MotifMarker Demo

This document contains a demo of MotifMarker used to study the structures of the CDK2,CDC2, CDK4 and CDK6 highlighting the 3 important domains/motifs, namely, the PSTAIRE region, ATP binding domain and kinase domain, in these cyclin-dependent kinases and try to relate the structural differences and similarities to their functions and cyclin partner specificity.

Methods & Programs Used

Finding Motifs in Protein Sequences and Marking Their Positions in Protein Structures

A java program *MotifMarker* is designed and implemented to fish out motifs in sequences contained in fasta files and generate a report in HTML format and corresponding scripts for each results according to the template script provided. The program uses simple console environment and requires Java 2 Platform Standard Edition (J2SE) 5.0 to compile and run. The program first reads in the fasta file containing the sequences, and loads in a consensus files containing the motifs to find and a script templates containing the template to create scripts (These templates can be written in different languages so that scripts can be generated for running in Rasmol, Spb and adapting *MotifMarker* to different scriptable bioinformatics software only requires writing the corresponding script templates.) The source codes and compiled program is available online at http://albertwcheng.inscyber.net/motifmarker.htm More details of *MotifMarker* are given in Appendix 1.

Obtaining Protein Structures

Protein structures were searched for on the NCBI homepage and Protein Data Bank(PDB) for existing experimentally determined structures. However, most of the structures of the protein of interest were studied as complexes and most of these structures are for *Homo sapiens*. In order to easily obtain a structure of the protein from the complex and do corrections on the structure to predict the protein structure not in complex and also to predict protein structure of the cell cycle proteins of species that no experimentally determined structures have been created and whose sequences have high sequence similarity to that of the available structures, Swiss-model homology modeling protein prediction was applied. Fasta sequences are inputted into the Swiss-model homology

modeling prediction server First Approach Mode online interface and results were emailed back and analyzed.

Visualizing and manipulating protein structure 3D models

The protein structure 3D models obtained from SWISS-MODEL server were analyzed and manipulated in DeepView/Swiss-PdbViewer. Superimposition of structures were performed by iterative magic fit function of the view, the sequence of the fits were determined by the phylogenetic trees and the distance matrixes of the proteins of interest, where the most similar proteins were fit together first, followed by less similar proteins, each fit into its own nearest neighbor until all structures were fitted.

Tree building and Presentation

The motifs marked by MotifMarker were analysised for phylogenetic relationship using PHYlogeny Inference Pakage (PHYLIP) and trees were visualized and presented in TreeView

Results

Table r1a PSTAIRE sequences of *Homo sapiens* CDKs

CDC2	EGVPSTAIREISLLKE
CDK2	EGVPSTAIREISLLKE
CDK4	GGLPISTVREVALLRR
CDK6	EGMPLSTIREVAVLRH

* Consensus sequence found by MotifMarker using consensus PSTAIRE described in Table r1d

Table r1b ATP binding sequences of *Homo sapiens* CDKs

CDC2	GEGTYGVVYK
CDK2	GEGTYGVVYK
CDK4	GVGAYGTVYK
CDK6	GEGAYGKVFK

* Consensus sequence found by MotifMarker using consensus ATPBINDING described in Table r1d

Table r2c Kinase sequences of *Homo sapiens* CDKs

CDC2	HRDLKPQNLLI
CDK2	HRDLKPQNLLI
CDK4	HRDLKPENILV
CDK6	HRDLKPQNILV

* Consensus sequence found by MotifMarker using consensus KINASE described in Table r2d

Table r1d Consensus for important domains in CDKs

Consensus	Consensus	Regular Expression
PSTAIRE	XxxPS/I/LT/SA/TI/VRExxxxxx	[A-Z]{3}P[SIL][TS][AT][IV]RE[A-Z]{6}
ATPBINDING	GE/VGA/TYGxVxK	G[EV]G[AT]YG[A-Z]V[A-Z]K
KINASE	HRDLKPQ/ENL/ILV/I	HRDLKP[QE]N[LI]L[VI]

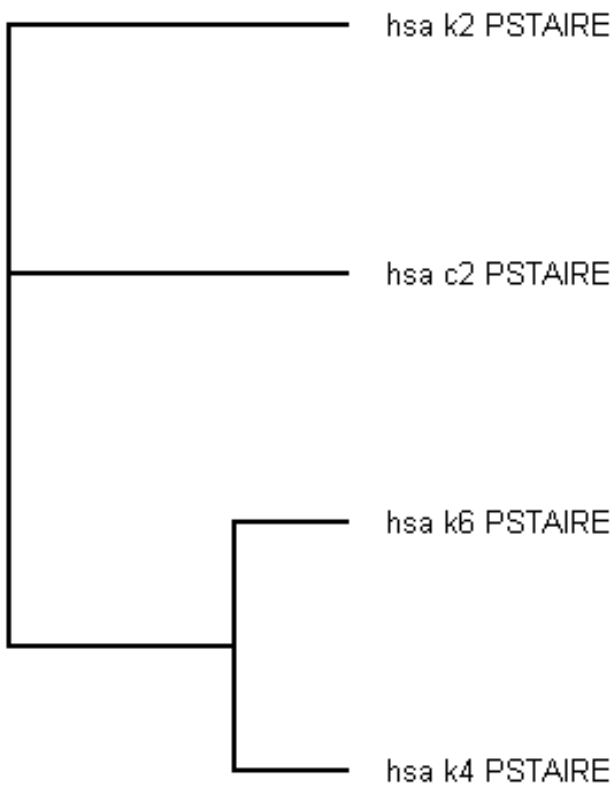

Figure r2 Cladogram of *Homo sapiens* CDK PSTAIRE sequences

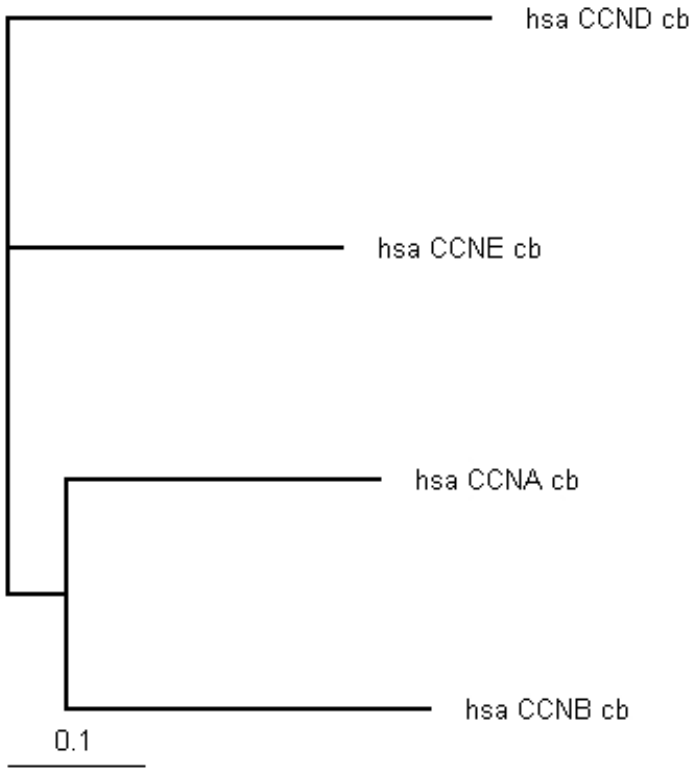

Figure r3 Phylogram of *Homo sapiens* Cyclin box sequences

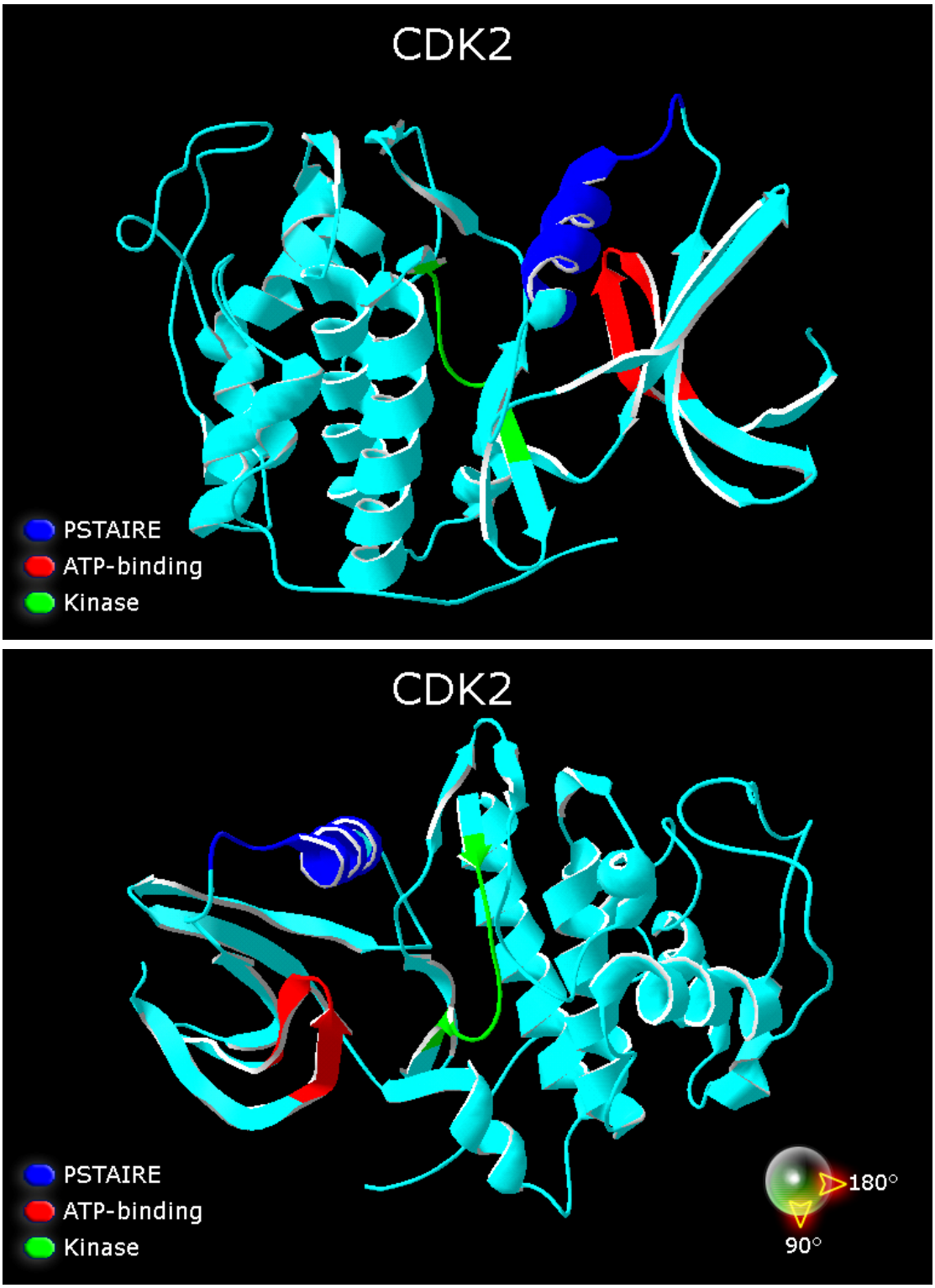

Figure r4 3D-structural model of *Homo sapiens* CDK2 highlighting the PSTAIRE region, ATP-binding domain and the kinase domain

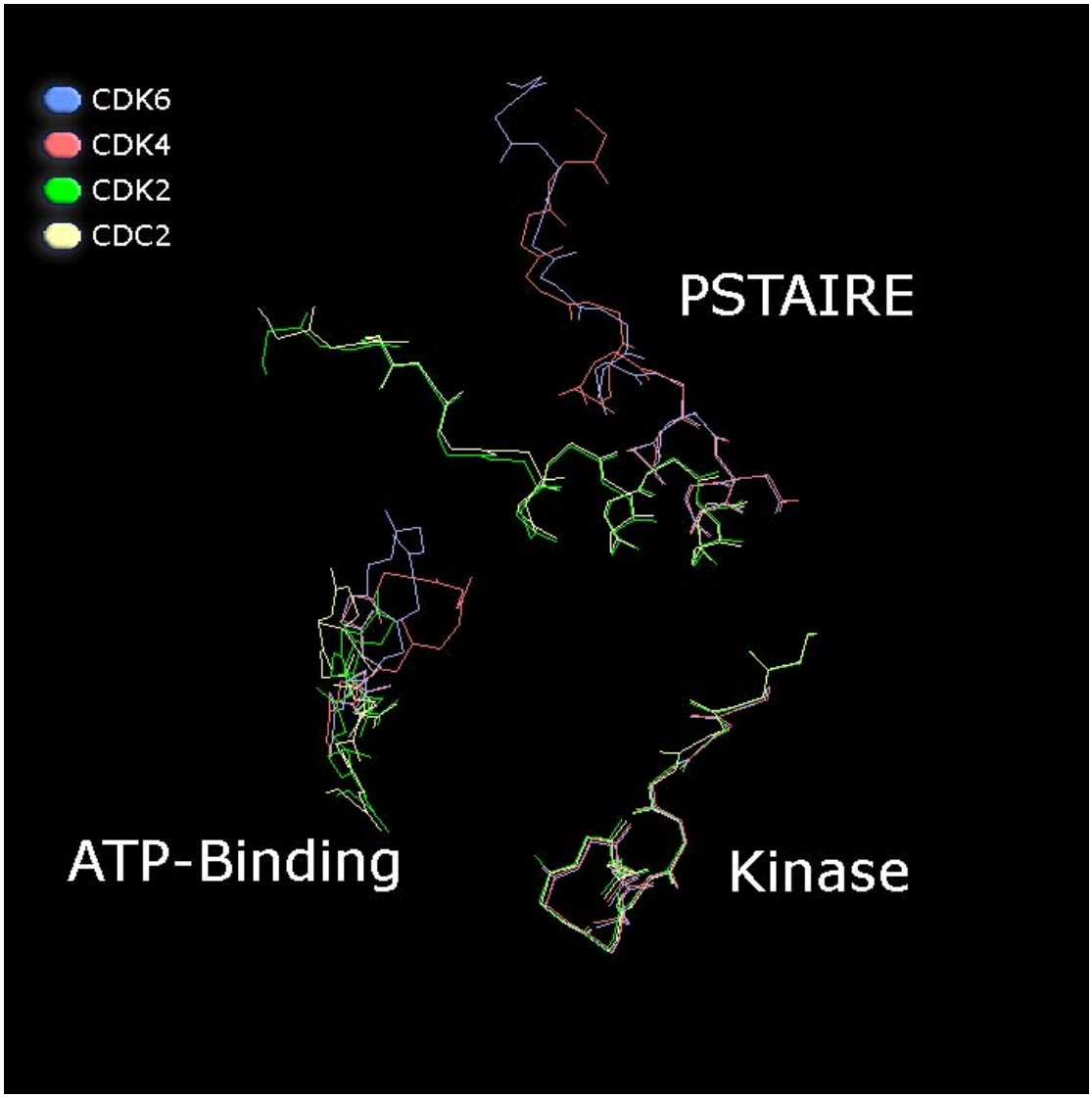

Figure r5 Superposition of Homo sapiens CDKs showing the three important domains

Discussions

Structure Superposition of CDKs revealed significant misfit in PSTAIRE region among other domains.

CDKs with their PSTAIRE region, ATP binding, and kinase domains marked by MotifMarker were superposed and analyzed (fig r4,r5) . In fig r5, only the backbone in the regions of interest was shown. Among the three regions, kinase domain showed highly fit backbone, ATP binding domain showed some degree of misfed, which may be responsible for the specificity of the activation of the kinase activity. The PSTAIRE region which is the interface for binding of the regulatory subunit Cyclin showed a significant misfit among the four kinases clustering into two major groups, CDK4/CDK6 in one group, CDK2/CDC2 in another. This observation was consistent in the fact that CDK2 and CDC2 are activated both by Cyclin A while CDK4 and CDK6 are both activated by Cyciln D. This structural clustering corresponding to the binding of the same cyclin and structural misfit corresponding to the binding of different cyclins suggested that the structure of the PSTAIRE regions confer specificity and selectivity of differential cyclin binding and subsequent activation of the kinase.

References

The RSBC Protein Data Bank (online) http://www.rcsb.org/pdb/ [Available: 12 May 2005]

Martz E. (n.d.) Rasmol Homepage (online) http://www.umass.edu/microbio/rasmol/index2.htm [Available: 12 May 2005]

Hanks S.K. et al. (1987) The Protein Kinase Family: Conserved features and deduced phylogeny of the catalytic domains. *Science* 241:42-52

Morgan D.O. (1996) The dynamics of cyclin dependent kinase structure. *Current Opinion in Cell Biology* 8:767-772

Garrett M.D. (2001) Cell cycle control and cancer. *Current Science* 81:515-522

Nugent J.H.A. et al. (1991) Conserved structural motifs in cyclins identified by sequence analysis. *Journal of Cell Sciences* 99:669-674

Guex N, and Schwede T. (n.d.) DeepView Scripting Language. (Online) http://tw.expasy.org/spdbv/text/script.htm [Available: 12 May 2005]

Schwede T, Kopp J, Guex N, and Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31**:** 3381-3385.

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis* 18**:** 2714-2723.

Peitsch, M. C. (1995) Protein modeling by E-mail *Bio/Technology* 13**:** 658-660.