing network, known as the lsr operon, which comprises a virulence mechanism in this pathogen. We determined that efficient application of our computational algorithms in high-throughput for detection of putative protein-protein complexes genome wide would require pre-computation of PDB domains and construction of a domain-domain association database.

Research Activities and Results/Technical Outcome

The scope of code development for this project limited our work to networks comprising 10 or fewer proteins, although a secondary goal was to determine how interacting partners could best be detected starting with an arbitrarily large (e.g., whole proteome) number of putatively interacting proteins. In this project we 1) wrote a code for detecting high-order (i.e., multiple) domain-fusion linkages and applied the method to a set of putative interacting proteins from the lsr operon (a genetic unit controlling group sensing) of a bio-defense pathogen, Yersinia pestis, the caustaive agent of plague; we 2) devised a protocol for tagging linkages with structural and functional IDs to enable results clustering,

and we 3) determined that the most feasible method for identifying binary/multiple DF linkages given an arbitrarily large number of input proteins will require generation of a pre-computed database of binary domain-domain linkages, generated by means of domain splitting of all protein structures in the Protein Data Bank (PDB) followed by binary DF analysis across all PDB domains. Furthermore, domain-splitting should be done using a sliding-window approach in performing NxN sturcture comparisons (e.g. using Local-global alignment (LGA) software; Zemla 2003) to define and associate domains across the data set. Here we provide a summary of an interesting discovery with respect to some of the proteins in the lsr operon. Specifically, in applying our code to identify putative multi-protein interactions, we were able to select from among several thousand domain-fusion hits, a short list of five that provided evidence for a three-way complex among three proteins within the network. Interestingly, these five domain-fusion templates represented five different interaction poses. This preliminary result implies that domain-domain associations tend to be highly conserved, but at the same time, display a remarkable diversity in relative positioning of the interacting domains.

Exit Plan

Data generated in our limited (10-protein) study will be included in a publication being prepared in association with another research project in which the Y. pestis lsr operon was studied using computational biology and experimental methods to elucidate protein structure, interactions, and function (Zemla et al. in preparation). We expect that results and conclusions drawn from this feasibility study will be included in future grant proposals to be submitted to DOD agencies (e.g., DTRA) and NIH.

Summary

Because protein-protein interactions frequently involve proteins encoded within different genetic units, it will ultimately be necessary to analyze an entire proteome, which may comprise 4000 or more proteins. This task will require development of an efficient algorithm for domain-splitting of the entire Protein Data Bank database of protein structures, application of binary domain-fusion analysis, storage of all putative interacting domains in a database, and generation of a code for association of interacting domains.

References

Chakicherla A, CLE Zhou, ML Dang, V Rodriguez, JN Hansen, and A Zemla. SpaK/SpaR two-component system characterized by a novel structure-driven domain-fusion method and in vitro phosphorylation studies. (in review at PLoS Computational Biology).

Zemla A. 2003. LGA—a method for finding 3D similarities in protein structires. Nucleic Acids Research 31: 3370-3374.

Zemla A et al. Domain-fusion analysis of the Yersinia pestis quorum sensing network. (in preparation).