Chapter II – Building a Functional Annotated Structural Loop Database of Kinases: *ArchKI*

2.1 Abstract

The annotation of protein function has become a crucial problem with the advent of sequence and structural genomic initiatives. A large body of evidence suggests that protein structural information is frequently encoded in local sequences, and that folds are mainly made up of a number of simple local units of super-secondary structural motifs, consisting of a few secondary structures and their connecting loops. Moreover, protein loops play an important role in protein function. Here we present ArchKI is a classification database of kinases loops with information of functional residues. Functional information have been extracted from three different sources: (i) functional residues (or regions) described in the literature, (ii) residues described in the SITE records of PDB files that specify residues comprising catalytic, cofactor, regulatory or other important sites and (iii) residues in contact with heteroatoms. ArchKI currently contains 1813 super-secondary elements classified into 133 motif subclasses. Functional loops such as P-loop or Gly-rich loop among others, were classified into structural motifs. The database provides an easy way to retrieve functional information from protein structures sharing a common motif, to search motifs found in a given SCOP family, super-family or fold, or to search by keywords on proteins with classified loops. The ArchKI database of loops is located at http://sbi.imim.es/archki.

2.2   Introduction

Loops are regions of non-repetitive conformation connecting regular secondary structures. There have been many attempts to classify loops, presenting topological clusters and consensus sequences(Efimov 1993; Kwasigroch et al. 1996; Wintjens et al. 1996; Oliva et al. 1997; Oliva et al. 1998; Burke et al. 2000). The reports of Salem et al.(Salem et al. 1999) and Wood and Pearson(Wood and Pearson 1999) suggested that folds are mainly made up of a number of simple local units of super-secondary structural motifs, formed by a few secondary structures connected by loops. An elementary super-secondary motif can be defined as one loop plus its bracing secondary structures. In particular, loops play an important role in the local conformation(Yang and Wang 2002) of the protein and are often related to its function.

Structural genomic initiatives attempt to infer details of protein function via 3D structure determination(Eisenberg et al. 2000; Shapiro and Harris 2000). If a new protein structure adopts a previously observed fold, then functional details might be inferred by considering the function of other proteins adopting the same fold(Murzin 1996; Russell et al. 1997; Dietmann et al. 2001; Dietmann et al. 2002). If fold similarities are ambiguous or if a protein adopts a new fold, it is still possible to infer function by comparison of key active site residues(Russell et al. 1998; Hegyi and Gerstein 1999). Common detected structural motifs contain particularly useful information on the conservation of specific residues across species, being occasionally involved in the protein function (i.e. the activation loop of some kinases) or in the folding nucleus(Mirny and Shakhnovich 2001). Moreover, loops are often the most difficult structures to model(Burke et al. 2000; Fiser et al. 2000) and thus a database of structurally classified protein loops will have widespread applications (i.e. in model building or to complete locally undefined regions from an X-ray diffraction map).

2.3   Material and Methods

2.3.1 Set of analyzed kinases

A database of protein kinases was recently published by Cheek et al.(Cheek et al. 2002). The initial database of structures of protein of kinases was extracted from the PDBFINDER database (April 2002)(Hooft et al. 1996) where structures with kinase or phosphotransferase function, with assigned Enzyme Commission (EC) number (2.7.X.X)(Kotyk 1999), were chosen. Sequence redundant proteins were removed and the protein coordinates were downloaded from the PDB database(Westbrook et al. 2003). Our search yielded 145 protein domains with kinase activity. These kinases belonged to 18 different folds, 20 different super-families and 39 different families (according to SCOP release 1.61)(Lo Conte et al. 2002) and represented a wide repertory of well-known kinases. The initial set of motifs was filtered in order to remove loops with low quality resolution and redundant. First, super-secondary structures poorly diffracted were removed. Second, super-secondary structures with 100% sequence identity and identical conformation in $[\phi,\psi]$ space were reduced to a single loop (taken from the protein structure with best resolution).

2.3.2   Clustering of motifs

The structural classification of kinase loops was obtained with an improved version(Espadaler et al. 2004) of the Arch-Type program(Oliva et al. 1997). We have used the loop clustering program Arch-Type to derive a fully automated loop classification of clusters with more than two loops. The algorithm clustering is based on a density search on the (ϕ,ψ) space of the loop conformation, henceforth

allowing for a second check by RMSD. Clusters were arranged as in the previous work. In the lowest level of the classification, structural motifs were grouped according to their geometry (motif *subclass* level).

At higher levels, motifs were grouped according to the loop size and (ϕ,ψ) conformation (motif *class* level). At the top of the classification, motifs were identified according to bracing secondary structure type ($\alpha\alpha$, $\beta\beta_{links}$, $\beta\beta_{hairpins}$, $\alpha\beta$ and $\beta\alpha$, the motif *type* level). At the *class* level, loops of similar size, with differences of ±1 residues, were allowed to cluster together to deal with the lax definition of the secondary structure ends.

Owing to the ±1 extension and to the wide definition around (ϕ,ψ) regions in l/g and in b/p conformations, loops were allowed to cluster into more than one group. A re-clustering protocol has been devised to deal with the overlap between clusters. Overlapping clusters are merged depending on the percentage of shared loops. The result is an optimized partition of the conformational space of loops that groups clusters (as obtained in Arch-Type(Oliva et al. 1997)) and containing at least two loops) into *subclasses* with the largest number of loops and the minimum overlap.

Finally, the averaged RMSD between loops in each *subclass* was checked in order to corroborate this procedure (see figure 2.3). Each *subclass* is identified in ArchKI by a three-number code as defined in the original paper(Oliva et al. 1997).

2.3.3 Assignation of function to loop sub-classes

Three different approaches were applied to identify functional residues of the loops of the sub-classes: (i) residues found within a cut-off distance of 6Å from an heteroatom, ligand, inhibitor, cofactor or complex partner molecule (protein or DNA), with the exception of D2O or crystallization buffer molecules; (ii) residues identified by functional information from ACTSITE and SITE records in the protein data bank(Westbrook et al. 2003); and (iii) residues identified by the functional annotation collected from the literature and assigned to specific motifs. See flowchart for assignation of function in figure 2.1.

Figure 2.1. Flowchart for functional annotation of loop sub-class. Functional residues were identified by literature, "PDB sites" or closeness to ligands and were mapped on the profile of aligned loops. The degree of conservation of functional residues and the percentage of loops from the same super-family of SCOP (more than 50% of loops of the sub-class) was used to assign a likely function for the sub-class.
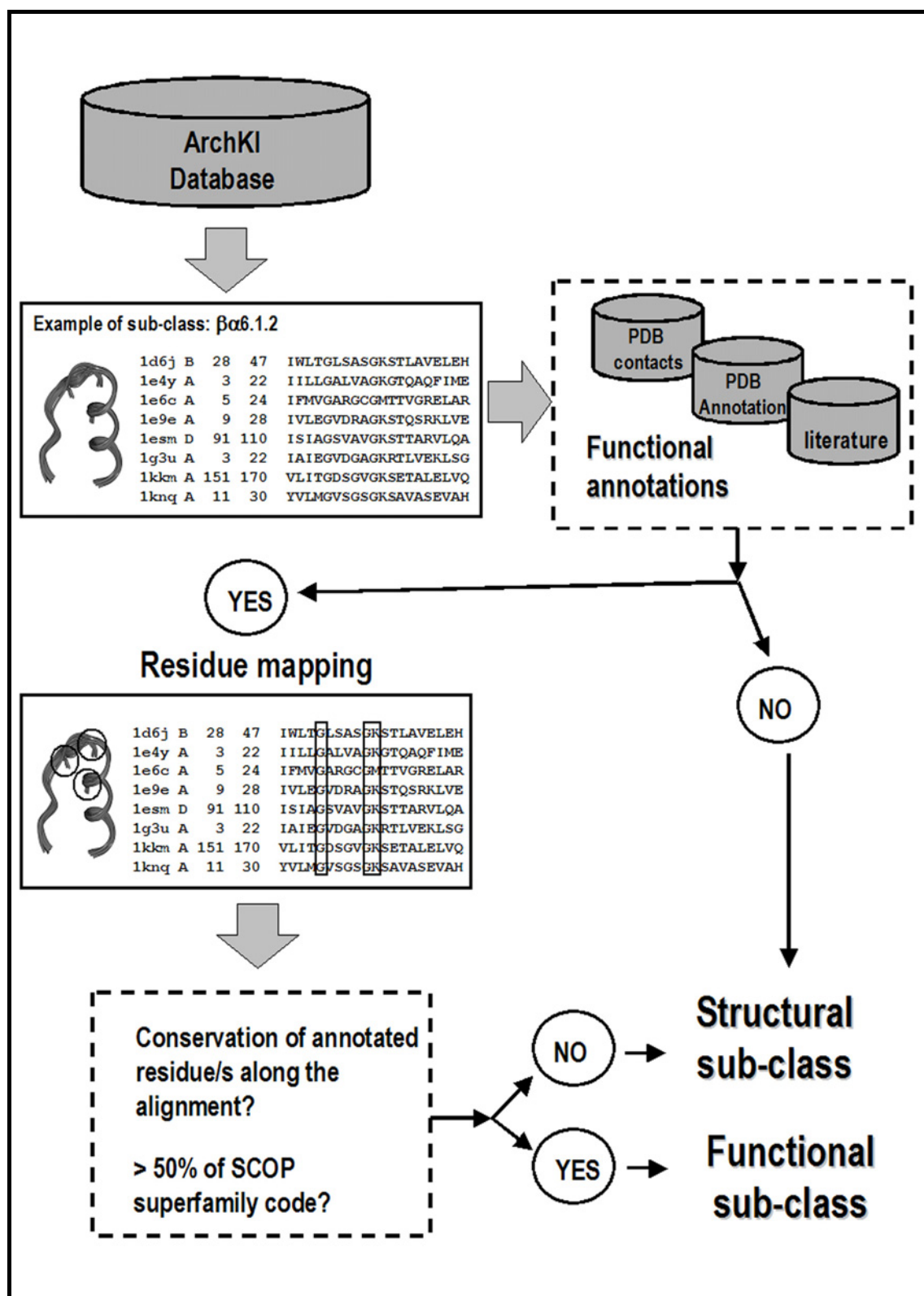
Figure 2.1 (legend opposite)

2.3.4 Web Server

The results obtained were presented in a web server (http://sbi.imim.es/archki). All the data was stored in *MySQL* tables and we used DBI-DBD (DataBase Interface-DataBase Driver) and related modules for communication between the scripts and the *MySQL* database server. We used a CGI (Common Gateway Interface) module to create the HTML (HyperText Markup Language) output. Users can browse the database or query it to: (i) search for structure motifs of a PDB structure by specifying the PDB identifier or SWISS-PROT accession code; (ii) browse through complete database levels; (iii) retrieve structural motifs satisfying some features (i.e. bracing secondary structure type, loop size and loop (ϕ,ψ) conformation); (iv) search for structural motifs of a SCOP family/super-family/fold; (v) search for SWISS-PROT keywords or Gene Ontology (GO) accession codes; (vi) search for motifs simultaneously found in two different PDB structures; (vii) looking for sub-classes or loops with bibliographical annotations, with residues in contact with ligands or with PDB annotation.

For each sub-class it is possible to acquire information about consensus sequence, conformation and geometry, with a link to the multiple alignment of sequences and secondary structures (as defined by DSSP(Kabsch and Sander 1983)) of the motifs. Additional information includes the average percentage of sequence identity, averaged RMSD of main-chain atoms, contacts between residues and heteroatoms within a range of 6Å, bibliographical annotations and functional information detailed at ACTSITE and SITE records (identified in the PDB file).

A PROSITE-like pattern was generated for every sub-class and the program AL2CO(Pei and Grishin 2001) was used to quantify the degree of conservation for each position in the multiple alignment. A PSSM profile was derived from the sequence multiple alignment. The Henikoff and Henikoff(Henikoff and Henikoff 1994) position-based method was used to weight the sequences. The structures involved

on different subclasses can be viewed using RASMOL or CHIME. Additional structural and functional information for each structure is also given, including resolution, R-factor, PDB source, SWISS-PROT keywords(Boeckmann et al. 2003), GO annotation(Ashburner et al. 2000), Enzyme annotation(Kotyk 1999) and SCOP domain classification(Lo Conte et al. 2002).

2.4 Results

2.4.1 General Overview

A total of 141 protein chains (145 SCOP domains) were used for the protein-kinase database (see the complete dataset in http://sbi.imim.es/archki -supplementary material-). The structures analyzed included complexes of kinases and yielded a total of 2755 motifs (693 $\alpha\alpha$, 682 $\alpha\beta$, 767 $\beta\alpha$, 368 $\beta\beta_{hairpins}$ and 245 $\beta\beta_{links}$) after removal of redundancies (motifs with the same sequence and the same structure). The clustering yielded a total of 139 classes (40 $\alpha\alpha$, 36 $\alpha\beta$, 30 $\beta\alpha$, 21 $\beta\beta_{hairpins}$ and 12 $\beta\beta_{links}$) with 1237 clustered loops. Classes were subdivided according to the bracing geometry, yielding 203 sub-classes (a summary is presented in Table 1.1 with 51 $\alpha\alpha$, 65 $\alpha\beta$, 46 $\beta\alpha$, 29 $\beta\beta_{hairpins}$ and 12 $\beta\beta_{links}$). Each *subclass* contains a minimum of two loops.

Table 2.1. Summary of clustered loops, classes and sub-classes by motif type.

Motif type	Number of loops	Clustered loops	Classes	Sub-classes
$\alpha\alpha$	693	243	40	51
$\alpha\beta$	682	417	36	65
$\beta\alpha$	767	288	30	46
$\beta\beta_{hairpins}$	368	218	21	29
$\beta\beta_{links}$	245	71	12	12
TOTAL	**2755**	**1237**	**139**	**203**

A total of 18 folds, 20 super-families and 30 families (as defined in SCOP v1.61) have a representative in ArchKI. Lengths of clustered motifs ranges from 0 to 30 residues long (as DSSP definition(Kabsch

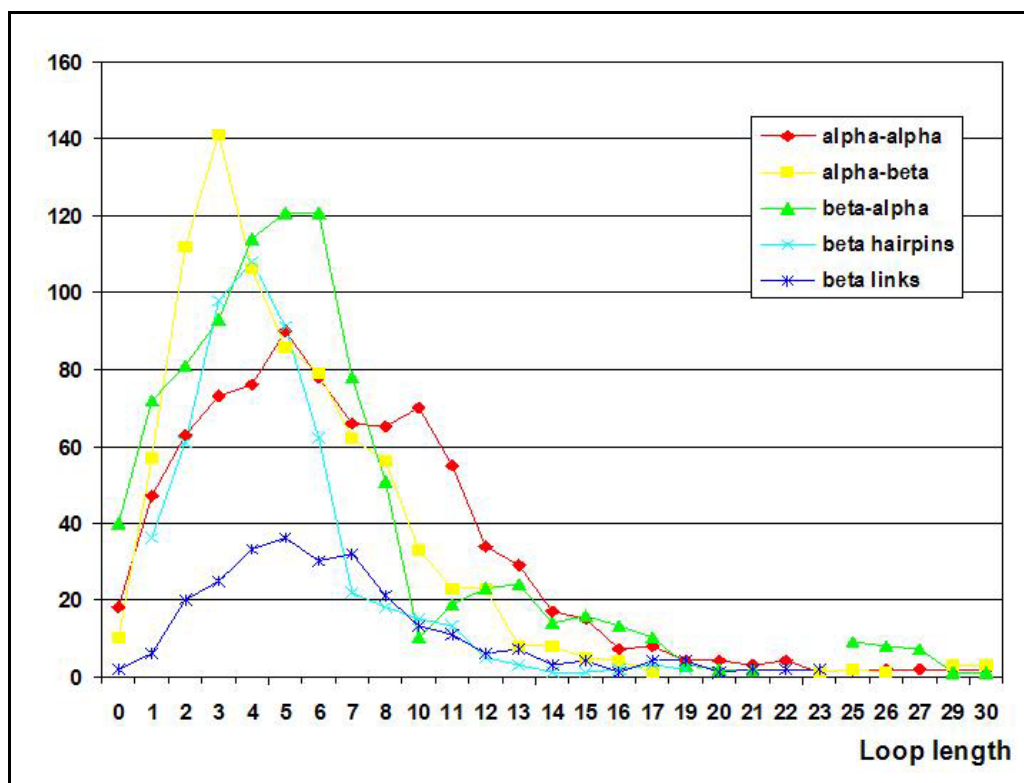and Sander 1983)). Figure 2.2 shows the distribution of clustered motifs by loop length and type of motif.

Figure 2.2. Distribution of motifs in relation to loop length. The loop length is the number of residues between the ends of the secondary structures as defined by Kabsch & Sander(Kabsch and Sander 1983).

Since Arch-Type clusters motifs based on a density search on the (ϕ,ψ) space of the loop conformation, a second check based in RSMD could be made in order to assess solidity and quality of the structural sub-classes. Figure 2.3 shows the values of averaged RMSD by loop length. All the RMSD values are below 1.2 Å, indicating a good compactness of sub-classes for all range of loop lengths.
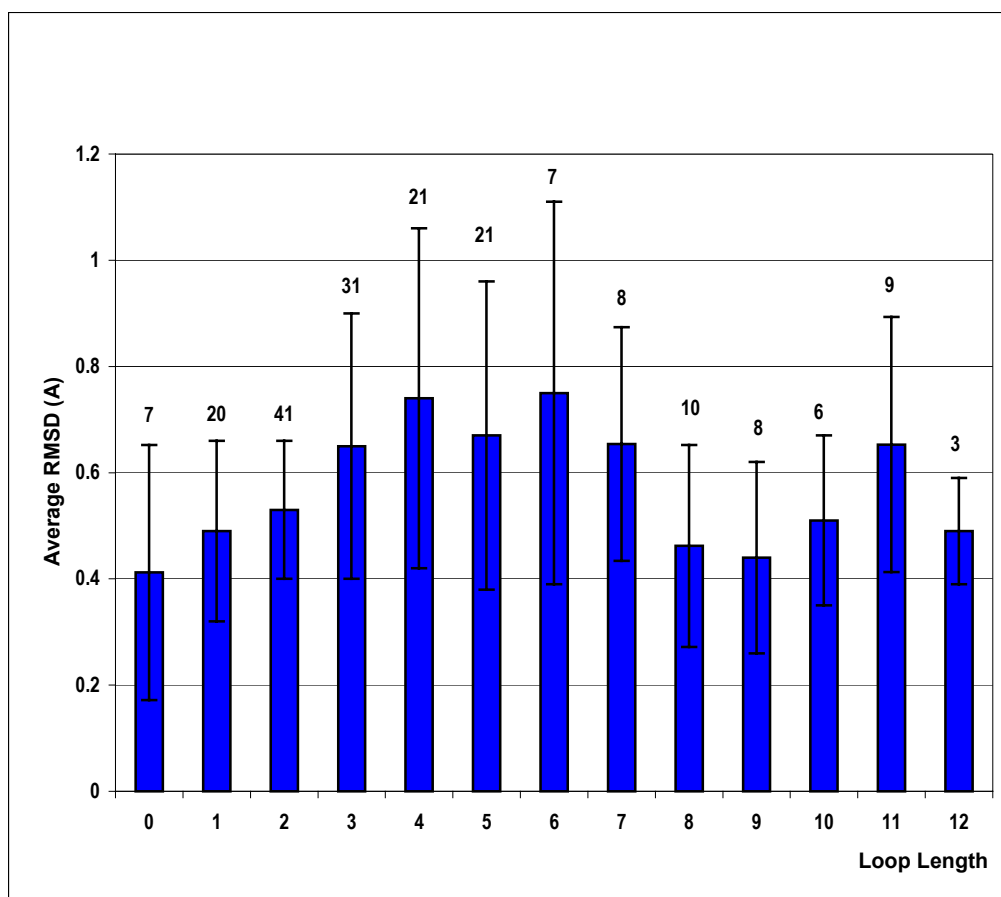
Figure 2.3. Averaged RMSD between loops in subclasses. The averaged RMSD of the sets of loop structures on each subclass was calculated with the main-chain atoms of the residues in the loop plus two bracing residues at each side. Additional extensions of the bars show the standard deviations of the averages. The number of sub-classes is shown at the top.

2.4.2 Web Database

The results obtained are presented in a web server (http://sbi.imim.es/archki) (see figure 2.4). For each

sub class it is possible to retrieve information about consensus sequence and consensus (ϕ, ψ)

conformation and a link to the multiple alignment of sequences and conformations of the loops forming

the sub class. Additional information includes the average percentage of sequence identity, RMSD of

main-chain atoms and functional information of residues (see material and methods). The PSSM profile

may help to search a sub-class or motif against a target sequence or for the comparative modeling of

loops of kinases. Additionally, the superimposition of the loops of the sub-classes can be viewed using

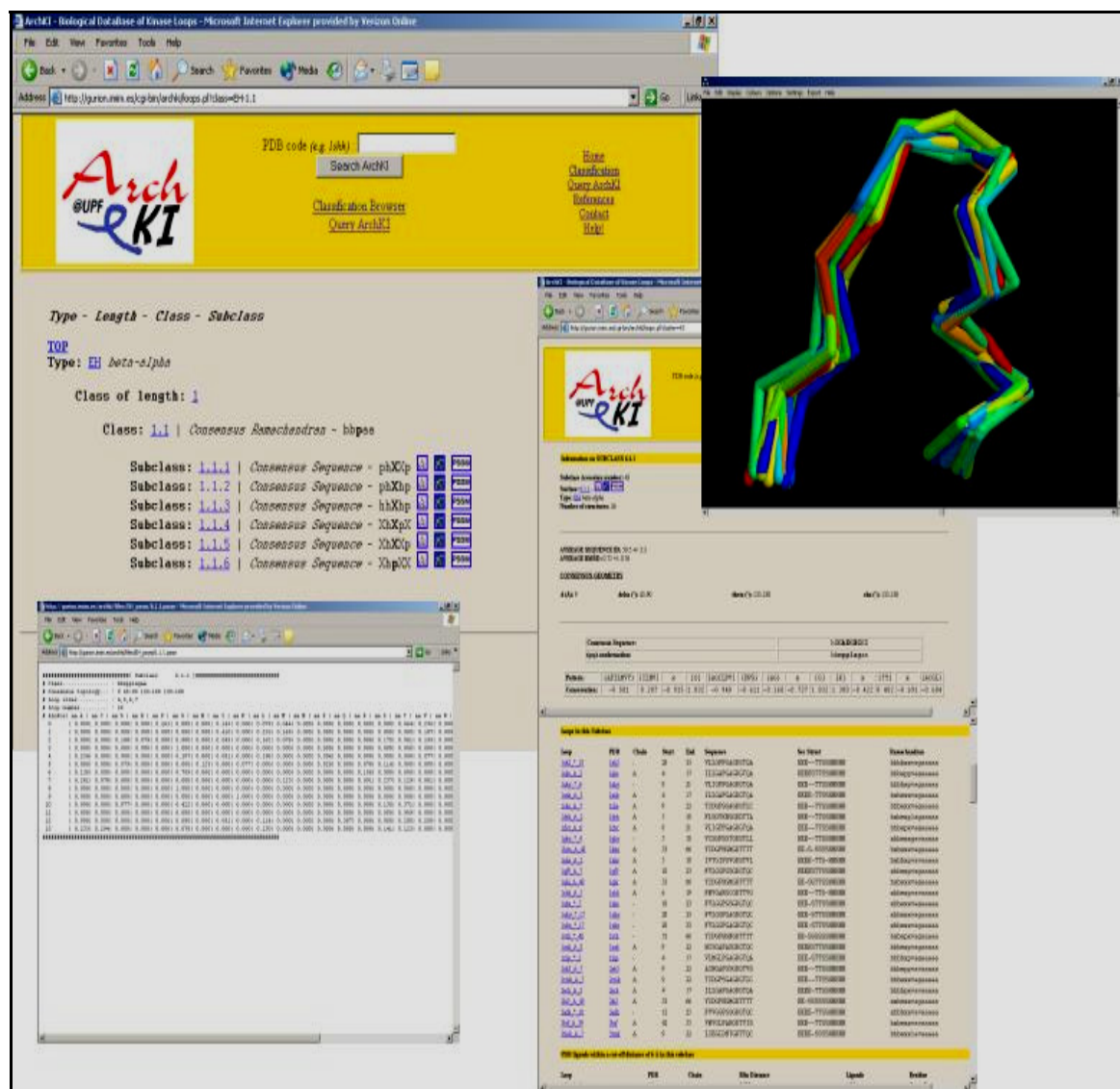RASMOL(Sayle and Milner-White 1995) or CHIME.

Figure 2.4. Screenshot of ArchKI information HTML pages. Classification browser with subclass information, multiple alignments of sequence and conformation, the profile pattern, and the image of structurally superimposed motifs as viewed with Rasmol(Sayle and Milner-White 1995).

2.5   Discussion and Conclusions

We have presented an improved and full automatic process of classification of loops. The new web site allows quick and easy access to the complete classification of loops within ArchKI database. The structural superposition of all of the loops is available for viewing. This permits inspection of not only the sequence conservation for each loop class or sub-class but also the conservation of ramachandran conformation and hydrogen bonding. Furthermore, ArchKI is linked to other biological database, such as SWISS-PROT(Boeckmann et al. 2003), GO(Ashburner et al. 2000), and SCOP(Murzin et al. 1995) giving to the user additional and useful information. Furthermore, functional information about loops is given.

The two major motivations for this study is to help in predicting loop conformation in comparative modeling and to make available a curated functional annotated loop structural classification of protein kinases. On one hand, we provide a single consistent source that classifies the conformation of standard loops with their associated sequences patterns and a PSSM profile for each structural alignment; together with the ability to search ArchKI database, this provides a powerful tool to aid the modeling of the loops. On the other hand, functional annotated sub-classes may help in the central problem of protein annotation derived of with the explosive increase of available sequences and structures as a result of sequencing projects and structural genomic initiatives.

2.6 References

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "*Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.*" Nat Genet **25**(1): 25-9.

Boeckmann, B., A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout and M. Schneider (2003). "*The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.*" Nucleic Acids Res **31**(1): 365-70.

Burke, D., C. Deane and T. Blundell (2000). "*Browsing the Sloop database of structurally classified loops connecting elements of protein secondary structure.*" Bioinformatics **16**: 513-516.

Cheek, S., H. Zhang and N. V. Grishin (2002). "*Sequence and structure classification of kinases.*" J Mol Biol **320**(4): 855-81.

Dietmann, S., F.-F. N. and L. Holm (2002). "*Automated detection of remote homology.*" Curr Opin Struct Biol **12**: 362-367.

Dietmann, S., J. Park, C. Notredame, A. Heger, M. Lappe and L. Holm (2001). "*A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3.*" Nucl Acid Res **29**: 55-57.

Efimov, A. V. (1993). "*Patterns of loops regions in proteins.*" Curr Opin Struct Biol **3**: 379-384.

Eisenberg, D., E. Marcotte, I. Xenarios and T. Yeates (2000). "*Protein function in the post-genomic era.*" Nature **405**: 823-826.

Espadaler, J., N. Fernandez-Fuentes, A. Hermoso, E. Querol, F. X. Aviles, M. J. Sternberg and B. Oliva (2004). "*ArchDB: Automated protein loop classification as a tool for Structural Genomics.*" Nucleic Acids Res **32**: D185-D188.

Fiser, A., R. K. Do and A. Sali (2000). "*Modeling of loops in protein structures.*" Protein Sci **9**: 1753-1773.

Hegyi, H. and M. Gerstein (1999). "*The relationship between protein structure and function: a comprehensive survey with application to the yeast genome.*" J.Mol.Biol. **288**: 147-164.

Henikoff, S. and J. G. Henikoff (1994). "*Position-based sequence weigths.*" J Mol Biol **243**(4): 574-578.

Hooft, R. W., C. Sander, M. Scharf and G. Vriend (1996). "*The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value.*" Comput Appl Biosci **12**(6): 525-9.

Kabsch, W. and C. Sander (1983). "*Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.*" Biopolymers **22**(12): 2577-637.

Kotyk, A. (1999). "*IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB). Newsletter 1999.*" Folia Microbiol (Praha) **44**(3): 243-6.

Kwasigroch, J. M., J. Chomilier and J. P. Mornon (1996). "*A global taxonomy of loops in globular proteins.*" J. Mol. Biol. **259**: 855-872.

Lo Conte, L., S. E. Brenner, T. J. Hubbard, C. Chothia and A. G. Murzin (2002). "*SCOP database in 2002: refinements accommodate structural genomics.*" Nucleic Acids Res **30**(1): 264-7.

Mirny, L. and E. Shakhnovich (2001). "*Evolutionary Conservation of the Folding Nucleus.*" J.Mol.Biol. **308**: 123-129.

Murzin, A. (1996). "*Structural classification of proteins: new superfamilies.*" Curr. Opin. Struct. Biol. **2**: 895-903.

Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia (1995). "*SCOP: a structural classification of proteins database for the investigation of sequences and structures.*" J Mol Biol **247**(4): 536-40.

Oliva, B., P. A. Bates, E. Querol, F. X. Aviles and M. J. Sternberg (1997). "*An automated classification of the structure of protein loops.*" J Mol Biol **266**(4): 814-30.

Oliva, B., P. A. Bates, E. Querol, F. X. Aviles and M. J. Sternberg (1998). "*Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction.*" J Mol Biol **279**(5): 1193-210.

Pei, J. and N. V. Grishin (2001). "*AL2CO: calculation of positional conservation in a protein sequence alignment.*" Bioinformatics **17**(8): 700-12.

Russell, R., M. Saqi, R. Sayle, P. Bates and M. J. E. Sternberg (1997). "*Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation.*" J.Mol.Biol. **269**: 423-439.

Russell, R., P. Sasieni and J. M. E. Sternberg (1998). "*Supersites within superfolds. Binding site similarity in the absence of homology.*" J.Mol.Biol. **282**: 903-918.

Salem, G. M., E. G. Hutchinson, C. A. Orengo and J. M. Thornton (1999). "*Correlation of observed fold frequency with the occurrence of local structural motifs.*" J Mol Biol **287**(5): 969-81.

Sayle, R. A. and E. J. Milner-White (1995). "*RASMOL: biomolecular graphics for all.*" Trends Biochem Sci **20**(9): 374.

Shapiro, L. and T. Harris (2000). "*Finding function through structural genomics.*" Curr. Opin. Biotechnol. **11**: 31-35.

Westbrook, J., Z. Feng, L. Chen, H. Yang and H. M. Berman (2003). "*The Protein Data Bank and structural genomics.*" Nucleic Acids Res **31**(1): 489-91.

Wintjens, R. T., M. J. Rooman and S. J. Wodak (1996). "*Automatic classification and analysis of alpha alpha-turn motifs in proteins.*" J Mol Biol **255**(1): 235-53.

Wood, T. and W. Pearson (1999). "*Evolution of protein sequences and structures.*" J Mol Biol **291**: 977-995.

Yang, A. S. and L. Y. Wang (2002). "*Local structure-based sequence profile database for local and global protein structure predictions.*" Bioinformatics **18**(12): 1650-7.

Chapter III - Mining ArchKI: On the Common and Characteristic Motifs with Functional Information.

3.1   Abstract

A structural classification of loops of kinases has been analyzed. Functional loops such as P-loop, Gly-rich-loop or activation loop of Ser-Thr kinases classified into structural motifs and studied . As a result, a common catalytic mechanism and substrate binding is proved for most kinases. A evolutionary relationship between Phospho-enol piruvate (PEP) carboxykinase and the family of HPr kinase through the P-loop superimposition suggested by Russell et al. (Russell et al. 2002), has been also revealed, validating the use of conserved structural patterns and it extension by neighbors to characterize a common evolutionary ancestor. We have recognized the location of important residues by conservation in sequence and structure among kinases, determining the relevance of particular residues. Additionally, the multiple-alignment of loop sequences made within each sub-class has been proved to be useful for comparative modeling of kinase loops.

3.2 Introduction

Deciphering protein function is becoming a major problem with the increase of newly sequenced genomes. The increase of accessible data has shown proteins with common folds for unrelated sequences and proteins with common functions shared by different folds. However, a link between sequence, structure and function exists. Among the important regions in a fold are those formed by motifs constituted by two secondary structures connected by a loop.

Conserved short stretches of amino acid sequences or motifs contain useful information on the conservation of specific residues involved in the protein function (catalysis or binding) or in the folding nucleus(Russell 1998; Copley et al. 2001; Lupas et al. 2001; Mirny and Shakhnovich 2001). Loops represent an important type of non-regular structure. They are commonly defined as non-periodic structures connecting two adjacent periodic secondary structural elements, $\alpha-$helices and $\beta-$strands(Venkatachalam 1968; Rose et al. 1985). The best characterized loop structures are short and geometrically well defined. Subsequent analyses did study the loop conformations for different types of protein motifs(Efimov 1991; Efimov 1994). Several classifications of protein loops, presenting topological clusters and consensus sequences were derived by automated methods(Donate et al. 1996; Wintjens et al. 1996; Oliva et al. 1997; Oliva et al. 1998). In particular, our previous work with the program Arch-Type was used to automatically classify a set of non redundant proteins(Oliva et al. 1997) and a special set of complementary determining regions of the heavy chain of immunoglobulins(Oliva et al. 1998).

The aim of the present work is to carry out an exhaustive analysis of a structural classification of a particular set of proteins with biological relevance containing some well studied functional motifs, for

which the set of protein-kinases has been chosen. In the present work, the relation on the set of kinases is described by analyzing these partial regions of their structures with an automated procedure derived from the program Arch-Type(Oliva et al. 1997). Protein-kinase loops have been classified and analyzed using our recently improved method(Espadaler et al. 2004). Functional residues (or amino-acid stretches) described in the literature were used to study the relationship between structure and function of the classified kinase-loops.

Kinases represent a ubiquitous group of enzymes that catalyze the transference of a phosphate group from one nucleotide to another, or to a small molecule or to a protein. Protein kinases represent one of the largest protein groups. For example, in the sequenced budding yeast genome, conventional protein-kinases genes represent about 2% of the total genome(Hunter and Plowman 1997) and in the recently sequenced human genome a functional annotation performed by Wright et al.(Wright et al. 2001) found that the largest functional group is related to the phosphor transfer and protein kinases. Phosphorylation was first identified as a mechanism for regulating protein activity in the 1950s by Fischer and Krebs(Fischer and Krebs 1955). At present, kinases have been described in many biochemical processes and probably phosphorylation is the most important mechanism for regulation in mammalian cells. Indeed, kinases are involved on signal transference pathways, in the regulation of cell growth, cell division, cell mobility, metabolism, membrane transport, gene expression, learning and memory(Johnson et al. 1998). Moreover, they represent, after proteases, a major subset of the known current drug targets.

3.3   Results

3.3.1 General Overview

A total of 141 protein chains (145 SCOP domains) were used to build the protein-kinase database (see chapter 1). The clustering yielded a total of 139 classes (40 $\alpha-\alpha$, 36 $\alpha-\beta$, 30 $\beta-\alpha$, 21 $\beta\beta_{hairpins}$ and 12 $\beta\beta_{links}$) and 203 sub-classes (a summary is presented in Table 3.1 with 51 $\alpha\alpha$, 65 $\alpha\beta$, 46 $\beta\alpha$, 29 $\beta\beta_{hairpins}$ and 12 $\beta\beta_{links}$), where 76 out of 203 sub-classes contained residues with functional annotation collected from the literature (see chapter 1). Functional annotation of residues was classified in four categories: (i) ATP binding: for residues involved on Adenosine Triphosphate (ATP) binding; (ii) substrate binding: for residues involved in substrate binding with the exception of ATP; (iii) ion interaction : for residues involved in binding of ions needed for the catalytic mechanism; (iv) catalytic: for residues involved in the catalytic reaction or the stabilization of a transition state. We considered sub-classes as "functional sub-classes", when there is a meaningful conservation of functional residues in the loops of the cluster and more than 50% of its loops belong to the same SCOP super-family.

Table 3.1. Summary of clustered loops, classes and sub-classes.

Loop type	Number of loops	Clustered loops	Classes	Sub-classes	Functional sub-classes[a]
$\alpha\alpha$	693	243	40	51	19 (11)
$\alpha\beta$	682	417	36	65	15 (9)
$\beta\alpha$	767	288	30	46	30 (11)
$\beta\beta_{hairpins}$	368	218	21	29	8 (3)
$\beta\beta_{links}$	245	71	12	12	4 (0)
TOTAL	2755	1237	139	203	76 (34)

[a]The number of sub-classes with motifs with functional annotation is indicated in the last column. Sub-classes where the functional residues of the motif were located in the secondary structure are indicated in parenthesis.

About 37% of the sub-classes obtained presented residues with functional annotation in the literature (see Table 3.2), although some functional residues were located in loops and some others in regular structures (i.e. a conserved Glu that forms a salt bridge with a Lys or Arg in the protein-kinase like fold(Kobe et al. 1996; Sicheri et al. 1997) is located in the Nt secondary structure of the sub classes $\beta\alpha8.1.1$ and $\beta\alpha8.2.1$ *).*

 For ATP binding we found 6 $\alpha\alpha$ sub-classes, 3 $\alpha\beta$ sub- classes, 16 $\beta\alpha$ sub-classes, 6 $\beta\beta_{hairpin}$ sub-classes and 2 $\beta\beta_{links}$ sub-classes; for substrate binding, we found 9 $\alpha\alpha$ sub-classes, 6 $\alpha\beta$ sub-classes and 11 $\beta\alpha$ sub-classes.

For ion binding we found 2 $\alpha\alpha$ sub-classes, 2 $\alpha\beta$ sub-classes and 2 $\beta\alpha$ sub-classes and for dimerization or oligomerization we found 1 $\alpha\alpha$ sub-class and 1 $\alpha\beta$ sub-class (see Table 3.2). Finally, for the catalytic mechanism we found 1 $\alpha\alpha$ sub-class, 3 $\alpha\beta$ sub-classes, 1 $\beta\alpha$ sub-classes, 2 $\beta\beta_{hairpin}$ sub-classes and 2 $\beta\beta_{links}$ sub-classes.

Clearly, motifs in $\beta\alpha$ show the largest percentage of functional sub-classes. Further analyses show that most of $\beta\alpha$ functional sub-classes were involved on substrate and ATP binding.

A total of 34 out of 76 subclasses with loops implicated on a function had their functional residues in the bracing secondary structures. However, most motifs involved on ATP binding contained the functional residues in the coiled region, while loops involved in substrate binding and catalytic mechanism indistinctly held the functional residues in one of the secondary structures or the loop coil.

Table 3.2. Sub-classes of loops with functional information.

Function	Sub-class	Scop code domain	cf (%)	sf (%)	fa (%)
Substrate binding	αα1.1.1	d.58.6.1	54861 (80)	54919 (80)	54920 (80)
	αα1.2.1	d.58.6.1	54861 (100)	54919 (100)	54920 (100)
	αα2.2.1	c.37.1.1	52539 (100)	52540 (100)	52641 (100)
	αα3.1.1	c.37.1.1	52539 (100)	52540 (100)	52641 (100)
	αα4.1.2	c.73.1.1	53747 (66)	53633 (66)	53634 (66)
	αα4.1.1	c.37.1.1	52539 (100)	52540 (100)	52641 (100)
	αα5.1.1	c.37.1.1	52539 (93)	52540 (93)	52641 (93)
	αα5.3.1	c.37.1.1	52539 (100)	52540 (100)	52641 (100)
	αα12.1.1	d.58.6.1	54861 (100)	54919 (100)	54920 (100)
	αβ1.2.2	c.37.1.1	52539 (100)	52540 (100	52641 (100)
	αβ3.1.7	c.73.1.1	53747 (100)	53633 (100)	53634 (100)
	αβ3.4.2	c.55.1.3	53066 (100)	53067 (100)	53083 (100)
	αβ5.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)
	αβ5.3.1	c.49.1.1	52934 (100)	52935 (100)	52936 (100)
	αβ7.2.1	c.1.12.1	51350 (100)	51621 (100)	51622 (100)
	βα0.1.1	c.37.1.1 c.37.1.2 c.37.1.7	52539 (88) 52539 (88) 52539 (88)	52540 (88) 52540 (88) 52540 (88)	52541 (78) 52566 (6) 52589 (6)
	βα1.1.3	d.58.6.1	54861 (57)	54919 (57)	54920 (57)
	βα1.1.5	c.37.1.3	52539 (50)	52540 (50)	52569 (50)
	βα2.3.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)
	βα2.4.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)
	βα3.1.1	c.49.1.1	52934 (100)	52935 (100)	52936 (100)
	βα5.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)
	βα12.1.1	c.86.1.1	53747 (100)	53748 (100)	53749 (100)
	βα14.1.1	c.1.12.1	51350 (100)	51621 (100)	51622 (100)
	βα16.1.1	c.55.1.3	53066 (100)	53067 (100)	53083 (100)
	βα17.1.1	c.55.1.3	53066 (100)	53067 (100)	53083 (100)

Function	Sub-class	Scop code domain	cf (%)	sf (%)	fa (%)
Catalytic	αα4.7.1	c.86.1.1	53747 (100)	53748 (100)	53749 (100)
	αβ8.1.1	d.144.1.1 d.144.1.2	56111 (100) 56111 (100)	56112 (100) 56112 (100)	56113 (68) 56150 (32)
	αβ8.2.1	d.144.1.1 d.144.1.2	56111 (100) 56111 (100)	56112 (100) 56112 (100)	56113 (50) 56150 (50)
	αβ11.1.1	d.144.1.1 d.144.1.2	56111 (100) 56111 (100)	56112 (100) 56112 (100)	56113 (63) 56150 (37)
	βα4.2.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)
	ββhairpin3.3.1	d.144.1.2	56111 (100)	56112 (100)	56150 (100)
	ββhairpin5.3.1	d.128.1.2	55930 (100)	55931 (100)	55935 (100)
	ββlink5.1.1	d.144.1.1 d.144.1.2	56111 (100) 56111 (100)	56112 (100) 56112 (100)	56113 (82) 56150 (18)
	ββlink8.1.1	d.144.1.1 d.144.1.2 d.144.1.6	56111 (100) 56111 (100) 56111 (100)	56112 (100) 56112 (100) 56112 (100)	56113 (77) 56150 (18) 64411 (5)

Function	Sub-class	Scop code domain	cf (%)	sf (%)	fa (%)
Ion binding	αα2.1.1	c.1.12.1	51350 (100)	51621 (100)	51622 (100)
	αα5.3.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)
	αβ2.1.1 αβ2.1.1	c.37.1.1 c.37.1.2	52539 (69) 52539 (69)	52540 (69) 52540 (69)	52541 (61) 52566 (4)
	αβ7.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)
	βα0.1.1	c.37.1.2	52539 (88)	52540 (88)	52541 (88)
	βα0.1.2	c.1.12.1	51350 (100)	51621 (100)	51622 (100)

Function	Sub-class	Scop code domain	cf (%)	sf (%)	fa (%)
Dimerization	αα8.3.1	a.83.1.1	48033 (100)	48034 (100)	48035 (100)
	αβ10.1.1	d.128.1.2	55930 (100)	55931 (100)	55935 (100)

Table 3.2 (Continued)

Function	Sub-class	Scop code domain	cf (%)	sf (%)	fa (%)	Function	Sub-class	Scop code domain	cf (%)	sf (%)	fa (%)
ATP binding	αα1.1.2	d.144.1.1	56111 (100)	56112 (100)	56113 (100)	ATP binding					
	αα1.2.1	d.58.6.1	54861 (100)	54919 (100)	54920 (100)		βα6.2.1	c.37.1.1	52539 (80)	52540 (80)	52541 (40)
	αα3.2.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)			c.37.1.2	52539 (80)	52540 (80)	52566 (10)
	αα9.4.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)			c.37.1.4	52539 (80)	52540 (80)	52572 (10)
	αα10.2.1	c.86.1.1	53747 (100)	53748 (100)	53749 (100)			c.37.1.6	52539 (80)	52540 (80)	52584 (10)
	αα12.1.1	d.58.6.1	54861 (100)	54919 (100)	54920 (100)			c.37.1.17	52539 (80)	52540 (80)	75195 (10)
	αβ0.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (100)			c.91.1.2	53794 (20)	53795 (20)	64186 (20)
	αβ8.4.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)		βα6.3.1	d.144.1.1	56111 (100)	56112 (100)	56113 (43)
	αβ10.2.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)			d.144.1.2	56111 (100)	56112 (100)	56150 (57)
	βα0.1.4	d.144.1.1	56111 (100)	56112 (100)	56113 (100)		βα6.3.2	d.144.1.2	56111 (100)	56112 (100)	56150 (100)
	βα1.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (95)		βα7.1.1	c.1.12.1	51350 (86)	51621 (86)	51622 (86)
		c.37.1.2	52539 (100)	52540 (100)	52566 (5)						
	βα1.2.1	c.86.1.1	53747 (88)	53748 (88)	53749 (88)		βα7.2.1	d.144.1.1	56111 (100)	56112 (100)	56113 (50)
								d.144.1.2	56111 (100)	56112 (100)	56150 (50)
	βα3.2.1	d.128.1.2	55930 (100)	55931 (100)	55935 (100)		βα8.1.1	c.86.1.1	53747 (100)	53748 (100)	53749 (100)
	βα4.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (95)		βα11.2.2	d.128.1.2	55930 (100)	55931 (100)	55935 (100)
		c.37.1.2	52539 (100)	52540 (100)	52566 (5)						
	βα4.3.1	c.37.1.1	52539 (100)	52540 (100)	52541 (75)		ββhairpin2.2.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)
		c.37.1.17	52539 (100)	52540 (100)	75195 (25)		ββhairpin2.2.2	d.144.1.1	56111 (100)	56112 (100)	56113 (100)
	βα5.2.1	c.1.12.1	51350 (100)	51621 (100)	51622 (100)		ββhairpin 2.3.2	d.144.1.1	56111 (75)	56112 (75)	56113 (75)
							ββhairpin 3.4.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)
	βα5.4.1	d.144.1.1	56111 (100)	56112 (100)	56113 (100)		ββhairpin 4.1.1	d.144.1.1	56111 (66)	56112 (66)	56113 (52)
								d.144.1.2	56111 (66)	56112 (66)	56150 (14)
	βα6.1.1	c.37.1.1	52539 (100)	52540 (100)	52541 (92)		ββhairpin 4.5.1	d.128.1.2	55930 (100)	55931 (100)	55935 (100)
		c.37.1.2	52539 (100)	52540 (100)	52566 (4)		ββlink11.1.1	b.58.1.1	50799 (100)	50800 (100)	50801 (100)
		c.37.1.7	52539 (100)	52540 (100)	52589 (4)		ββlink12.1.1	d.128.1.2	55930 (100)	55931 (100)	55935 (100)

Second and third column identify the sub-class and the major SCOP domain code (more than 50% of the loops of the sub-class). In the last three columns the SCOP code of fold, super-family and family are indicated, and within parenthesis it is indicated the percentage of loops that belong to a particular SCOP code (cf, sf and fa) within the sub-class.

69

Besides, 6 sub-classes clustered functional loops from different folds (sub-classes $\alpha\beta1.1.1$, $\alpha\beta2.1.1$, $\beta\alpha$ 0.1.3, $\beta\alpha0.1.6$, $\beta\beta_{hairpin}$ 2.3.2 and $\beta\alpha$ 6.2.1 see Table 3.3). No specific conservation of functional residues was observed for sub-classes $\alpha\beta1.1.1$, $\alpha\beta2.1.1$, $\beta\alpha$ 0.1.3, $\beta\alpha0.1.6$ and $\beta\beta_{hairpin}$ 2.3.2 where also function was ambiguous. On the other hand sub class $\beta\alpha6.2.1$ (formed by P-loops) presents all functional residues conserved and a well-defined function. Within this sub-class the P-loop from "P-loop containing nucleotide triphosphate hydrolases" fold is clustered together with the P-loop from "PEP carboxykinases-like" fold. This result agrees with the recent work of Russell et al. where it is shown the conservation of the $\alpha\beta$ segment that contains the P-loop suggesting a common evolutionary ancestor for all known P-loop containing proteins(Russell et al. 2002).

Table 3.3. Sub-classes with functional residues involved in specific functions and shared by different folds are indicated.

Sub-class	Functional residues	Function described	SCOP domain code	% cf	% sf	% fa
$\alpha\beta1.1.1$	R (-A15)	Catalitic	c.86.1.1	48	48	48
	D (B5)	Ion binding (Mg)	c.72.1.1	22	22	14
	A (-A7) L (-A4) M (B3)	Substrate binding (Thiazola)	c.72.1.2	22	22	5
	E (-A8)	Substrate binding (NMP)	c.37.1.1	24	24	19
$\alpha\beta2.1.1$	H (B2) S (B4) E(B5)	Ion binding (Mg)	c.37.1.1	69	69	61
	T (-A11) D (B4)	Ion binding (Mg)	c.37.1.2	69	69	4
	A (-A7) L (-A3) M (B4)	Substrate binding (Thiazola)	c.72.1.2	4	4	4
$\beta\alpha0.1.3$	T (A1) D (A3) R (A6)	Substrate binding (NMP)	c.37.1.1	33	33	33
	K (L1)	Catalytic	d.143.1.2	33	33	33
$\beta\alpha0.1.6$	K (-B1)	Catalytic	c.143.1.2	33	33	33
	D (-B4)	Substrate binding (Adenosine)	d.72.1.1	33	33	33
$\beta\alpha6.2.1$	G (L1) G (L6) K (A1)	ATP binding (p-loop)	c.37.1.1	80	80	40
	G (L1) G (L6) K (A1)	ATP binding (p-loop)	c.37.1.2	80	80	10
	G (L1) G (L6) K (A1)	ATP binding (p-loop)	c.37.1.4	80	80	10
	G (L1) G (L6) K (A1)	ATP binding (p-loop)	c.37.1.6	80	80	10
	G (L1) G (L6) K (A1)	ATP binding (p-loop)	c.37.1.17	80	80	10
	G (L1) G (L6) K (A1)	ATP binding (p-loop)	c.91.1.2	20	20	20
$\beta\beta_{hairpin}2.3.2$	D (B7)	ATP binding	d.144.1.1	75	75	75
	L (B4) C (-B5)	Substrate binding	c.72.1.1	25	25	25

In the second column the functional residues and its location is the motif is described. The function is described in the third column and the SCOP domain code in the fourth. The last three columns indicate the percentage of loops of a particular SCOP code (cf, sf ,fa) within the subclass.

3.3.2   Data Collection of protein kinase folds

We have analyzed the loops clustered among the sub-classes of the classification for the different folds of kinases. We have focused on loops with functional annotation and in functional sub-classes, identifying those residues involved on function, binding and specificity. The fold for which we have found loops on the classification are:

3.3.2.1   Protein kinase like

On the protein kinase like fold there are six families of kinases (SCOP v.1.61 codes of family have been indicated within parenthesis): Ser-Thr kinases (56113), Tyr kinases (56150), Actin-fragmin kinase catalytic domain (56168), Miosyn Heavy Chain Kinase 7 Elongation Factor 2 (MHCK7-EF2) kinases (64408), Type IIIa 3'-5'-aminoglycoside phosphotransferase (64411) and Phosphoinositide 3-kinase catalytic domain (56171). They have a central conserved core of about 200-250 residues. The overall kinase domain folds into a two-lobed structure: the Nt lobe that anchors and orients the ATP, and the Ct lobe responsible for substrate binding and transference of the phosphate group. The active site is placed in a deep cleft between the two lobes(Schenk and Snaar-Jagalska 1999).

The family of Ser-Thr kinases is composed by kinases that transfer a phosphate group to a Ser or Thr residue of a target protein. These kinases are mainly involved on signal transduction, regulation of cell growth, cell division, cell mobility, metabolism, membrane transport, gene expression, and learning and memory(Johnson et al. 1998). Among the Ser-Thr family we found: cyclin dependent kinases (CDK) involved on cycle cell control and regulated by cycline proteins; Cyclic adenosine monophosphate

dependent kinases (cAPK) involved on signal transduction and activated by cyclic adenosine monophosphate (cAMP); Calcium-Calmodulin dependent kinases (CaMK) involved on signal transduction; Titin kinases, responsible of sarcomer-assemble control on muscle cells; Twitchin kinases, involved on the movement of the cytoskeleton; Mitogen Activated Protein Kinases (MAPK) involved on cycle cell and activated by mitogenic factors or cellular stress factors; Casein kinases, that phosphorylate a wide repertory of proteins including transcriptional factors like CREM (cAMP-responsive element modulator) or p53; the cytoplasmatic domain of the transforming grown factor beta (TGFß) receptor and others. Several sequence motifs have been described for Ser-Thr kinases: Gly-rich-loop (GXGXXGV) acting as a flexible clamp that anchors the non-transferable phosphates of ATP (found in sub-classes $\beta\beta_{hairpin}2.2.1$, $\beta\beta_{hairpin}2.2.2$ and $\beta\beta_{hairpin}3.4.1$, see Figure 3.5 for sub-class $\beta\beta_{hairpin}2.2.2$); AxK motif responsible for anchoring and orienting ATP (found in sub-class $\beta\alpha4.2.1$) and the catalytic loop HRDLKXXN (in sub-classes $\alpha\beta11.1.1$ and $\beta\beta_{links}8.1.1$). The activation region comprises the region between the DFG triplet, a highly conserved triplet involved in the binding of ATP (found in sub-class $\beta\beta_{links}5.1.1$), and a second conserved triplet of APE(Schenk and Snaar-Jagalska 1999) that was not found among the clusters because of its high flexibility. This activation region undergoes a large conformational change when switching between inactive and active states(Johnson et al. 1996).

The family of Tyr kinases is divided in two groups: Receptor Tyr Kinases (RTKs) and Non Receptor Tyr Kinases (NRTKs). The fold shares similar features with Ser-Thr kinases: the characteristic catalytic loop (found in sub-classes $\alpha\beta11.1.1$ and $\beta\beta_{links}8.1.1$), and a DFG triplet at the beginning of an activation loop, found in sub-class $\beta\beta_{links}5.1.1$ (see Figure 3.6 for sub-class $\beta\beta_{links}8.1.1$ and $\beta\beta_{links}5.1.1$), homologue to the activation loop of Ser-Thr kinases(Hubbard et al. 1994; Hubbard 1999).

The family of the catalytic domain of Actin-fragmin kinase contains kinases with F-actin capping activity involved on temporal and spatial regulation of the cytoskeleton. These kinases have a Gly-rich-loop shorter than in Ser-Thr and Tyr kinases(Steinbacher et al. 1999), that was not clustered in our database.

The MHCK7-EF2 kinase family contains the catalytic domain of Transient Receptor Potential (TRP) channel proteins (proteins involved on intracellular calcium level modulation). The catalytic domain is divided in two lobes: The Nt lobe shares the key features of the active site, with a phosphate binding loop and a catalytic loop, while the Ct lobe resembles ATP-grasp proteins (see further) and it contains a GXA(G)XXG motif located within the region that corresponds to the activation loop of the protein kinase like fold(Yamaguchi et al. 2001). Unfortunately, not enough motifs were found to produce a cluster of this motif, although the data can be retrieved from the loop database as a single loop.

Finally, the IIIa 3'-5''-aminoglycoside phosphotransferases form the family APH(3')-IIIa.This enzyme catalyzes the phosphorylation of 4,5-disubstituted antibiotics at the 5'' positions(Burk et al. 2001). The enzyme has a catalytic loop (found in sub-class $\beta\beta_{links}8.1.1$) and a Gly-rich-loop motif of protein kinases similar to motif found in cAMP-Dependent Protein Kinase (cAPK)(Burk et al. 2001).

3.3.3.2   P-loop containing nucleotide triphosphate hydrolases

Seven families in one super-family form the fold. The common feature is a highly conserved loop named P-loop or Walker-A-motif (GXXGXGK). Additionally, these kinases may have a second motif named Walker-B-motif (ZZZD being Z a hydrophobic residue). Both motifs are involved on the interaction with ATP(Saraste et al. 1990). The nucleotide and nucleoside protein kinase family (NMPKs)(52541) is the most populated family on this fold. These enzymes play an important role in the synthesis of nucleotides(Lavie et al. 1998; Van Rompay et al. 2000)). They are composed by three parts: (i) the CORE, a highly conserved central domain formed by a P-loop or Walker-A-motif (found in sub-classes

$\beta\alpha$6.1.1 and $\beta\alpha$6.2.1, see Figure 3.3) and occasionally a Walker-B-motif(Krell et al. 1998) (in sub-classes $\alpha\beta$2.1.1, $\alpha\beta$3.1.1 , $\beta\alpha$4.1.1, $\beta\alpha$5.1.1 and $\beta\alpha$7.1.1); (ii) the NMP-binding domain, a domain containing a nucleoside monophosphate binding site (with loop conformations found in sub-classes $\alpha\alpha$3.1.1, $\alpha\alpha$ 4.1.1, $\alpha\alpha$ 5.1.1, $\alpha\alpha$ 5.3.1, $\alpha\beta$1.2.2, $\alpha\beta$ 5.1.1, $\beta\alpha$0.1.1, $\beta\alpha$0.1.3, $\beta\alpha$2.3.1, $\beta\alpha$2.4.1 and $\beta\alpha$5.1.1; see Figure 3.1); (iii) the LID domain, a domain of variable length (between 11 and 64 residues) with a mobile loop(Briozzo et al. 1998) (found in sub-classes $\alpha\alpha$9.4.1, $\alpha\beta$0.1.1, $\beta\alpha$1.1.1, $\beta\beta_{hairpin}$4.1.1 and $\beta\beta_{hairpin}$4.2.1). The enzyme is inactive in the open conformation of the LID domain and the binding of substrates (ATP and NMP) causes the approach between CORE and NMP-binding domains. Finally, the LID domain undergoes a conformational change and it remains over the active site seized by positively charged residues, allowing for the transference of a phosphate group(Scheffzek et al. 1996; Sinev et al. 1996).

Cloramphenicol phosphotransferase (52569), Adenosine 5'-phopshosulfate (52572) , Gluconate kinase (75195), phoshoribulokinase/pantothenate kinase (52584), Shikimate kinases (52566) and 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (52589) share a common mononucleotide-binding fold as well as the Walker-A-motif(Izard and Ellis 2000; MacRae et al. 2001; Kraft et al. 2002) (found in sub-classes $\beta\alpha$6.1.1 and $\beta\alpha$6.2.1) and occasionally the Walker-B-motif. This motif is also found for Shikimate kinases and in 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (in sub-class $\alpha\beta$3.1.1)(Krell et al. 1998; Yuen et al. 1999).

3.3.3.3   Ribokinase like

Included in the Ribokinase like SCOP fold there are three families: (i) Ribokinase like (53614); (ii) Thiamin biosynthesis kinases (53620); and (iii) ADP-dependent glucokinase (64147). All families include

enzymes that catalyze the transference of a phosphate group from ATP to a cyclic carbohydrate molecule(Cheng et al. 2002).

The Ribokinase like family includes two types: i) adenosine kinase , and ii) ribosin kinase. The enzyme is monomeric and its active site is formed by two binding sub-sites for adenosine and Mg^{2+}(Mathews et al. 1998). The active form of Ribosin kinase is dimeric and it catalyses the transference of a phosphate group from ATP to a molecule of D-ribose. On the structure of the enzyme it is found a particular arrangement of β-strands, named β-clasp-lid, clustered in sub-class $\beta\beta_{hairpin}$4.1.1, which has been described as a functional motif. In its inactive form, the ß-clasp-lid is situated over the active site. This β-strand may participate on the enzyme dimerization(Sigrell et al. 1998). Functional residues related with ATP binding and substrate binding have been described in the literature and were mapped in sub-classes $\alpha\alpha$4.5.1, $\alpha\beta$1.1.1 and $\beta\beta_{hairpin}$2.3.2.

The Thiamin biosynthesis kinase family is formed by thiazole kinases (THZK) and 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinases (HMPPK). The active form of THZK is a homo-trimer(Campobasso et al. 2000) while the active form of HMPPK is a dimer(Cheng et al. 2002). Some functional residues were mapped in sub-classes $\alpha\beta$1.1.1, $\alpha\beta$2.1.1 and $\alpha\beta$3.3.2.

3.3.3.4   Phosphoglycerate kinase

Kinases belonging to this fold and family catalyze the transference of a phosphate group from 1,3-bi-phosphateglycerate to ADP. Their protein structures and sequences were highly conserved in prokaryotes and eukaryotes(Davies et al. 1993; Bernstein et al. 1998). The structure of the active enzyme is monomeric and it is formed by two domains linked by a pivoting α-helix, which articulates the movement between open and close conformations (found in sub-class $\alpha\beta$7.4.1)(Davies et al. 1993; Auerbach et al. 1997). The active site is located between both domains and it is formed by four binding sites: 3-phosphoglycerate binding site (clustered in sub-class $\beta\alpha$12.1.1), ATP binding site (loops for

this binding site were splitted between sub-classes $\alpha\alpha$10.2.1, $\alpha\beta$7.4.1 and $\beta\alpha$1.2.1), and two binding

sites for anionic ligands (usually sulphate(Sherman et al. 1990)). On the catalytic mechanism Arg (

found in sub class $\beta\alpha$12.1.1) and Asp residues have been described to be involved by means of a salt

bridge on the approaching movement between both domains(Auerbach et al. 1997).

3.3.3.5   Ribonuclease-H-motif

All kinases belonging to this fold: glycerol kinases (53089), acetate kinases (53080) and hexokinases

(53083), share a conserved structural core that consist of a duplicated βββαβαβα secondary structure

with insertions of sub-domains between particular elements of the β-sheet(Buss et al. 2001). Moreover,

a set of motifs are conserved in these kinases: the "adenosine" motif that interacts with ribosyl and α-

phophoryl group of ATP, the "Phosphate 1" motif that interacts with Mg^{2+} and the "Phosphate 2" motif

that interacts with β and γ-phosphoryl group of ATP. These motifs were not clustered into sub classes

because there were not enough loops with similar conformation (due to a large flexibility) to produce a

cluster. On the other hand, we found loops of hexokinases family in sub-classes $\beta\alpha$16.1.1 and

$\beta\alpha$17.1.1, involved in the substrate interaction (see Figure 3.2).

3.3.3.6   ß/α Tim barrel SCOP

The pyruvate kinase (51622) and Pyruvate phosphate dikinase Ct-domain (51629) are included inside

the ß/α Tim barrel fold. Pyruvate kinases catalyze the last step in glycolysis, yielding ATP and pyruvate

from ADP and P-enolpyruvate(Larsen et al. 1994), and on the decarboxylation of oxalacetate and the

enolization of pyruvate(Larsen et al. 1997). The active enzyme is an oligomer constituted by four

subunits(Larsen et al. 1994). The enzyme shows two conformations depending on the catalytic state,

open (inactive) and closed (active)(Larsen et al. 1998). Four functional regions have been described : (i)

a divalent cation binding site (found in sub-classes $\alpha\alpha$2.1.1, $\beta\alpha$0.1.2) , which is needed for the

coordination of ADP; (ii) a substrate binding site for pyruvate and ADP (found in sub-classes $\alpha\beta$7.2.1, $\beta\alpha$14.1.1,); (iii) a monovalent cation binding site, typically K+ or Na+ (in sub-class $\beta\alpha$5.2.1), also needed on its enzymatic activity(Larsen et al. 1998); and (iv) the allosteric site, formed by two loops of variable length where fructose-1,6-bisphosphate is accommodated (clustered in sub-classes $\alpha\beta$5.3.1 and $\beta\alpha$3.1.1)(Jurica et al. 1998). The comparison between pyruvate/phosphoenolpyruvate binding site of pyruvate phosphate dikinase kinases and the active site of pyruvate kinase reveals the conservation of several key residues involved in substrate binding(Herzberg et al. 2002).

3.3.3.7 Glutamine synthase/guanido kinase catalytic domain

The guanido kinases (55935) are involved on the reversible transfer of a phosphoryl group from phosphocreatine (creatine kinase) or phosphoarginine (arginine kinase) to ADP. The fold is formed by two domains: 1) a small α-helical N-terminal domain (fold 48033 without kinase activity); and 2) a larger C-terminal α/ß domain that contains the active site(Fritz-Wolf et al. 1996; Rao et al. 1998; Zhou et al. 1998). Several loops involved in ATP binding of these kinases were clustered in sub-classes $\beta\alpha$3.2.1, $\beta\alpha$7.1.1, $\beta\alpha$11.2.2, $\beta\beta_{hairpin}$4.5.1, $\beta\beta_{hairpin}$5.3.1 and $\beta\beta_{link}$12.1.1 being involved in ATP binding (see Figure 3.4 for sub-class $\beta\alpha$11.2.2). Loops involved in monomer-monomer interaction were found in sub-classes $\alpha\alpha$8.3.1 and $\alpha\beta$10.1.1 (see Figure 3.7 for sub-class $\alpha\beta$10.1.1).

3.3.3.8 PRTase-like

The enzymes of this family are phosphoribosylpyrophosphate (PRPP) synthetases (53296). The active site of these enzyme is located between two domains(Eriksen et al. 2000). The PRPP binding site is formed by three loops of conserved sequence(Sinha and Smith 2001): the "PPi loop", the "PRPP loop" and a third flexible loop. There is only one protein in the classification but some of its loops were

clustered in sub-class $\beta\beta_{hairpin}$4.2.1 (motif involved in ATP binding), and in sub-class $\alpha\beta$3.1.2 (motif involved in the binding of the regulatory ADP).

3.3.3.9   Ferrodoxin-like

Three super-families form the Ferrodoxin-like fold: Nucleoside diphosphate kinase , 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) and Galactose Homoserine Mevalonate Phosphomevalonate (GMPH) kinases. The nucleotide diphosphate kinases (54920) are kinases involved in the exchange of a γ-phosphate between nucleoside di- and triphosphate(Cherfils et al. 1994). We found three out of four conserved motifs in these kinases: (i) Kpn loop, a mobile loop involved in the interaction with the nucleotide(Williams et al. 1993) (found in sub-class $\alpha\alpha$12.1.1); (ii) GXXGK motif analogous to the p-loop but with different function(Williams et al. 1993) (in sub-class $\alpha\alpha$1.1.1); (iii) DXXG motif involved in nucleotide binding(Williams et al. 1993) (found in sub-class $\alpha\alpha$12.1.1); and (iv) HSGD motif that contains the catalytic His involved in the transfer of the phosphate group to the nucleoside diphosphate substrate(Cheek et al. 2002) .

The active site of the HPPK family (55084) includes two Mg^{2+} binding sites coordinated by two Asp. The comparison between apo and holo forms reveals a remarkable conformational change of three flexible loops(Blaszczyk et al. 2000).

The GMPH kinase super-family is formed by three families: (i) the Homoserine kinase Ct domain family (55061)that catalyzes the phosphorilation of L-homoserine from ATP(Krishna et al. 2001); (ii) the mevalonate kinase family (75540) that catalyzes the transference of the γ-phosphate group from ATP to mevalonate(Yang et al. 2002); and (iii) the phosphomevalonate kinase family (75454) for the transference of the γ-phosphoryl group of ATP to the phosphate oxygen of (R)-5-phosphomevalonate(Romanowski et al. 2002). The active site of GMPH kinases lies along the cleft

between Nt-domain and Ct-domain. A Gly-rich-loop in close contact with the phosphate-binding loop (located at the Nt-domain) is found in the Ct domain of the GMPH kinases(Yang et al. 2002).

3.3.3.10   Ribosomal protein S5 domain 2-like

The fold is formed by the Ribosomal protein S5 domain 2-like super-family and the GMPH kinase Nt-domain (54232) family that includes the Nt-domain of the above-mentioned GMPH kinases. The Nt-domain of these enzymes contains two motifs identified in all GMPH kinases: Motif I, at the beginning of the Nt-domain; and Motif II, PXXXGLGSSAA that forms a novel phosphate-binding loop. The orientation of the ATP is different in the Motif II than in the Walker-A-motif (P-loop), where a conserved glutamate coordinates a magnesium cation involved in the coordination the γ-phosphate. On the other hand, motif II interacts extensively with the α- and β-phosphates of the nucleotide(Krishna et al. 2001). We found a functional loop clustered in sub-class $\alpha\beta 3.1.6$ preceding motif II. Nevertheless, both characteristic motifs, Motif I and Motif II, were not clustered in the database and appeared as single loops.

3.3.3.11   Carbamate kinase like

Carbamate kinases (53634) produce Carbamoyl phosphate (CP) using ATP, bicarbonate and ammonia as precursors. The active form is a dimer and each monomer is split by a large cleft between Nt- and Ct-domains. The Nt domain binds the carbomoyl moiety of CP while the Ct domain is involved on ADP binding(Ramon-Maiques et al. 2000). A catalytic loop was clustered in sub-class $\alpha\beta 3.1.5$ and a loop for substrate binding in sub-class $\alpha\alpha 4.1.2$.

3.3.3.12   PEP carboxykinase-like (PCK)

The Ct domain of Histidine phosphocarrier protein (HPr) kinase/phosphorilase belongs to this fold. This enzyme catalyzes   the ATP-dependent phosphorylation/dephosphorylation in serine of HPr. These kinases contain the characteristic Walker A (P-loop) (clustered in sub-class $\beta\alpha6.2.1$)(Fieulaine et al. 2001). This loop is found included in the common structural core of Hpr, all PCK kinases and P-loop containing proteins, as much of the active site, with a number of identities associated to a P-value of 10^{-10}. This has suggested a common ancestor relationship(Russell et al. 2002) (see above).

3.3.3.13   SAICAR synthase-like

The Phosphoinositide kinases (PIP3-kinase-II-ß) (56108) play an important role in signal transduction by phosphorylating either phosphoinositol 3-phosphaste (PI3P) or phosphoinositol 5-phosphate (PI5P) on the 4-hydroxyl. PIPK-II-ß consists of two α+ß domains(Rao et al. 1998) and has some motifs shared with protein kinase like kinases: a Gly-rich-loop and a hypothetical activation loop that has been proposed as substrate specificity loop(Hurley et al. 2000) not clustered in our database. On the other hand, we have found two loops in sub-classes $\beta\alpha0.1.3$ and $\beta\alpha0.1.6$ with annotations in PDB and related with the active site.   In both cases, the loops have clustered with loops of other folds and involved in substrate binding.

3.3.3.14   Phosphofructokinase

The fold of phosphofructokinase (53785) is formed by two domains (Ct and Nt). The nucleotide-binding domain, is located in the Nt domain and the active site in the cleft between both domains(Shirakihara and Evans 1988). We have found loops of this family in sub classes $\alpha\beta1.1.2$, $\alpha\beta1.1.6$, $\alpha\beta3.1.2$ $\alpha\beta3.1.4$ and $\alpha\beta3.3.1$. These loops have residues in contact with ADP and Mg within a cut-off distance

of 6Å, suggesting its participation in binding and catalysis even though the sub-class was not considered functional (the function was not proved for more than 50% of the loops of the sub-class).

3.3.3.15   Thiamin pyrophosphokinase catalytic domain

The fold is formed by the Ct-domain of thiamin pyrophosphate (TPP) enzymes. The active form of the enzyme is a homo-dimer and each subunit has two domains, Nt and Ct, being the Ct domain the catalytic domain (64000) and the Nt domain the substrate binding domain (63863). The active site is located in a cleft at the interface between Nt domain of one monomer and the Ct domain of the other(Timm et al. 2001). We have found loops of this family in sub-classes $\alpha\beta1.1.2$ and $\beta\alpha1.1.2$. In both sub-classes we have found loops with annotated function in substrate, ADP and Mg^{2+} binding. However, the annotation was obtained for different folds and the sub-classes contained less than 50% of loops with function. Consequently, a similar function could be suggested but not inferred.

3.3.4   Non-clustered functional motifs

Not all the analyzed kinases are equally represented in our database, for some kinases there is a large variety of homologs while for others there is only one representative in the PDB. One cluster is formed by at least two loop conformations. Consequently, we may have functional annotation for residues of loops that could not be clustered because of the lack of enough representative members of the family. Some loops belong to flexible regions; or are found complexed with a substrate (or other molecules), producing changes of the loop conformation; or are poorly diffracted and the chain breaks. Therefore, these loops were not classified either.

The information for these loops is added on the loop database as single non-clustered loops. Still, they can be retrieved by querying the server by the protein name or searching for functional loops. We found

in this situation several folds, for which the functional information was not mapped on the sub-class but on a single loop conformation. A few of these cases were previously mentioned. In these cases some loops were classified because of being common to other folds, while others loops were specific of the family and were not clustered because the lack of members of the family. Also, avoiding redundancies implied the removal of identical loops yielding the banishment of those cluster with insufficient number of non-identical loops.

Here, we have included those folds for which no common loops involved in function could be found to yield a cluster. This is the case of the folds: ATP grasp, ATPase domain of HSP90 chaperone/DNA topoisomerase II/ Histidine kinase.

3.3.4.1 ATP grasp

The ATP grasp fold is formed by the super-family Glutathione synthease ATP-binding domain-like and the family Pyruvate phosphate dikinase (PPDK) Nt-domain (56085). The Nt domain of PPDK contains the ATP binding site(Herzberg et al. 2002).

3.3.4.2 ATPase domain of HSP90 chaperone/DNA topoisomerase II/ Histidine kinase

The protein histidine kinases (PHKs) (55884) are key components of many signaling pathways. There are two classes of PHKs according to their domain organization. In class I the domain that contains the histidine phosphorylation site (domain H) is followed by the ATP-domain, whereas in class II one or more domains separate the H-domain from the ATP-domain(Bilwes et al. 2001). There are four conserved regions in the ATP binding pocket that typify the histidine kinases: (i) the N box. (ii) the G box. (iii) the F box and (iv) the G2 box and a mobile loop, called ATP-lid, that couples the ATP-binding

and the motion between domains in the catalysis(Bilwes et al. 2001). These conserved regions and the ATP-lip loop also appears at the Ct domain of the α-ketoacid dehydrogenase (69804) (BCK) (also called K domain) and in the Ct-domain pyruvate dehydrogenase kinase(Machius et al. 2001; Steussy et al. 2001).

3.3.5   Selected functional sub-classes.

We have selected a few examples of sub classes of loops where it can be shown the relation between sequence and structure. These examples are also used as a proof that the relation between one special motif and its function is maintained for different folds and families. Residue positions within clusters are denoted by the secondary structure (B for β-sheet, A for $\alpha-$helix, and L for coil) and counting backwards from the C-terminal residue of the first secondary structure and forwards from the N-terminal residue of the loop.

3.3.5.1   Substrate binding loops

Sub-classes $\alpha\alpha$5.1.1 and $\beta\alpha$5.1.1 (see Figure 3.1) contain loops involved in substrate binding from the nucleotide and nucleoside protein kinase family. Gly in position L1 and Val in position L4 of sub-class $\alpha\alpha$5.1.1 are the residues necessary for the interaction with the nucleoside monophosphate(Berry et al. 1994; Muller-Dieckmann and Schulz 1994). Residues located in A4 are involved in substrate recognition, where a Val is found for the adenylate kinase and a Thr for uridilate kinase(Scheffzek et al. 1996). On the other hand, sub-class $\beta\alpha$5.1.1 presents a Gly in position L1 and Arg in position L4 implicated in substrate binding. The specificity of this sub-class is given by a Gln residue in position A3 for adenilate kinase and Asn for uridilate kinase (see Figure 3.1 for details)(Scheffzek et al. 1996).
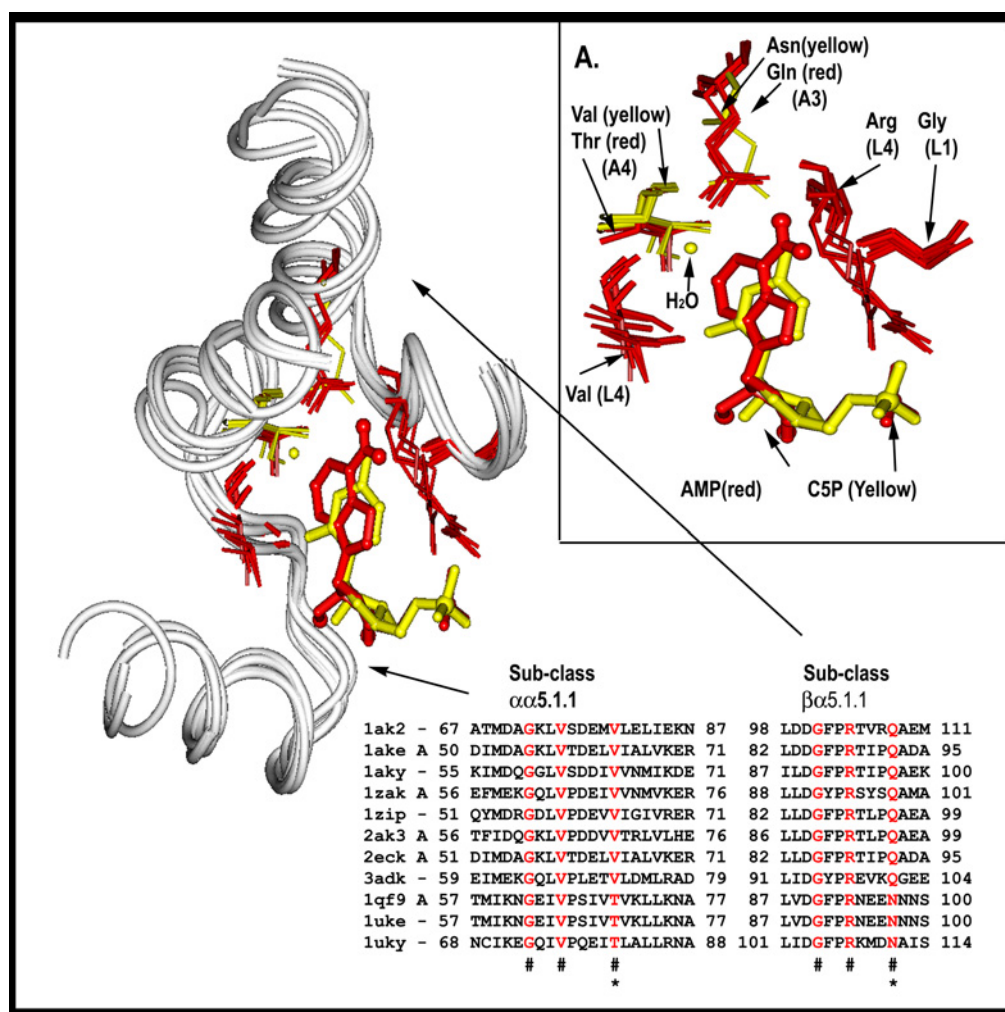
Figure 3.1. Composition to show a general overview of sub-class $\alpha\alpha5.1.1$ and sub-class $\beta\alpha5.1.1$. Side-chain residues involved on substrate specificity are indicated in red (for adenine kinases) and in yellow (for uridilate kinase). Residues specific for substrate binding are highlighted in red (plus #) in the multiple sequence alignment, while positions implicated on specificity are marked with an asterisk. Substrate molecules of AMP (extracted from the structure with PDB code 2ak3) and C5P (taken from the structure with PDB code 1qf9) were included on the superimposition of loops together with the side-chains of the loops involved in substrate binding. A detailed view is presented in the right top corner (A). All the Figures were generated with PREPI v.0.9 S.A. Islam & M.J.E. Sternberg (available upon request on: http://www.sbg.bio.ic.ac.uk/prepi/index.html).

Sub-classes $\beta\alpha16.1.1$ and $\beta\alpha17.1.1$ are specific for the hexokinase family. We found specific contacts of Thr (in position L4) and Lys (in position L5) of sub-class $\beta\alpha16.1.1$ and Asn (in position –B1) and Glu (in position L3) of sub-class $\beta\alpha17.1.1$ with a glucose molecule (see Figure 3.2 for details). Besides, a non–crystallographic dimer of wild-type human hexokinase complexed with glucose and glucose-6-phosphate (PDB code 1hkb) has the corresponding annotation for the region covered by both sub-classes.
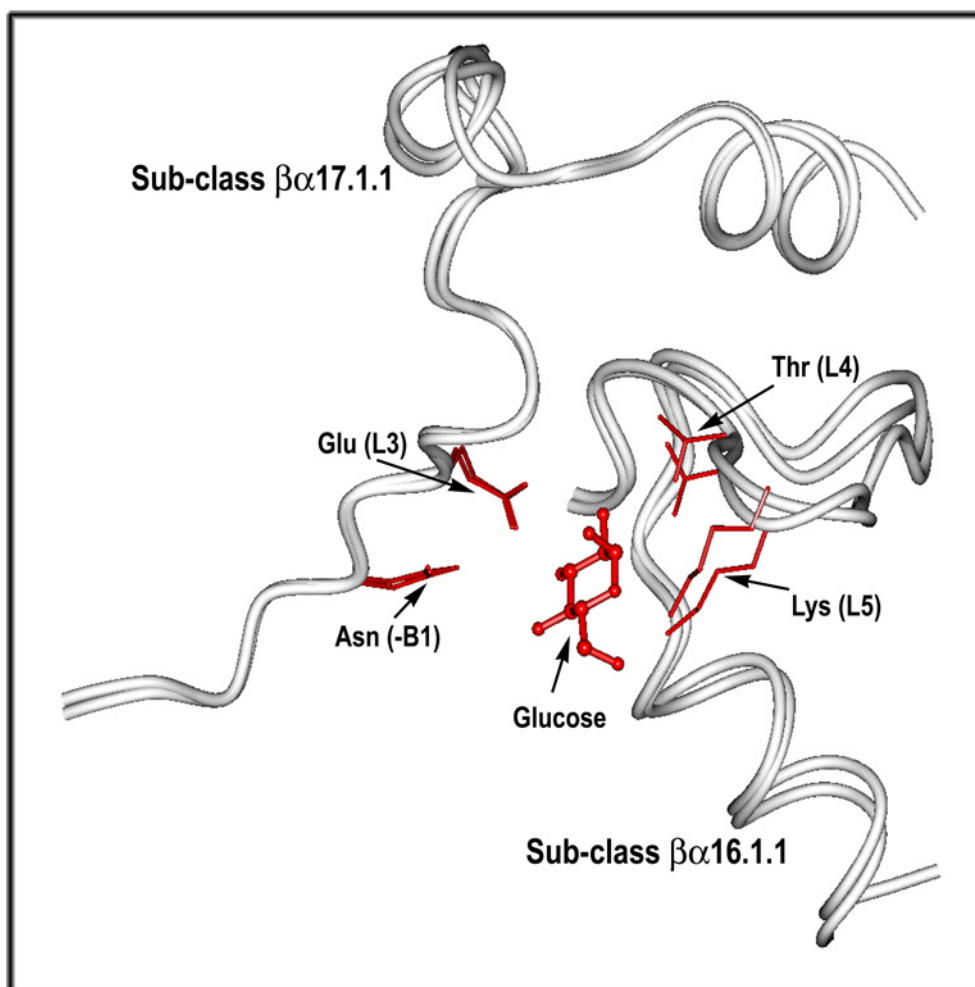
Figure 3.2. Coiled trace of the loops of sub-class βα16.1.1 and βα17.1.1. Side chains of conserved and functional residues are shown. The structure with PDB code 1bdg of a hexokinase in complex with glucose was used to superimpose the glucose molecule with the rest of loops from this sub class.

3.3.5.2   ATP binding

Sub-class βα6.1.1 (see Figure 3.3) and sub-class βα6.1.2 have clustered the P-loop motif (or Walker

A motif)(Saraste et al. 1990), which is conserved among all kinases of the "P-loop containing nucleotide

triphosphate hydrolases", "Nucleotide diphosphate kinases" and "Hpr kinases". A lysine residue is

conserved in position A1, near the γ-phosphate of ATP and involved on the transference of γ-phosphate

to the substrate(Matte et al. 1998). Conserved Gly residues playing a role on the binding of ATP are

also noteworthy. Additionally, in the C-terminal direction of the P-loop it is often found another motif,

clustered in sub-class βα0.1.1, which is involved in substrate binding. Sub-class βα6.1.1 mainly

contains loops of proteins that belong to the "Nucleoside and nucleotide kinases" family while $\beta\alpha6.2.1$

contains loops of the rest of families of the "P-loop containing nucleotide triphosphate hydrolases" fold

and the proteins of "Nucleotide diphosphate kinases" super-family and "Hpr kinases". The principal

difference between the proteins of both sub-classes the type of substrate that they catalyze: while the

family of "Nucleoside and nucleotide kinases" phosphorylate nucleotide/nucleoside substrates,

members of sub-class $\beta\alpha6.2.1$ phosphorylate non-nucleotide substrates, such as D-gluconate acid,

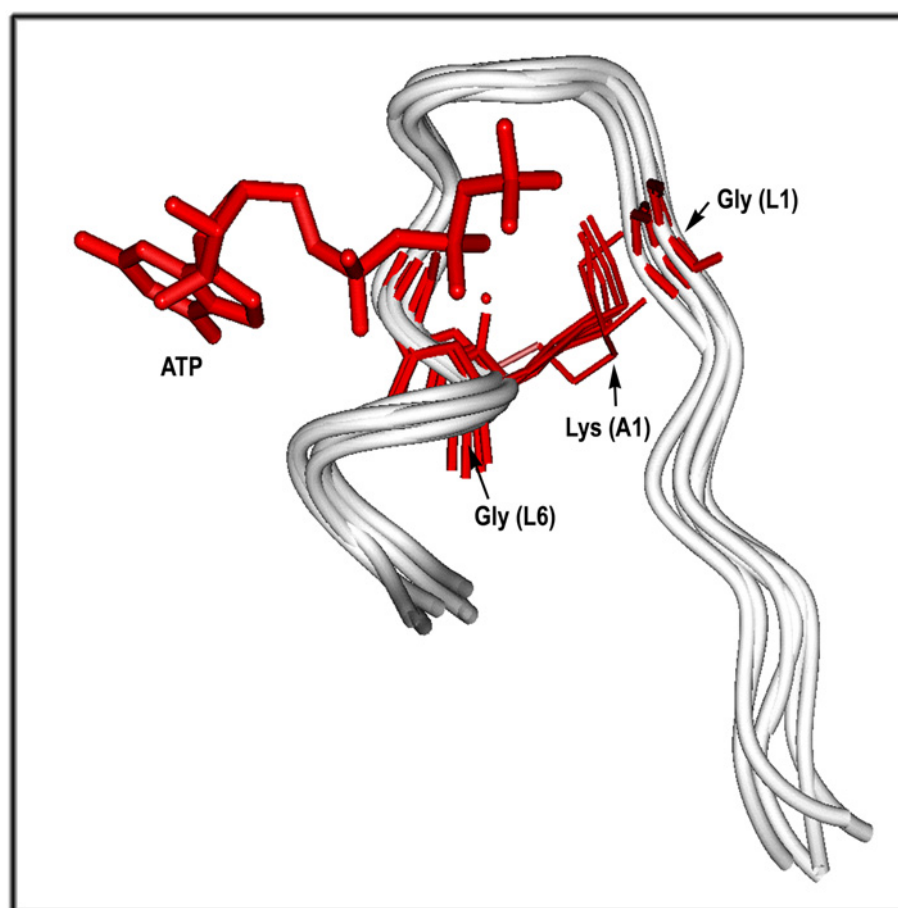Adenosine 5'-phosphosulfate, pantothenic acid or the HPr protein.

Figure 3.3. Coiled coil representation of the P-loop motifs of sub-class $\beta\alpha6.1.1$. The structure of a protein in complex with ATP (PDB code 1dvr) was used to superimpose the ATP molecule with the set of loops of this subclass. The side-chain of Lys in position A1 is included to show the binding of ATP.

Also, the work of Via et al.(Via et al. 2000) showed four surface regions associated with the P-loop

containing proteins. Some of the structures used in their analysis were present in our database. These

conserved regions are (according to the nomenclature used by the author): (i) the GK P-loop element (described above); (ii) the first negative charge; (iii) the conserved positive charge and (iv) the second negative charge. We found functional sub-classes containing the residues described in these regions. The GK P-loop element was found in sub-classes $\beta\alpha6.1.1$ and $\beta\alpha6.1.2$; the first negative charge in sub-classes $\alpha\beta5.1.1$, $\beta\alpha0.1.1$ and $\beta\alpha2.3.1$; the conserved positive charge in sub-classes $\alpha\alpha9.4.1$, $\alpha\beta0.1.1$, $\beta\alpha1.1.1$, $\beta\alpha2.1.1$ and $\beta\alpha4.1.1$; and the second negative charge in sub-class $\beta\alpha4.1.1$.

In a recent comparison of conserved motif sequences defined as: Walker A, strand 2, helix 2, Walker B, strand 4 and helix 4) and using structural information, Leipe et al.(Leipe et al. 2003) has proposed a classification for the kinases of the "*p-loop kinases class* in 8 groups" (named A to H) and approximately 40 families according to the sequence and structural conservation of these defined motifs. Furthermore, the authors has suggested the evolutionary process followed by theses kinases.

We have found some of these sequences motifs: Walker A motif in two sub-classes of the same class, sub-class $\beta\alpha6.1.1$ and $\beta\alpha6.2.1$. Strand 2 in groups A, C and D is found in sub-classes $\beta\alpha0.1.1$ and $\alpha\beta2.1.1$ while for group B this is in $\alpha\beta3.1.1$ and $\beta\alpha2.4.1$ sub-classes. Helix 2 of the groups A,C and D is found in sub-classes $\beta\alpha0.1.1$ and for group B in sub-class $\beta\alpha2.4.1$. The Walker B motif for group A is found in sub-classes $\beta\alpha2.1.1$, $\alpha\beta7.1.1$ and $\beta\alpha5.1.1$; for group B in sub-class $\beta\alpha2.1.1$ and for group D in subclass $\alpha\beta3.1.1$. Strand 4 motif is found in sub-class $\beta\alpha1.1.1$ for groups A, B, C and D; sub-class $\beta\alpha4.1.1$ for groups A and B; and sub-class $\beta\alpha1.1.2$ in groups B and C. Finally, the defined helix 4.1 motif is found in sub-class $\beta\alpha1.1.1$ for groups A, B and C; sub-class $\beta\alpha4.1.1$ for group A and B; sub-class $\beta\alpha1.1.2$ for groups C and D.

Figure 3.4 shows the sub-class $\beta\alpha$11.2.2 . This sub-class is specific for the "Guanido kinases" family. A Ser in position --B7 and Arg in position –B3 and –B4 are engaged in ATP binding(Fritz-Wolf et al. 1996; Rao et al. 1998; Zhou et al. 1998). Additionally, we have found a conserved Gly located in L5 and a Pro in L7 probably conserved for structural reasons.
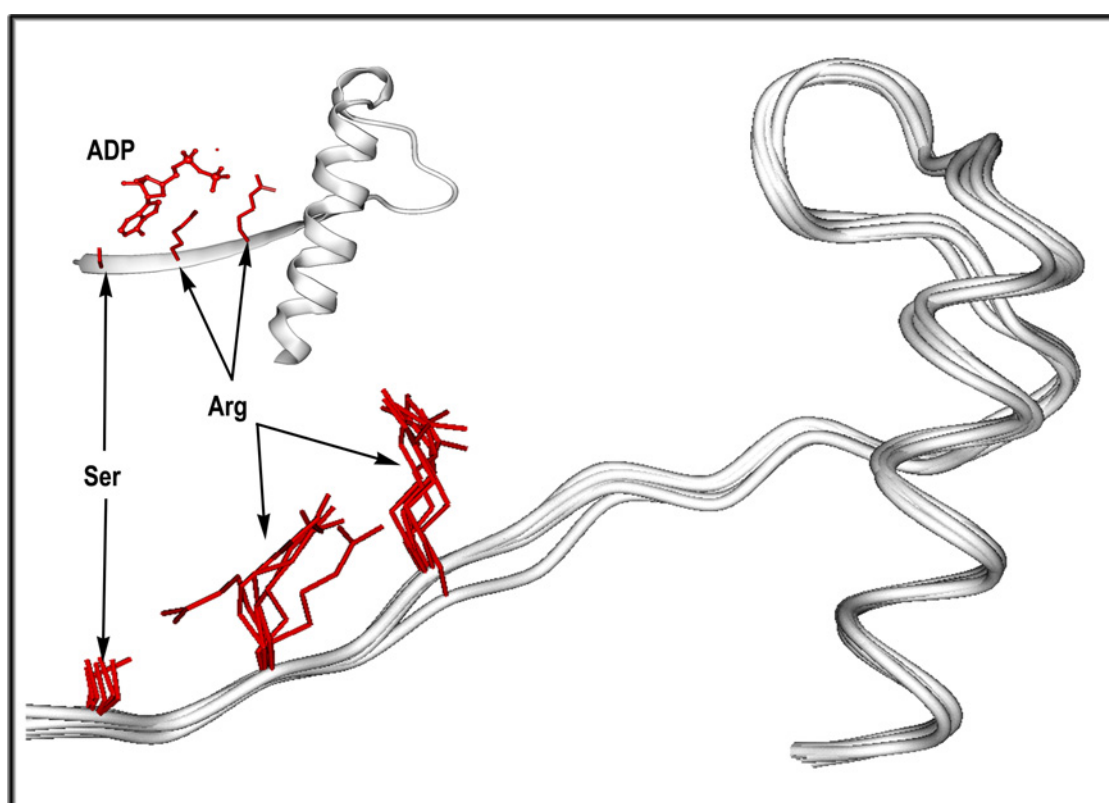

Figure 3.4. Coil representation of the main-chain of loops from sub-class $\beta\alpha$11.2.2. Only residues with functional annotations are shown (Ser position -B8 and Arg position -B4 and -B6). The structure of a protein in complex with ADP (PDB code 1bg0) was taken to illustrate the pattern of interaction between the conserved residues and the ADP molecule.

Three $\beta\beta_{hairpin}$ sub-classes, 2.2.1, 2.2.2 and 3.4.1 (see Figure 3.5) clustered the Gly-rich-loop from the "protein kinase like" fold. The Gly-rich-loop act as a flexible clamp that anchors ATP phosphates(Schenk and Snaar-Jagalska 1999). Consequently, due to a high flexibility these loops appeared in three different sub classes with different length (between two and four residues).
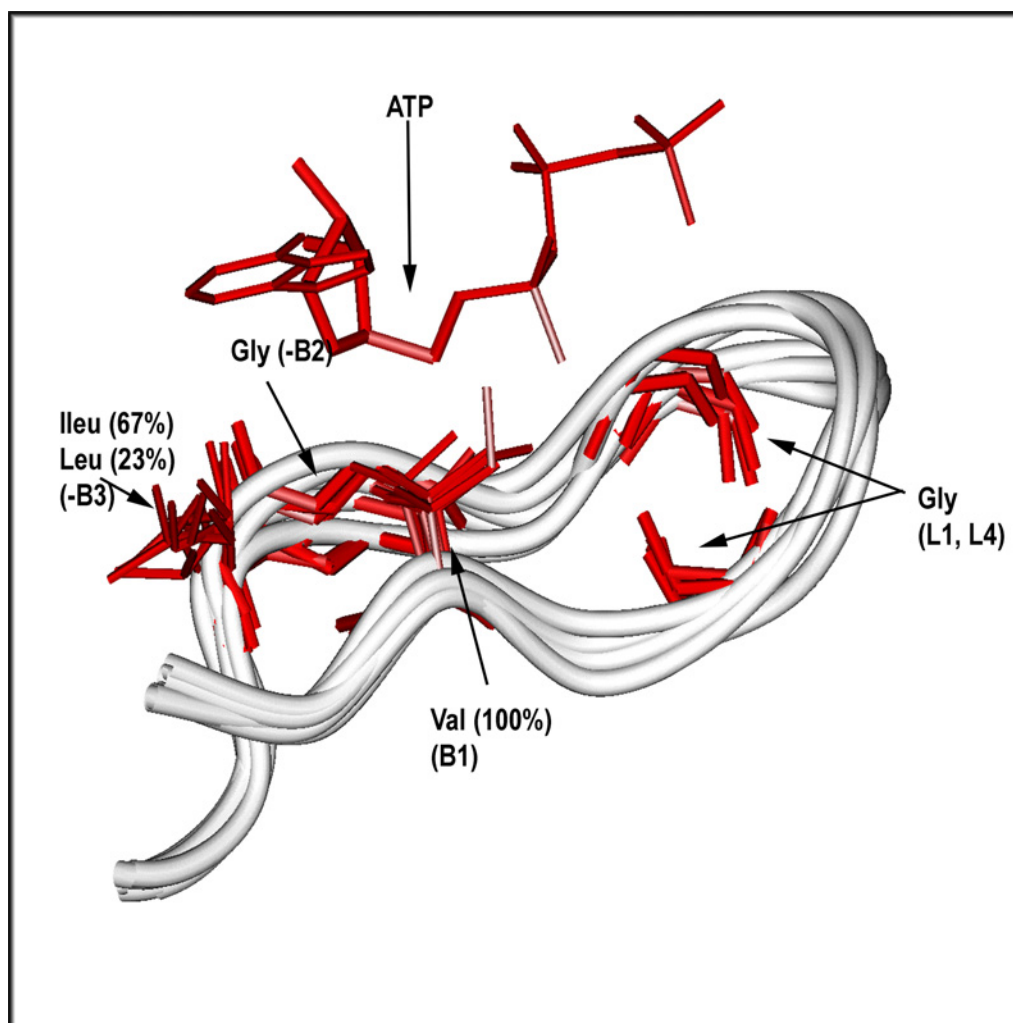
Figure 3.5. Trace plate of the sub-class ββhairpin2.2.2 showing the Gly-rich-loop motif found in kinases of protein kinases like fold. The molecule of ATP, taken from the structure with PDB code 1b38, was superimposed with the rest of loops of the sub -class to show the pattern of interactions between the adenosine ring and the hydrophobic residue (Ile or Leu) on the Nt β-strand and a Val in the Ct β-strand.

3.3.5.3   Catalytic loops

We found two sub-classes (Figure 3.6) involved in the binding of ATP and transference of phosphate:

Sub-class ββ$_{link}$8.1.1 (with catalytic residues responsible for the transference of a phosphate group) and

sub-class $\beta\beta_{links}$5.1.1 with the DFG motif of the activation loop. Both sub-classes contain loops from the same fold of kinases (protein kinase like) because they are functionally related on the mechanism of phosphorylation.

In sub-class $\beta\beta_{link}$8.1.1, Asp (in position L3) is the most likely candidate to be the catalytic base(Bossemeyer et al. 1993; Schenk and Snaar-Jagalska 1999), and Lys (in position L5) binds the ATP γ-phosphate. Lys (at L5) and Asn (at L8) are involved in the stabilization of the transition state(Bossemeyer et al. 1993; Goldberg et al. 1996), also Asn (in position L8) binds a Mg^{2+}(Xu et al. 1995; Xu et al. 1996; Xie et al. 1998). In position B1 of the $\beta-$strand we found a conserved hydrophobic residue that interacts with the adenine ring of ATP . No functional annotations were found on the literature for a conserved Pro in position L6, but its conservation can be explained by structural reasons producing a turn that places Asn (L8) with the correct orientation.

On the other hand, sub-class $\beta\beta$links5.1.1 contains the triplet of DFG, which is a fragment of the activation loop of the protein-kinase like fold. It is postulated that the conserved Asp (in position L2) is responsible for ATP binding and metal binding(Bossemeyer et al. 1993; Xu et al. 1995; Goldberg et al. 1996; Xu et al. 1996). Additionally, we have found a conserved hydrophobic residue in position –B1 that may interact with the adenine ring of ATP.

 It has to be noted that sub-class $\alpha\beta$11.1.1 is homologous to sub-class $\beta\beta$links8.1.1 with identical catalytic residues, but different bracing secondary structures. A detailed observation of both sub-classes proves that the loop structure is identical in both sub-classes, but the program DSSP(Kabsch and Sander 1983) failed to find the βstrand at the Ct site for sub-class $\beta\alpha$11.1.1 .
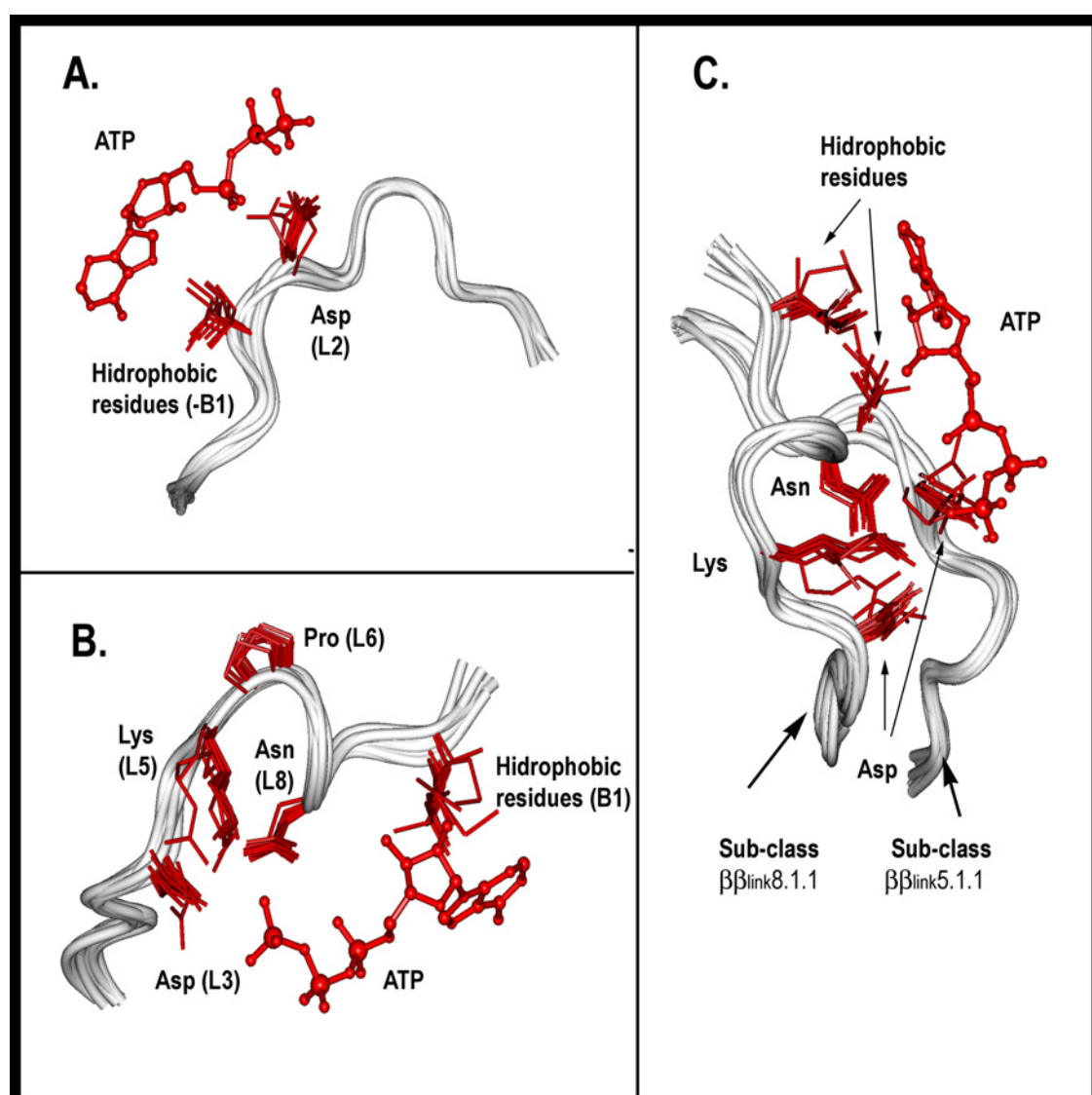
Figure 3.6. A) Coil trace of the sub-class ββlinks5.1.1 showing the DFG triplet motif. Side-chains of the Asp of the triplet and conserved residues located in –B1 are included in the picture. B) Coiled coil representation of the loops of sub-class ββlinks8.1.1 with the characteristic set of catalytic loops of the protein-kinase like fold. Side-chains for conserved residues are depicted (in red). The ATP of the structure with PDB code 1phk was superimposed with the rest of loops of the sub-class. C). Coiled plate with the combination of sub-classes ββlinks8.1.1 and ββlinks5.1.1 showing the main interactions with ATP.

3.3.5.4 Loops for dimerization/oligomerization

Sub-class $\alpha\beta$10.1.1 is specific for guanido kinases (see Figure 3.7). This sub-class is involved in the dimerization of guanido kinases, being the active form oligomeric. Residues in positions L5 and L7 are an invariant Arg and Trp respectively and both residues have been proposed to be involved in the oligomerization process of the enzyme(Fritz-Wolf et al. 1996).
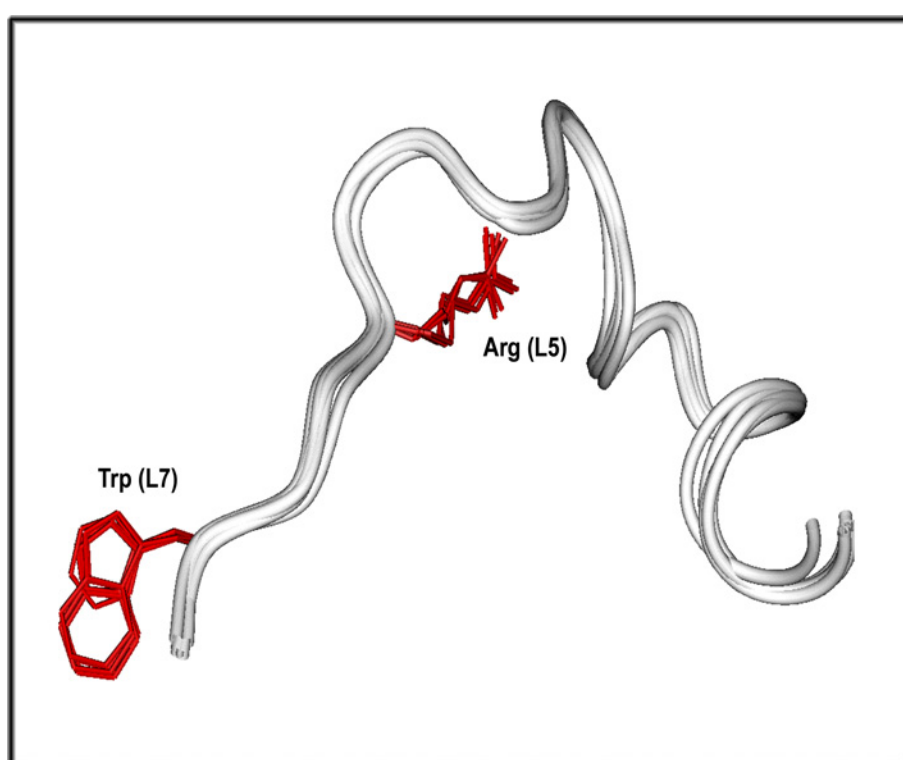

Figure 3.7. Representation of the loops of sub-class $\alpha\beta$10.1.1. Only side-chains with functional annotation are shown (Arg in position L6 and Trp in positon L8).

3.3.6 Application in kinase loops modeling

To test the potential application of PSSMs for loop modeling, a prediction test was applied to the following sub-classes: $\beta\beta$links5.1.1, $\beta\beta$links8.1.1, $\beta\alpha$6.1.1, $\beta\alpha$6.2.1 and $\alpha\alpha$5.1.1. One loop of each

sub-class was extracted to be used as a query and the PSSMs were recalculated without the query loop. The query loop was aligned with the profiles of the same type of bracing secondary structures using the algorithm of FUGUE(Shi et al. 2001) and conformation substitution matrices derived from a more general classification of loops(Espadaler et al. 2004). The Zscore was calculated for a query loop with the mean value (μ) and standard deviation (σ) obtained with all the scores of the alignments of the query.

The results of the predictions are shown in table 3.4. It is noteworthy that this procedure could correctly predict in the first rank the two sub-class of p-loops ($\beta\alpha$6.1.1 and $\beta\alpha$6.2.1).

Table 3.4 Test for prediction of sub-class assignment.

PDB code	Chain	Fragment	Rank	Zscore	Number of sub-classes	Sub-class
3lck	-	378-389	1	6.128	12	$\beta\beta_{links}$5.1.1
1k3a	A	1101-1114	1	5.695	12	$\beta\beta_{links}$8.1.1
1zip	-	4-18	1	7.489	46	$\beta\alpha$6.1.1
1jb1	A	151-171	1	5.603	46	$\beta\alpha$6.2.1
2ak3	A	56-77	1	5.070	51	$\alpha\alpha$5.1.1

The sequence fragment of the loop used as query is indicated by the PDB code of the protein (first column), the chain (second column, where hyphen means none), and the interval of residues (third column). The rank and Zscore for the correct prediction are indicated in the fourth and fifth columns. The last two columns indicate the total number of sub-classes with the same bracing secondary structure of the loop used for the prediction and the correct sub-class of the query-loop.

3.4   Discussion and Conclusions

The present study has shown the classification of loops of kinases. Besides the practical utility of the classification in comparative modeling or in the resolution of fragments of protein kinases when the experimental information is scarce, we have substantiated one of the questions raised in previous works(Cheek et al. 2002; Russell et al. 2002): different folds can share similar catalytic mechanism in kinases. Moreover, we have investigated the relation between this mechanism and the presence of a particular sequence for structural motifs with functional information.

Five main functions were identified (for catalysis, ATP-binding, substrate binding, ion binding and, additionally, oligomerization) and specific loops for these functions were characterized. A set of these loops had already been described in the literature and identified as kinase-motifs (i.e. P-loop, Gly-rich-loop and LID domain, among others). The work has automatically identified, clustered and classified the common structures of these motifs.

We were able to recognize the location of important residues for the function within the loops of kinases. These were found by conservation in sequence and structure among kinases belonging to the same fold, super-family or family and differentiated by comparison of its function. This procedure allowed us to determine the relevance of particular residues. Besides, the identification by close distance of ions, heteroatoms, substrates or other macromolecules to the residues of a loop permitted to better characterize its hypothetical function and the function of its residues. This clearly helps on the study of the catalytic mechanism of kinases.

Most sub-classes with functional information belonged to the $\beta\alpha$ type motif (about a 50% of the total). Also some loops of type $\alpha\alpha$ and $\alpha\beta$ were noteworthy, but few were $\beta\beta_{hairpins}$ or $\beta\beta_{links}$. It is clear that the main parts of the kinase domains were α/β and the Rossmann fold was likely the most used domain involved on kinase activity, consequently most loops were $\beta\alpha$ and $\alpha\beta$, i.e. those involved in the catalytic mechanism and substrate or ATP binding.

Additionally, we found the P-loop repeated in at least two different folds and six different families, this being a classical on the catalytic mechanism of kinases. Besides, the Gly-rich-loop and the DGF motifs were found in several domains too. However, most loops with functional activity were classified with members of the same fold and, with the exception of the P-loop, those shared by different folds were short and with ambiguous functions. We conclude from this result that different folds shared these loops for structural, not functional, reasons.

On the other hand, loops within the same family with specific activity were clustered together and residues for specific binding were easily identified. On the basis that clusters are obtained on structure, the specificity for binding was characterized by changes in sequence, and common features were used for inferring similar function. Therefore, the case of P-loop, with Gly residues conserved at specific distances forming a particular sequence pattern, performs the same function in several families, producing a conformation to accommodate the phosphate chain of ATP, while Lys and Ser at the end of the pattern orient the position of ATP by binding the phosphate.

The P-loop region is influenced by other residues at close distance that come from other loops. This was already presented by Via et al.(Via et al. 2000) showing four regions associated to the P-loop. We have obtained clusters of loops for these regions, proving that some of these residues can be provided by different types of loops, while others are more explicit. The GK P-loop element and the second

95

negative charge are found in few sub-classes ($\beta\alpha6.1.1$, $\beta\alpha6.1.2$ and $\beta\alpha4.1.1$). Sub-classes $\beta\alpha6.1.1$ and $\beta\alpha6.2.1$ contain loops from two different folds and six families, while $\beta\alpha4.1.1$ is more specific for fold "P-loop containing nucleotide triphosphate hydrolases" and is found in 2 families. But the sub-classes containing the first negative charge and the conserved positive charge show a wider distribution of folds. A total of 8 sub-classes with these functions were identified with different types of bracing secondary structures ($\alpha\alpha$, $\alpha\beta$ and $\beta\alpha$). The assortment probes the diversity by which kinases yield a similar function, but identifies the characteristic common motifs for specific folds and families. Consequently, the classification helps to relate function and structure and to facilitate the pattern that can be used to determine ancestor relationships.

Some of these characteristics were already used in the work of Leipe et al.(Leipe et al. 2003) to reconstruct the major stages of the evolution of P-loop kinases by acquisition of additional elements from the common ancestor that contained a Walker A and a Walker B motifs and agrees with our results. Two sub-classes (of the same loop class) for Walker A motif that implicate high conservation but five sub-classes (with different loop class and type, and larger difference in sequence pattern) for Walker B motif point to a less conservation through the evolutionary process.

The presence of more than one common motif in two proteins has been used to characterize and extend the regions of a common core between different families of kinases (and also folds). This was used to link by evolutionary relationship the members of PEP carboxykinase and the family of HPr kinase through the P-loop superimposition. Here, the example suggested the use of these patterns and its extension by neighbors to characterize a common evolutionary ancestor pattern(Russell et al. 2002).

This classification may help to facilitate the use of patterns for identifying shared features between families on the general classification of kinases. In particular as suggested by Russell et al.(Russell et

al. 2002), we have connected folds of P-loop containing proteins together with additional shared loops.

That leads us to classify these kinases within the same fold and super-family.

3.6 References

Auerbach, G., R. Huber, M. Grattinger, K. Zaiss, H. Schurig, R. Jaenicke and U. Jacob (1997). "*Closed structure of phosphoglycerate kinase from Thermotoga maritima reveals the catalytic mechanism and determinants of thermal stability.*" Structure **5**(11): 1475-83.

Bernstein, B. E., D. M. Williams, J. C. Bressi, P. Kuhn, M. H. Gelb, G. M. Blackburn and W. G. Hol (1998). "*A bisubstrate analog induces unexpected conformational changes in phosphoglycerate kinase from Trypanosoma brucei.*" J Mol Biol **279**(5): 1137-48.

Berry, M. B., B. Meador, T. Bilderback, P. Liang, M. Glaser and G. N. Phillips, Jr. (1994). "*The closed conformation of a highly flexible protein: the structure of E. coli adenylate kinase with bound AMP and AMPPNP.*" Proteins **19**(3): 183-98.

Bilwes, A. M., C. M. Quezada, L. R. Croal, B. R. Crane and M. I. Simon (2001). "*Nucleotide binding by the histidine kinase CheA.*" Nat Struct Biol **8**(4): 353-60.

Blaszczyk, J., G. Shi, H. Yan and X. Ji (2000). "*Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 A resolution.*" Structure Fold Des **8**(10): 1049-58.

Bossemeyer, D., R. A. Engh, V. Kinzel, H. Ponstingl and R. Huber (1993). "*Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 A structure of the complex with Mn2+ adenylyl imidodiphosphate and inhibitor peptide PKI(5-24).*" Embo J **12**(3): 849-59.

Briozzo, P., B. Golinelli-Pimpaneau, A. M. Gilles, J. F. Gaucher, S. Burlacu-Miron, H. Sakamoto, J. Janin and O. Barzu (1998). "*Structures of escherichia coli CMP kinase alone and in complex with CDP: a new fold of the nucleoside monophosphate binding domain and insights into cytosine nucleotide specificity.*" Structure **6**(12): 1517-27.

Burk, D. L., W. C. Hon, A. K. Leung and A. M. Berghuis (2001). "*Structural analyses of nucleotide binding to an aminoglycoside phosphotransferase.*" Biochemistry **40**(30): 8756-64.

Buss, K. A., D. R. Cooper, C. Ingram-Smith, J. G. Ferry, D. A. Sanders and M. S. Hasson (2001). "*Urkinase: structure of acetate kinase, a member of the ASKHA superfamily of phosphotransferases.*" J Bacteriol **183**(2): 680-6.

Campobasso, N., Mathews, II, T. P. Begley and S. E. Ealick (2000). "*Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from Bacillus subtilis at 1.5 A resolution.*" Biochemistry **39**(27): 7868-77.

Cheek, S., H. Zhang and N. V. Grishin (2002). "*Sequence and structure classification of kinases.*" J Mol Biol **320**(4): 855-81.

Cheng, G., E. M. Bennett, T. P. Begley and S. E. Ealick (2002). "*Crystal structure of 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase from Salmonella typhimurium at 2.3 A resolution.*" Structure (Camb) **10**(2): 225-35.

Cherfils, J., S. Morera, I. Lascu, M. Veron and J. Janin (1994). "*X-ray structure of nucleoside diphosphate kinase complexed with thymidine diphosphate and Mg2+ at 2-A resolution.*" Biochemistry **33**(31): 9062-9.

Copley, R. R., R. B. Russell and C. P. Ponting (2001). "*Sialidase-like Asp-boxes: sequence-similar structures within different protein folds.*" Protein Sci **10**(2): 285-92.

Davies, G. J., S. J. Gamblin, J. A. Littlechild and H. C. Watson (1993). "*The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent.*" Proteins **15**(3): 283-9.

Donate, L. E., S. D. Rufino, L. H. Canard and T. L. Blundell (1996). "*Conformational analysis and clustering of short and medium size loops connecting regular secondary structures: a database for modeling and prediction.*" Protein Sci **5**(12): 2600-16.

Efimov, A. V. (1991). "*Structure of coiled beta-beta-hairpins and beta-beta-corners.*" FEBS Lett **284**(2): 288-92.

Efimov, A. V. (1994). "*Favoured structural motifs in globular proteins.*" Structure **2**(11): 999-1002.

Eriksen, T. A., A. Kadziola, A. K. Bentsen, K. W. Harlow and S. Larsen (2000). "*Structural basis for the function of Bacillus subtilis phosphoribosyl-pyrophosphate synthetase.*" Nat Struct Biol **7**(4): 303-8.

Espadaler, J., N. Fernandez-Fuentes, A. Hermoso, E. Querol, F. X. Aviles, M. J. Sternberg and B. Oliva (2004). "*ArchDB: Automated protein loop classification as a tool for Structural Genomics.*" Nucleic Acids Res **32**: D185-D188.

Fieulaine, S., S. Morera, S. Poncet, V. Monedero, V. Gueguen-Chaignon, A. Galinier, J. Janin, J. Deutscher and S. Nessler (2001). "*X-ray structure of HPr kinase: a bacterial protein kinase with a P-loop nucleotide-binding domain.*" Embo J **20**(15): 3917-27.

Fischer, E. H. and E. G. Krebs (1955). "*Conversion of phosphorylase b to phosphorylase a in muscle extracts.*" J Biol Chem **216**(1): 121-32.

Fritz-Wolf, K., T. Schnyder, T. Wallimann and W. Kabsch (1996). "*Structure of mitochondrial creatine kinase.*" Nature **381**(6580): 341-5.

Goldberg, J., A. C. Nairn and J. Kuriyan (1996). "*Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I.*" Cell **84**(6): 875-87.

Herzberg, O., C. C. Chen, S. Liu, A. Tempczyk, A. Howard, M. Wei, D. Ye and D. Dunaway-Mariano (2002). "*Pyruvate site of pyruvate phosphate dikinase: crystal structure of the enzyme-phosphonopyruvate complex, and mutant analysis.*" Biochemistry **41**(3): 780-7.

Hubbard, S. R. (1999). "*Src autoinhibition: let us count the ways.*" Nat Struct Biol **6**(8): 711-4.

Hubbard, S. R., L. Wei, L. Ellis and W. A. Hendrickson (1994). "*Crystal structure of the tyrosine kinase domain of the human insulin receptor.*" Nature **372**(6508): 746-54.

Hunter, T. and G. D. Plowman (1997). "*The protein kinases of budding yeast: six score and more.*" Trends Biochem Sci **22**(1): 18-22.

Hurley, J. H., Y. Tsujishita and M. A. Pearson (2000). "*Floundering about at cell membranes: a structural view of phospholipid signaling.*" Curr Opin Struct Biol **10**(6): 737-43.

Izard, T. and J. Ellis (2000). "*The crystal structures of chloramphenicol phosphotransferase reveal a novel inactivation mechanism.*" Embo J **19**(11): 2690-700.

Johnson, L. N., E. D. Lowe, M. E. Noble and D. J. Owen (1998). "*The Eleventh Datta Lecture. The structural basis for substrate recognition and control by protein kinases.*" FEBS Lett **430**(1-2): 1-11.

Johnson, L. N., M. E. Noble and D. J. Owen (1996). "*Active and inactive protein kinases: structural basis for regulation.*" Cell **85**(2): 149-58.

Jurica, M. S., A. Mesecar, P. J. Heath, W. Shi, T. Nowak and B. L. Stoddard (1998). "*The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate.*" Structure **6**(2): 195-210.

Kabsch, W. and C. Sander (1983). "*Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.*" Biopolymers **22**(12): 2577-637.

Kobe, B., J. Heierhorst, S. C. Feil, M. W. Parker, G. M. Benian, K. R. Weiss and B. E. Kemp (1996). "*Giant protein kinases: domain interactions and structural basis of autoregulation.*" Embo J **15**(24): 6810-21.

Kraft, L., G. A. Sprenger and Y. Lindqvist (2002). "*Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography.*" J Mol Biol **318**(4): 1057-69.

Krell, T., J. R. Coggins and A. J. Lapthorn (1998). "*The three-dimensional structure of shikimate kinase.*" J Mol Biol **278**(5): 983-97.

Krishna, S. S., T. Zhou, M. Daugherty, A. Osterman and H. Zhang (2001). "*Structural basis for the catalysis and substrate specificity of homoserine kinase.*" Biochemistry **40**(36): 10810-8.

Larsen, T. M., M. M. Benning, I. Rayment and G. H. Reed (1998). "*Structure of the bis(Mg2+)-ATP-oxalate complex of the rabbit muscle pyruvate kinase at 2.1 A resolution: ATP binding over a barrel.*" Biochemistry **37**(18): 6247-55.

Larsen, T. M., M. M. Benning, G. E. Wesenberg, I. Rayment and G. H. Reed (1997). "*Ligand-induced domain movement in pyruvate kinase: structure of the enzyme from rabbit muscle with Mg2+, K+, and L-phospholactate at 2.7 A resolution.*" Arch Biochem Biophys **345**(2): 199-206.

Larsen, T. M., L. T. Laughlin, H. M. Holden, I. Rayment and G. H. Reed (1994). "*Structure of rabbit muscle pyruvate kinase complexed with Mn2+, K+, and pyruvate.*" Biochemistry **33**(20): 6301-9.

Lavie, A., M. Konrad, R. Brundiers, R. S. Goody, I. Schlichting and J. Reinstein (1998). "*Crystal structure of yeast thymidylate kinase complexed with the bisubstrate inhibitor P1-(5'-adenosyl) P5-(5'-thymidyl) pentaphosphate (TP5A) at 2.0 A resolution: implications for catalysis and AZT activation.*" Biochemistry **37**(11): 3677-86.

Leipe, D. D., E. V. Koonin and L. Aravind (2003). "*Evolution and Classification of P-loop Kinases and Related Proteins.*" J Mol Biol **333**: 781-815.

Lupas, A. N., C. P. Ponting and R. B. Russell (2001). "*On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world?*" J Struct Biol **134**(2-3): 191-203.

Machius, M., J. L. Chuang, R. M. Wynn, D. R. Tomchick and D. T. Chuang (2001). "*Structure of rat BCKD kinase: nucleotide-induced domain communication in a mitochondrial protein kinase.*" Proc Natl Acad Sci U S A **98**(20): 11218-23.

MacRae, I. J., I. H. Segel and A. J. Fisher (2001). "*Crystal structure of ATP sulfurylase from Penicillium chrysogenum: insights into the allosteric regulation of sulfate assimilation.*" Biochemistry **40**(23): 6795-804.

Mathews, II, M. D. Erion and S. E. Ealick (1998). "*Structure of human adenosine kinase at 1.5 A resolution.*" Biochemistry **37**(45): 15607-20.

Matte, A., L. W. Tari and L. T. Delbaere (1998). "*How do kinases transfer phosphoryl groups?*" Structure **6**(4): 413-9.

Mirny, L. and E. Shakhnovich (2001). "*Protein folding theory: from lattice to all-atom models.*" Annu Rev Biophys Biomol Struct **30**: 361-96.

Muller-Dieckmann, H. J. and G. E. Schulz (1994). "*The structure of uridylate kinase with its substrates, showing the transition state geometry.*" J Mol Biol **236**(1): 361-7.

Oliva, B., P. A. Bates, E. Querol, F. X. Aviles and M. J. Sternberg (1997). "*An automated classification of the structure of protein loops.*" J Mol Biol **266**(4): 814-30.

Oliva, B., P. A. Bates, E. Querol, F. X. Aviles and M. J. Sternberg (1998). "*Automated classification of antibody complementarity determining region 3 of the heavy chain (H3) loops into canonical forms and its application to protein structure prediction.*" J Mol Biol **279**(5): 1193-210.

Ramon-Maiques, S., A. Marina, M. Uriarte, I. Fita and V. Rubio (2000). "*The 1.5 A resolution crystal structure of the carbamate kinase-like carbamoyl phosphate synthetase from the hyperthermophilic Archaeon pyrococcus furiosus, bound to ADP, confirms that this thermostable enzyme is a carbamate kinase, and provides insight into substrate binding and stability in carbamate kinases.*" J Mol Biol **299**(2): 463-76.

Rao, J. K., G. Bujacz and A. Wlodawer (1998). "*Crystal structure of rabbit muscle creatine kinase.*" FEBS Lett **439**(1-2): 133-7.

Rao, V. D., S. Misra, I. V. Boronenkov, R. A. Anderson and J. H. Hurley (1998). "*Structure of type IIbeta phosphatidylinositol phosphate kinase: a protein kinase fold flattened for interfacial phosphorylation.*" Cell **94**(6): 829-39.

Romanowski, M. J., J. B. Bonanno and S. K. Burley (2002). "*Crystal structure of the Streptococcus pneumoniae phosphomevalonate kinase, a member of the GHMP kinase superfamily.*" Proteins **47**(4): 568-71.

Rose, G. D., L. M. Gierasch and J. A. Smith (1985). "*Turns in peptides and proteins.*" Adv Protein Chem **37**: 1-109.

Russell, R. B. (1998). "*Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution.*" J Mol Biol **279**(5): 1211-27.

Russell, R. B., J. A. Marquez, W. Hengstenberg and K. Scheffzek (2002). "*Evolutionary relationship between the bacterial HPr kinase and the ubiquitous PEP-carboxykinase: expanding the P-loop nucleotidyl transferase superfamily.*" FEBS Lett **517**(1-3): 1-6.

Saraste, M., P. R. Sibbald and A. Wittinghofer (1990). "*The P-loop--a common motif in ATP- and GTP-binding proteins.*" Trends Biochem Sci **15**(11): 430-4.

Scheffzek, K., W. Kliche, L. Wiesmuller and J. Reinstein (1996). "*Crystal structure of the complex of UMP/CMP kinase from Dictyostelium discoideum and the bisubstrate inhibitor P1-(5'-adenosyl) P5-(5'-uridyl) pentaphosphate (UP5A) and Mg2+ at 2.2 A: implications for water-mediated specificity.*" Biochemistry **35**(30): 9716-27.

Schenk, P. W. and B. E. Snaar-Jagalska (1999). "*Signal perception and transduction: the role of protein kinases.*" Biochim Biophys Acta **1449**(1): 1-24.

Sherman, M. A., B. K. Szpikowska, S. A. Dean, A. M. Mathiowetz, N. L. McQueen and M. T. Mas (1990). "*Probing the role of arginines and histidines in the catalytic function and activation of yeast 3-phosphoglycerate kinase by site-directed mutagenesis.*" J Biol Chem **265**(18): 10659-65.

Shi, J., T. L. Blundell and K. Mizuguchi (2001). "*FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties.*" J Mol Biol **310**: 243-257.

Shirakihara, Y. and P. R. Evans (1988). "*Crystal structure of the complex of phosphofructokinase from Escherichia coli with its reaction products.*" J Mol Biol **204**(4): 973-94.

Sicheri, F., I. Moarefi and J. Kuriyan (1997). "*Crystal structure of the Src family tyrosine kinase Hck.*" Nature **385**(6617): 602-9.

Sigrell, J. A., A. D. Cameron, T. A. Jones and S. L. Mowbray (1998). "*Structure of Escherichia coli ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insights into a new family of kinase structures.*" Structure **6**(2): 183-93.

Sinev, M. A., E. V. Sineva, V. Ittah and E. Haas (1996). "*Domain closure in adenylate kinase.*" Biochemistry **35**(20): 6425-37.

Sinha, S. C. and J. L. Smith (2001). "*The PRT protein family.*" Curr Opin Struct Biol **11**(6): 733-9.

Steinbacher, S., P. Hof, L. Eichinger, M. Schleicher, J. Gettemans, J. Vandekerckhove, R. Huber and J. Benz (1999). "*The crystal structure of the Physarum polycephalum actin-fragmin kinase: an atypical protein kinase with a specialized substrate-binding domain.*" Embo J **18**(11): 2923-9.

Steussy, C. N., K. M. Popov, M. M. Bowker-Kinley, R. B. Sloan, Jr., R. A. Harris and J. A. Hamilton (2001). "*Structure of pyruvate dehydrogenase kinase. Novel folding pattern for a serine protein kinase.*" J Biol Chem **276**(40): 37443-50.

Timm, D. E., J. Liu, L. J. Baker and R. A. Harris (2001). "*Crystal structure of thiamin pyrophosphokinase.*" J Mol Biol **310**(1): 195-204.

Van Rompay, A. R., M. Johansson and A. Karlsson (2000). "*Phosphorylation of nucleosides and nucleoside analogs by mammalian nucleoside monophosphate kinases.*" Pharmacol Ther **87**(2-3): 189-98.

Venkatachalam, C. M. (1968). "*Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units.*" Biopolymers **6**(10): 1425-36.

Via, A., F. Ferre, B. Brannetti, A. Valencia and M. Helmer-Citterich (2000). "*Three-dimensional view of the surface motif associated with the P-loop structure: cis and trans cases of convergent evolution.*" J Mol Biol **303**(4): 455-65.

Williams, R. L., D. A. Oren, J. Munoz-Dorado, S. Inouye, M. Inouye and E. Arnold (1993). "*Crystal structure of Myxococcus xanthus nucleoside diphosphate kinase and its interaction with a nucleotide substrate at 2.0 A resolution.*" J Mol Biol **234**(4): 1230-47.

Wintjens, R. T., M. J. Rooman and S. J. Wodak (1996). "*Automatic classification and analysis of alpha alpha-turn motifs in proteins.*" J Mol Biol **255**(1): 235-53.

Wright, F. A., W. J. Lemon, W. D. Zhao, R. Sears, D. Zhuo, J. P. Wang, H. Y. Yang, T. Baer, D. Stredney, J. Spitzner, A. Stutz, R. Krahe and B. Yuan (2001). "*A draft annotation and overview of the human genome.*" Genome Biol **2**(7): RESEARCH0025.

Xie, X., Y. Gu, T. Fox, J. T. Coll, M. A. Fleming, W. Markland, P. R. Caron, K. P. Wilson and M. S. Su (1998). "*Crystal structure of JNK3: a kinase implicated in neuronal apoptosis.*" Structure **6**(8): 983-91.

Xu, R. M., G. Carmel, J. Kuret and X. Cheng (1996). "*Structural basis for selectivity of the isoquinoline sulfonamide family of protein kinase inhibitors.*" Proc Natl Acad Sci U S A **93**(13): 6308-13.

Xu, R. M., G. Carmel, R. M. Sweet, J. Kuret and X. Cheng (1995). "*Crystal structure of casein kinase-1, a phosphate-directed protein kinase.*" Embo J **14**(5): 1015-23.

Yamaguchi, H., M. Matsushita, A. C. Nairn and J. Kuriyan (2001). "*Crystal structure of the atypical protein kinase domain of a TRP channel with phosphotransferase activity.*" Mol Cell **7**(5): 1047-57.

Yang, D., L. W. Shipman, C. A. Roessner, A. I. Scott and J. C. Sacchettini (2002). "*Structure of the Methanococcus jannaschii mevalonate kinase, a member of the GHMP kinase superfamily.*" J Biol Chem **277**(11): 9462-7.

Yuen, M. H., X. L. Wang, H. Mizuguchi, K. Uyeda and C. A. Hasemann (1999). "*A switch in the kinase domain of rat testis 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase.*" Biochemistry **38**(38): 12333-42.

Zhou, G., T. Somasundaram, E. Blanc, G. Parthasarathy, W. R. Ellington and M. S. Chapman (1998). "*Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions.*" Proc Natl Acad Sci U S A **95**(15): 8449-54.