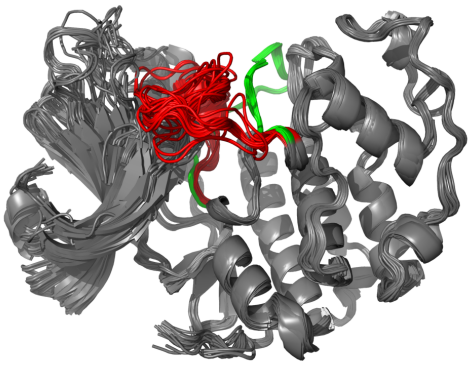
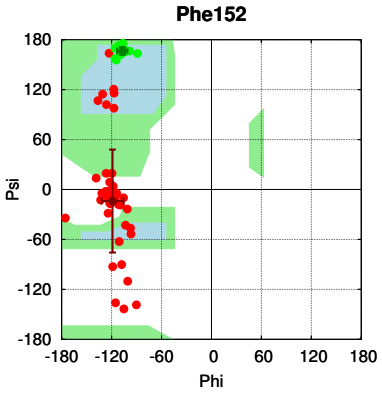
Automatic classification of CDK2 conformations

Thomas Gaillard
thomas.gaillard@polytechnique.edu

1 Context

Cyclin dependent-kinase 2 (CDK2) belongs to the serine-threonine cyclin dependent kinase (CDK) family. CDK2 is a catalytic subunit of the CDK complex and plays an important role in the regulation of the cell cycle. CDK2 activation requires the formation of a complex with a cyclin (regulatory subunit), and the phosphorylation of a conserved threonine at position 160 in the sequence, and located in what is called the activation loop or T-loop. Two states of CDK2, active and inactive, can be found, depending on the phosphorylation of Thr160. Two conformations, open and closed, are observed, depending on the presence or not of a cyclin partner. CDKs are important actors for the cell cycle. They play a role in the different phases of this cycle and are activated or inhibited by the different control processes. These roles make of CDKs interesting therapeutic targets, in particular against cancers. It has indeed been observed that CDK function was deregulated in several cancers, causing excessive proliferation of cancer cells. These properties have made of CDK2 an interesting protein to study. A large number of structures of CDK2 in complex with small molecules are available. The goal of the project is to automatically classify these conformations in different groups of interest.





2 Progression

- classify human CDK2 conformations based on annotations
  - make a list of PDB entries of human CDK2 structures
  - annotate open/closed based on the presence/absence of a cyclin or inhibitor partner
  - annotate active/inactive based on the presence/absence of a phospho-tyrosine
- classify human CDK2 conformations based on structures
  - align the sequences of the human CDK2 structures
  - superpose the human CDK2 structures
  - classify the structures with PSS dihedral-based clustering
  - propose other classification approaches
- does structure-based classification correspond to annotation-based classification?

3 Resources

- protein sequence database:
  UniProt (`http://www.uniprot.org`)

- macromolecule structure database:
  PDB (`http://www.rcsb.org`)

- sequence visualization programs:
  Jalview (`http://www.jalview.org`)
  SeaView (`http://doua.prabi.fr/software/seaview`)

- structure visualization program:
  Pymol (`https://pymol.org`)
  freely accessible legacy binaries:
  `https://sourceforge.net/projects/pymol/files/Legacy`

- Protein Structural Statistics (PSS) program:
  `http://thomasgaillard.fr/pss`

- Protein Structural Statistics web server (PSSweb):
  `http://pssweb.org`

- PSS article:
  Protein structural statistics with PSS.
  T. Gaillard, B.B.L. Schwarz, Y. Chebaro, R.H. Stote, A. Dejaegere
  Journal of Chemical Information and Modeling, **2013**, 53, 2471-2482.
  doi:10.1021/ci400233j

- Biopython:
  tutorial: `http://biopython.org/DIST/docs/tutorial/Tutorial.html`