

Variability Patterns in Kinase Subfamilies

Jacek Kuśka¹, Jacek Leluk², Bogdan Lesyng^{2,3}

Keywords: mutational variability, protein diversity, correlated mutations, homology, kinases

Four homologous kinase subfamilies (hexokinases, RTKs, JAKs, cAMP dependent kinases) were studied with respect to their variability spectrum and distribution within the molecule. The sequences were taken from Swiss-Prot database. Multiple sequence alignments as well as consensus sequences for each group were created and verified with the aid of the genetic semihomology algorithm [1-3]. The resources of the database Ligand-Protein Contact and Contacts of Structural Units (<http://bip.weizmann.ac.il/oca-bin/lpccsu>) were used to detect residues being in close, direct contact in the molecule structure at the secondary as well as tertiary level. The analysis and visualization of variability in three analyzed groups was carried out with the use of three dimensional models obtained from the Protein Data Bank. Tertiary structures of JAKs representatives were created *in silico* in the Interdisciplinary Centre for Mathematical and Computational Modelling at Warsaw University. Each subfamily was also checked for the occurrence of correlated mutations. For this purpose the program Corm, developed at ICM UW, was applied.

The following observations can be formulated for the studied kinase subfamilies:

- the variability range varies from 1 to 17 residue types per position;
- a representative three-dimensional molecule can be divided into regions of relatively high and low variability. This allows to establish precisely the molecule parts of different variability characteristics, and to predict/confirm their putative physiological role;
- the residues that form the catalytic centre reveal relatively low variability level, but residues surrounding the centre may reveal significant variability and they are probably responsible for very narrow specificity of each individual enzyme;
- the N-terminal leader peptide is very variable;
- the residues which are in close, direct contact show the same or very similar degree of variability;
- many of the identified and described correlated mutations may be considerably distant in the primary structure but close in three dimensional representation. This phenomenon confirms the commonly accepted compensation mechanism. However some correlated residues are located far away from each other at both primary and tertiary structural levels. The phenomenon of the very distant residues correlation (with respect to the sequence and three dimensional model) is still unclear and needs further research;

Generally the results of this analysis confirm the hypothesis concerning the correlation among levels of variability for the residues which are in contact. It refers to the statement that proteins' structure as well as function is more conserved than the primary protein structure. On the other

¹ Faculty of Biology, Warsaw University, Warsaw, Poland. E-mail: jackk@grid.icm.edu.pl

² Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University, Poland. E-mail: lulu@icm.edu.pl

³ Department of Biophysics, Faculty of Physics, Warsaw University, Poland. E-mail: B.Lesyng@icm.edu.pl

hand the importance of relatively highly variable regions for structural and functional specificity were indicated. These data are very helpful in understanding the mechanism and tendencies leading to the diversity within the homologous protein family. They also can be used in explaining the common differentiation properties (divergent) in structures from non-homologous groups that do not resemble each other at the amino acid sequence.

References

- [1] Leluk, J. 1998. A new algorithm for analysis of the homology in protein primary structure. In: *Computers & Chemistry* 22:123-131.
- [2] Leluk, J. 2000 A non-statistical approach to protein mutational variability. In: *BioSystems* 56:83-93.
- [3] Leluk, J. 2000 Regularities in mutational variability in selected protein families and the Markovian model of amino acid replacement. In: *Computers & Chemistry*, 24:659-672