16. Molecular Recognition and Partner Prediction for Transient Protein Complexes: CDK-Cyclin Homologue Interactions

Xueping Quan, Peter Doerner, and Dietlind L.Gerloff

Biocomputing Research Unit & Structural Biochemistry Group, University of Edinburgh

To predict which specific CDK homologue interacts with which cyclin homologue in Arabidopsis thaliana, a comparative modelling strategy was applied to model the 3-D structures of these proteins. All-by-all docking of the model structures using the program ZDOCK, combined with additional selection criteria, yielded 15 most likely interacting CDK-cyclin pairs.

Introduction

CDKs (cyclin-dependent kinases) are Ser/Thr protein kinases that play an essential role in the regulation of eukaryotic cell proliferation, neuronal and thymus functions, and transcription in animals. Monomeric CDKs are inactive and require both association with a positive regulatory subunit, a cyclin, and phosphorylation of a conserved threonine residue that lies within the activation loop of the CDK, for full activity. The 3-D structures of CDKs and cyclins coming from other species, and/or other subfamilies, can be modelled using known and modelled structures of human CDKs and cyclins as templates. Previous studies and sequence sub-classification of the cyclin and CDK multigene families in the genome of the model plant Arabidopsis thaliana have revealed a minimum of 50 cyclin-like, and 35 CDK-like putative gene products. However, the sequence-structure relationships that determine the specificity of protein-protein interactions are not sufficiently understood at present to predict which specific CDK-cyclin pairings are likely to occur, and which are not. By focusing on this specific prediction challenge, we are working towards a better understanding of the biophysical principles governing protein-protein interactions in general, and towards developing computational methodology combining multiple components from several existing methods, that can be applied generally in this field.

Methods

ALL-BY-ALL DOCKING: Comparative models for 33 putative CDKs and 35 putative cyclins from Arabidopsis thaliana were produced with the program Modeller6, based on carefully adjusted multiple sequence alignments of each family [1]. The resulting structures were subjected to a large scale molecular docking experiment with ZDOCK [2] in which all CDK-cyclin combinations were considered.

ADDITIONAL SELECTION CRITERIA: Automated protein-protein docking results typically contain too many false positive complexes to be directly useful in practice. We therefore applied two additional criteria to select the best complexes from the ZDOCK result lists, based on:

(i) Orientation - correct relative orientation of CDK and cyclin subunits in the complex;
(ii) Interface – the electrostatic and hydrophobic properties at the interaction surface. Interface electrostatic correlation coefficients (ECC) and hydrophobic correlation coefficients (HCC) were calculated using the program MolSurfer [3].

CALIBRATION: For calibrating the interface criterion we used a positive, and a negative set of control data. The positive set consisted of 104 non-homologous, transient hetero-dimer complexes [4]. For the negative set 70 “non-complexes” were generated by using ZDOCK to combine proteins that do not normally interact (ZDOCK score > 60; interface size > 600 square angstroms). An optimal cutoff value

for CEH (a combination of ECC and HCC) was chosen, based on the interface properties of these two sets.

Results and Discussion

• Cross-validation, by randomly selecting 80% of data from each control set to derive the CEH cutoff, consistently yielded separation accuracies around 80% for the other 20% of control complexes. When we applied the entire modelling and interaction prediction approach to the well-characterized set of human CDKs and cyclins, 80% of the resulting predictions were in agreement with HRPD and Swiss-Prot annotation.

• All-by-all docking between 33 CDK models and 35 cyclin models yielded 40 pairs with outstanding ZDOCK scores (> 60). Of these pairs, 37 are supported by correct orientation between the interacting subunits. Using CEH as an additional selection criterion, we retained 15 most likely interacting CDK-cyclin pairs in Arabidopsis thaliana. The most strongly predicted complex is formed between a close homologue of human CDK1/2/3, and a sequence most similar to human cyclinA (human CDK1/2/3-cyclinA are natural pairs). Another predicted complex has recently been confirmed experimentally.

References

1. Modeller: Sali & Blundell, 1993, J.Mol.Biol. 234: 779-815
2. ZDOCK: Chen et al., 2003, Proteins 52: 80-87
3. MolSurfer: Gabadoulline et al., 2003, Nucl.Acids Res., 31:3349-3351
4. Heterodimer Complexes: Ofran & Rost, 2003. J.Mol.Biol. 325:377-387.