Biol. Stud. 2015: 9(1); 5–14 • DOI: https://doi.org/10.30970/sbi.0901.415
www.http://publications.lnu.edu.ua/journals/index.php/biology

UDC 577.322.23

COMPARATIVE ANALYSIS OF SH2 DOMAIN STRUCTURES

V. V. Hurmach[1]***, M. O. Platonov***[2]***, O. M. Boyko***[1]***, Yu. I. Prylutskyy***[1]

[1]*Taras Shevchenko National University of Kyiv, 64, Volodymyrska St., Kyiv 01601, Ukraine*
e-mail: gyrmach@gmail.com

[2]*Institute of Molecular Biology and Genetics, NAS of Ukraine*
150, Akademik Zabolotnyi St., Kyiv 03680, Ukraine

Src Homology 2 (SH2) is the compact globular domains, which is involved in intracellular signaling pathways and play an important role in mediating specific protein-protein interactions. It consists of about 100 amino acids and include sevens β-sheets and two α-helices. The SH2 domains comprise two highly conservative parts of binding pocket – pTyr and pTyr +3. The knowledge of the protein complexes structure is an important step forward to understanding of the mechanisms of their functioning. The binding site of SH2 domains and surrounding to binding site sequences were analyzed by using *in silico* methods. All SH2 domains were divided into the groups by sequence similarity. The parts of sequences which are common to all domains and the unique parts of certain domains were found within the framework of conservation and similarity analysis of SH2 domains. Furthermore, the surface area analysis displays that the highly conservative structures occupy the smallest area. These results indicate the ability of the SH2 domains to recognize not only linear phosphopeptide sequences. It opens new insights on the interpretation of possible mechanisms of interaction between SH2 domains and ligands/protreins (e. g. possibility of binding between protein or ligand with SH2 domain not only within binding site area).

Keywords: SH2 domain, binding site, conservativeness.

INTRODUCTION

Proteins are characterized by sequence similarity relationships. The understanding of protein relationships is important for the genome sequences annotation [12]. Proteins with high sequence identity and high structural similarity commonly have functional similarity and evolutionary relationships. Examples of proteins deviating from this general relationship of sequence/structure/function homology are well-recognized. For example, high sequence identity and low structure similarity can occur from conformational plasticity, mutations, solvent effects and ligand binding.

SH2 domains were identified across structure and sequence similarity. These domains are highly structurally conserved motifs comprising of about 100 amino acids; they participate in large number of signal-transduction between proteins. Human genome

encodes about 120 SH2 domains which are present in 110 proteins such as kinases (Src, Lck), phosphatases (SHP2, SHIP2), phospholipases (PLCγ1), transcription factors (STAT), regulatory proteins (SOCS), adapter proteins (Grb2), structural proteins (SHC) and others. Wide spread of SH2 domains in animals and their almost complete absence in microorganisms (e.g. primitive SH2 fragments in yeast) may testify to their appearance related to complication of signal transduction mechanisms in multicellular organisms [14].

Latest research has shown that it is possible to divide SH2 domains according to the recognition specificity of pTyr residue with C-terminus. Such recognition may take place in residues position +1, +2 and +3, bind to pTyr [10]. So, each SH2 domain binds only to specific phosphotyrosine-containing fragments. For example, the Src SH2 domain mostly recognizes Glu-Glu-Ile (binding fragment pYEEI), whereas the Grb2 SH2 domain binds to another fragment – pYVNV. However, complete understanding of this effect requires detailed study of thermodynamic peculiarities of the interaction between SH2 domains and phosphopeptides.

The 3D and 1D structure of SH2 domain was studied by computer modeling methods. They are analyzed according to sequence and structure similarity. The conservation which is not present across whole SH2 domains, was found. It shows a difference between groups at the amino acid level. So, conversation of some groups is larger than in others. Also, the several common motifs of binding pocket and spatial close part were selected. Surface analysis identified potential opportunity of binding in selected conserved regions in binding site and other parts of domain.

The aim of this work was to select the main conservative regions of SH2 domain, which can take part in protein-protein and ligand-protein interactions, and to compare the selected sequences within every separate group and across whole domain.

MATERIALS AND METHODS

3D dataset of SH2 domain. 219 (66 – Nuclear magnetic resonance, 153 – X-ray crystallography) 3D crystal structures of SH2 domains were retrieved from PDB (Protein Databank) [2]. It was taken into account that there is more than one SH2 domain per file. Consequently, they were divided in 1129 separate coordinate structures. Obtained structures were used in further calculation steps.

3D structural environment analysis of SH2 domain binding pocket. 3D structure analysis of the SH2 domain binding site was done by using Chimera [5] and helixweb [6, 9] software tools.

Protein residues with at least one atom takes part in the binding pocket creating (even if residue atom occupies minimal area of pocket), were included in the binding site definition. The residue selection was done by evaluating ASA (Accessible Surface Area) of protein residue. So, if amino acid occupies more than 5 $Å^2$ and locates within atomic distance of 5 Å or less from the nearest ligand atom (makes strong contacts), it is considered as amino acid with a large contribution to the binding pocket creating. However, amino acid with less area or bigger distance between the nearest atoms of interacting structures is considered as a support of amino acid binding in pocket (without significant contribution to the binding pocket creating).

The comparison of SH2 domain structures. Step by step, whole available coordinate files were compared. Firstly, the comparison was provided within every separate ID by using ClustalX [15] (sequence comparison) and Chimera [4] (comparison of rmsd between structures) software. Then, the same procedure was applied for PDB struc-

tures with different ID, 56 structures were selected. Finally, utilizing ClustalX sequences were aligned and edited using Jalview [3] and divided to six groups.

Residue columns analysis. Residue consistency and sequence similarity of selected groups were assessed using PRALINE software [13].

The conservation score were calculated in the Scorecons Server [1] by following entropy scores method calculation. Entropy scores used here normalize Shannon's entropy so that conserved (low entropy) columns score 1 and diverse (high entropy) columns score 0. The entropy scores for every position within the alignment defined as

$$C_{\text{entropy}} = \frac{-\sum_a^K P_a \log_2 P_a}{\log_2(\min(N, K))},$$

where N is a number of residues in column, K is a number of residue types and $P_a = n_a/N$ (n_a is a number of residues of type a). In this part of work two variants of calculation were done: (1) amino acids were classified into one of K = 21 types: 20 standard amino acid types + 1 gap type; (2) amino acids were classified into one of K = 7 types: aliphatic (AVLIMC), aromatic (FWYH), polar (STNQ), positive (KR), negative (DE), special conformations (GP) and gaps.

Also, it should be note that all calculations present above are done in BLOSSUM62 substitution matrix [11]. However, before a substitution matrix is used, it must first be transformed into a convenient range. Therefore a Karlinlike transformation was applied [8].

RESULTS AND DISCUSSION

Selection of every separate group. The selection of every SH2 domain group was done in such order: (1) 346 PDB structures were retrieved from protein data bank [2]; (2) they were divided to 1129 structures; (3) then those obtained coordinate files were compared among every separate PDB ID of structures and if there were any differences between compared structures, they were considered as different structures. As a result 56 distinctive structures were obtained (see below in SH2 domain characteristics within the groups and Fig. 2); (4) in the next stage whole obtained structures of SH2 domains have been divided into six groups based on amino acid similarity (Fig. 2, Table).

For example, based on this algorithm the domain structure of 2K7A (group 5) presents information about seven others PDB (1LUI. (residues 238–344), 1LUK.A (residues 238–344), 1LUM.A (residues 238–344), 1LUN.A (residues 238–344), 2K79.B, 2ETZ.A and 2EU0.A (all of them are Tyrosine-protein kinase ITK/TSK and contains in Mus musculus) structures and contains 18 similar coordinate files. It means that only one structure was selected to next comparison stage.

As one can, see the structures are almost 100 % similar: alignment score per residue pair is 18.74, the average RMSD between compared domains is 1.6 Å (Fig. 1).

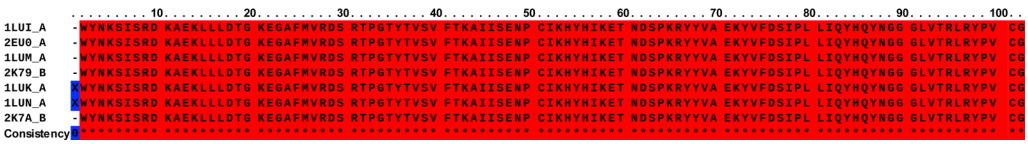

Fig. 1. An example of the sequence comparison of structures, represented by 2K7A coordinate (PDB) structure

Рис. 1. Приклад порівняння послідовностей структур, які представлені репрезентативною структурою 2K7A (PDB)

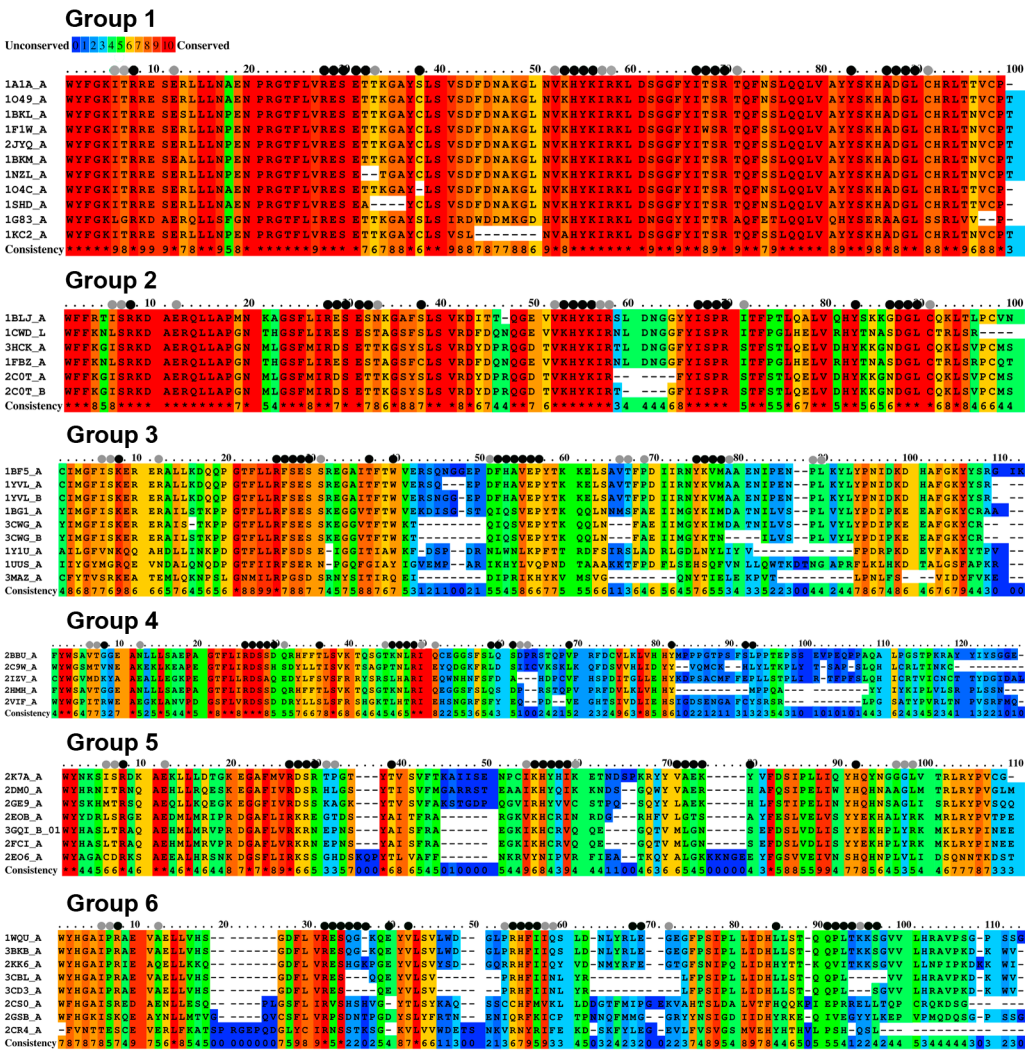

Fig. 2. Multiple sequence alignments of SH2 domains. The key residues in the binding pocket are highlighted in big black points (creates a main part of binding pocket). Grey indicates residue without a significant contribution in binding pocket creating

Рис. 2. Множинне вирівнювання послідовностей доменів SH2. Основні амінокислотні залишки сайту зв'язування позначені чорними крапками (утворюють основні частину сайту зв'язування). Сірим кольором відмічені амінокислотні залишки, які не роблять значного внеску в утворення сайту зв'язування

SH2 domain characteristics within the groups. In all cases the quality parameters are much more in selected groups than such parameters in all SH2 domains (e. g., the percent of sequence identity across whole domains is 31; in turn, in group 1 – 91 % and in binding pocket of group 2 – 92 %) (Table).

It is significant that protein types in every separate group are unique, group 1 – c-src tyrosine kinase (*Homo sapiens* – 1A1A,1SHD, Rous sarcoma virus – 1BKL, 1KC2), Proto-oncogene tyrosine-protein kinase src (*Homo sapiens* – 1O4C, 1O49, *Gallus gallus* – 1F1W),

The list of amino acids in SH2 domain binding pocket for different groups. The parameters of every separate group of SH2 domains and its binding pocket are shown as Poc.-binding site and Pro.-whole proteins. The position in sequence in Fig. 2 is shown in (...) and the number of motif is shown in [...]. Conservation score parameters are presented in one column (K = 21/K =7)

Список розташування амінокислот сайту зв'язування доменів SH2. Параметри окремих груп доменів SH2 та їх сайту зв'язування показані як Poc.-binding site і Pro.-whole proteins. Точки на рис. 2 показані в (...), номер мотиву показаний в [...]. Значення консервативності наведені в одній колонці (K = 21/K = 7)

Group and shab PDB	Environment of pocket	Sequence identity %/ similarity %		Average RMSD Å		Average surface area Å^2		Conservation score	
		Poc.	Pro.	Poc.	Pro	Poc.	Pro	Poc.	Pro
Group 1 1A1A	[1]. Ile 156 (6), Thr 157 (7), Arg 158 (8). [2]. Glu 162 (12). [3]. Arg 178 (28), Glu 179 (29), Ser 180 (30), Glu 181 (31), Thr 182 (32), Thr 182 (33). [4]. Ser 188 (38). [5]. Val 202 (52), Lys 203 (53), His 204 (54), Tyr 205 (55), Lys 206 (56). Ile 207 (57), Arg 208 (58). [6]. Ile 217 (67), Thr 218 (68), Ser 219 (69), Arg 220 (70), Thr 221 (71). [7]. Tyr 233 (83). [8]. Ala 237 (87), Asp 238 (88), Gly 239 (89), Leu 240 (90), Cys 241 (91)	0.92	0.91	0.72	0.79	1268	5876	0.94/ 0.93	0.92/ 0.91
Group 2 1BLJ	[1]. Ile 19 (6), Ser 20 (7), Arg 21 (8). [2]. Ser 25 (12). [3]. Arg 41 (28), Glu 42 (29), Ser 43 (30), Glu 44 (31), Ser 45 (32), Asn 46 (33). [4]. Ser 51 (38). [5]. Val 64 (52), Lys 65 (53), His 66 (54), Tyr 67 (55), Lys 68 (56), Ile 69 (57), Arg 70 (58). [6]. Ile 79 (67), Ser 80 (68), Pro 81 (69), Arg 82 (70), Ile 83 (71). [7]. Tyr 95 (83). [8]. Gly 99 (87), Asp 100 (88), Gly 101 (89), Leu 102 (90), Cys 103 (91).	0.85	0.78	1.14	1.32	1491	5856	0.89/ 0.96	0.81/ 0.87
Group 3 1BF5	[1]. Ile 582 (6), Ser 583 (7), Lys 584 (8). [2]. Arg 588 (12). [3]. Arg 602 (26), Phe 603 (27), Ser 604 (28), Glu 605 (29), Ser 606 (30), Ser 607 (31). [4]. Trp 616 (40). [5]. Asp 627 (51), Phe 628 (52), His 629 (53), Ala 630 (54), Val 631 (55), Glu 632 (56), Pro 633 (57). [6]. Val 642 (66), Thr 643 (67), Pro 645 (69). [7]. Tyr 651 (75), Lys 652 (76), Val 653 (77), Met 654 (78), Ala 655 (79). [8]. Pro 633 (89), Leu 634 (90). [9]. Thr 613 (37)	0.5	0.47	1.53	1.65	1552	6945	0.59/ 0.67	0.56/ 0.62

End of the Table

Group 4 2C9W	[1]. Met 53 (6) Thr 54 (7), Val 55 (8). [2]. Lys 59 (12). [3]. Arg 73 (26), Asp 74 (27), Ser 75 (28), Ser 76 (29), His 77 (30), Ser 78 (31). [4]. Thr 83 (36). [5]. Pro 92 (45), Thr 93 (46), Asn 94 (47), Leu 95 (48), Arg 96 (49), Ile 97 (50), Glu 98 (51). [6]. Leu 106 (59), Asp 107 (60), Ser 108 (61), Ile 109 (62), Ile 110 (63). [7]. Tyr 129 (82). [8]. His 149 (92), Leu 150 (93). [9]. Leu 116 (69)	0.45	0.32	1.82	1.99	1355	7161	0.49/ 0.60	0.37/ 0.51
Group 5 2DMO	[1]. Ile 23 (6), Thr 24 (7), Arg 25 (8). [2]. Glu 29 (12). [3]. Arg 44 (27), Asp 45 (28), Ser 46 (29), Arg 47 (30), His 48 (31), Leu 49 (32). [4]. Thr 53 (39). [5]. Ile 68 (54), Lys 69 (55), His 70 (56), Tyr 71 (57), Gln 72 (58), Lys 73 (59), Ile 74 (60). [6]. Val 83 (71), Ala 84 (72), Glu 85 (73), Arg 86 (74), His 87 (80). [7]. His 79 (92). [8]. Ala 104 (97), Glu 105 (98), Lleu 106 (99)	0.39	0.39	1.81	2.00	1645	6499	0.53/ 0.70	0.50/ 0.68
Group 6 1WQU	[1]. Ile 23 (6). Pro 24 (7), Arg 25 (8). [2]. Ala 29 (12). [3]. ARG 41 (32), GLU 42 (33), SER 43 (34), GLN 44 (35), GLY 45 (36), -//- (37). [4]. Val 50 (42). [5]. Pro 59 (53), Arg 60 (54), His 61 (55), Phe 62 (56), Ile 63 (57), Ile 64 (58), Gln 65 (59). [6]. Leu 73 (68), Glu 74 (69), Gly 75 (72). [7]. Leu 88 (85). [8]. Gln 93 (91), Pro 94 (92), Leu 95 (93), Thr 96 (94), Lys 98 (96), Ser 99 (97). [9]. Glu 48 (40).	0.45	0.45	1.27	1.45	1639	5429	0.53/ 0.59	0.51/ 0.61

Tyrosine-protein kinase transforming protein src (Rous sarcoma virus – 1NZL, 2JYQ), Proto-oncogene tyrosine-protein kinase fyn (*Homo sapiens* – 1G83), v-src tyrosine kinase transforming protein (Rous sarcoma virus – 1BKM); group 2 comprises p55 blk protein tyrosine kinase (*Mus musculus* – 1BLJ), p56lck tyrosine kinase (Homo sapiens – 1CWD), Proto-oncogene tyrosine-protein kinase lck (*Homo sapiens* – 1FBZ), Tyrosine-protein kinase hck (*Homo sapiens* – 2C0T), hck SH2 (*Homo sapiens* – 3HCK); group 3 contains DNA(5′D(AP CP AP GP TP TP TP CP CP CP GP TP AP AP AP TP G P C)-3′) (*Homo sapiens* – 1BF5), Signal transducer and activator of transcription 1-alpha/beta (*Homo sapiens* – 1YVL), DNA(5′D(TP GP CP AP TP TP TP CP CP CP GP TP AP AP AP TP CPT)-3′) (*Mus musculus* – 1BG1), Signal transducer and activator of transcription 3 (*Mus musculus* – 3CWG), Signal transducer and activator of transcription 5A (*Mus musculus* – 1Y1U), Stat protein (Dictyostelium discoideum – 1UUS), Signal-transducing adaptor protein 1 (Homo sapiens – 3MAZ); group 4 – Suppressor of cytokine signaling (*Mus musculus* – 2BBU, 2HMH, *Homo sapiens* – 2IZV, 2VIF);

group 5 – SH3 domain of Tyrosine-protein kinase ITK/TSK (Mus musculus – 2K7A), Tyrosine-protein kinase TXK (*Homo sapiens* – 2DM0), Tyrosine protein kinase BTK (*Homo sapiens* – 2GE9), B-cell linker protein (*Mus musculus* – 2EO6), Basic fibroblast growth factor receptor 1 (*Homo sapiens* – 3GQI), Doubly phosphorylated peptide derived from Syk kinase (*Bos taurus* – 2FCI), 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma 2 (*Rattus norvegicus* – 2EOB); group 6 – Proto-oncogene tyrosine-protein kinase FES/FPS (*Homo sapiens* – 1WQU, 3BKB, 3CBL, 3CD3), Proto-oncogene tyrosine-protein kinase FER (*Homo sapiens* – 2KK6), Hematopoietic SH2 domain containing (*Homo sapiens* – 2CS0), Ras GTPase-activating protein 1 (*Homo sapiens* – 2GSB).

Accessible surface area determination has shown that SH2 domain binding site occupies about ¼ of whole domain area. It is not surprising because in all cases the binding site is flat with two main grooves (pTyr-binding part and hydrophobic part) [7] (Fig. 3). Besides, the matching of whole available structures shows similarity increasing of the binding site in comparison with whole domain (Table). However, if take into consideration the group 1, only. The difference between pocket and whole domain parameters are not significant (e.g. sequence similarity increasing is only 1 %).

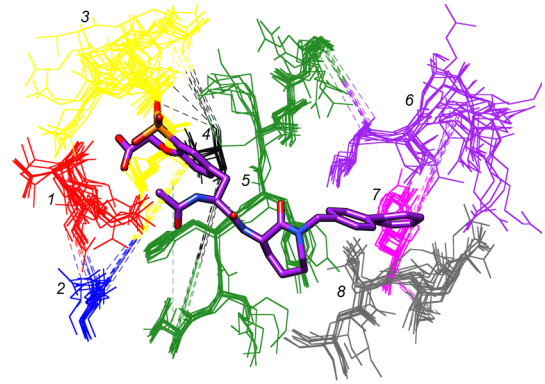

Fig. 3. The binding site of group 1 with the native ligand and the binding site motif distribution: motif *1* (ITR), motif *2* (E), motif *3* (RESETT), motif *4* (S), motif *5* (VKHYKIR), motif *6* (ITSRT), motif *7* (Y) and motif *8* (ADGL). Motives are numbered according to table

Рис. 3. Сайт зв'язування групи 1 із нативним лігандом і розподіл мотивів сайту зв'язування: мотив *1* (ITR), мотив *2* (E), мотив *3* (RESETT), мотив *4* (S), мотив *5* (VKHYKIR), мотив *6* (ITSRT), мотив *7* (Y) і мотив *8* (ADGL). Мотиви пронумеровані відповідно до таблиці

Groups 1 and 2 are highly conservative, especially in ligand binding positions: the increasing of structures identity (if compare with whole SH2 domains) is 61 and 47 % and of similarity is 55.1, 35.2 %, respectively. These groups exhibit the largest conserved surface area, as might be expected from the high level of sequence similarity. But it affects on overall domain surface area and overall surface area of its parts. For example, group 1 has the lowest area of binding pocket, which is highly conserved while less conserved binding pocket of group 2 occupies a bigger surface area (Table). It is achieved by a variable domain amino acid composition.

Groups 3-6 are much less conservative than the groups 1 and 2: the increasing of structures identity (if compare with whole SH2 domains) is 29, 14, 8 and 14 % and of similarity is 14.5, 12.2, 15.5, 11.63%, correspondingly. It leads to significant surface area increasing.

Of course, increasing of similarity and identity of groups 3–6 is not such significant as in groups 1 and 2. But the quality of above parameters is good in the most interesting regions of domains – binding pocket and surrounding areas (Fig. 2, Table) (e. g. similarity increasing in binding pocket in group 3 is 21% and identity increasing is 32 %). In all cases conservative and identity parameters within every separate group is not lower than 40 % (such results are acceptable in bioinformatics research).

Moreover, the conserved regions are located in binding site almost in all cases. However, there are few highly conserved sequences not within binding site, but spatially close to one (For example, the positions 1–5 and 10–11 in all sequences) (Fig. 2). Such conserved parts, which are not involved in a binding site, challenge the notion that SH2 domains recognize only a linear phosphopeptide sequences.

3D structural environment of SH2 domain binding pocket. The eight main separate motifs and single amino acids are present in SH2 domains binding pocket environment. However, the groups 3, 4 and 6 comprise one additional binding point (number 9) (Table).

SH2 domain binding site comprises 20–24 amino acids. It was defined by determining accessible surface area of whole amino acid which is forming SH2 domain binding site and spatially close to one. The mean conservation is much greater in the binding sites than complete domain structure. It confirms that these sites are more conserved than the rest of the domain (Table 1). A lot of domain secondary structures are present in binding site environment (β-B sheet, β-G sheet, β-G sheet and α-B helix) (Figs 1, 3; Table). The pTyr forming parts of pocket are ITR – motif 1, FLVRESETT – motif 3 and C – motif 4. The second half of the motif 3 (RESETT) is involved in the possible ligand interaction, only. Central part of pocket is formed by E – motif 2 and VKHYKIR – motif 5 (β-Dsheet and α-Ahelix). The third part includes ITSRT (motif 6), Y (motif 7), ADGLC (motif 8), a hydrophobic part of pocket (β-G sheet and α-B helix and α-A helix), position 67–70 and 87–91.

Besides, four highly conserved positions (>70 %) through all SH2 domains were selected (position 3 in motif 1, positions 1 and 3 in motif 3, position 3 in motif 5). The other binding site parts are less conservative (<50 %).

However, there are few positions of binding site, which are not involve in any binding motif describe above (e. g., T (position 37 – group 3), L (position 69 – group 4) and Q (position 40 – group 6)). This diversity, located in the known binding interface, may be important in the recognition of ligands.

CONCLUSIONS

Protein members could be studied using the framework created from protein classification data. It is a starting point for examining similarity and diversity inside large protein families. Such kind of studies can indicate functionally important part within the protein structure. It is difficult to make useful conclusion until the protein family is divided to groups.

It is likely that some parts of protein especially binding sites are defined more completely than others. Structure and similarity parameters difference between groups and lack of conservation across whole binding site suggest that while the binding site is conserved in all family, region spatially close to this site can be diverse. However, it is possible to assume that the residue conservation within groups of similar domains outside binding site area might correspond to the conservation of protein-protein interface between SH2 domain and phosphorylated protein. In turn, it casts doubt notion that SH2 domain recognizes short linear peptide motifs only. That assumption based on literature and presented in this article dates. Native ligands mostly are short peptide structures which contact with SH2 domains within pTyr binding site. But, when bound to proteins or long peptide, there are interactions outside pTyr binding site.

1. *Asieh S., Rodziah B. A., KhairinaT. A.* et al. Comparison and Evaluation of Multiple Sequence Alignment Tools in Bininformatics. **Int. J. Computer Sci. Network Security**, 2009; 9: 51–56.
2. *Berman H. M., Westbrook J., Feng Z.* et al. The protein databank. **Nucleic Acids**, 2000; 28: 235–242.
3. *Clamp M., Cuff J., Searle SM.* et al. The Jalview Java alignment editor. **Bioinformatics,** 2004; 20: 426–427.
4. *Erick F.P., Thomas D.G., Conrad C.H.* et al. UCSF Chimera A visualization system for exploratory research and analysis. **J. Comput. Chem,** 2004; 25: 1605–1612.
5. *Goddard T.D., Huang C.C., Ferrin T.E.* Software Extensions to UCSF Chimera for Interactive Visualization of Large Molecular Assemblies. **Structure,** 2005; 13: 473–482.
6. *Guntur F., Joanne L., Scott A.* et al. A potent anti-dengue human antibody preferentially recognizes the conformation of E protein monomers assembled on the virus surface. **EMBO Mol. Med,** 2014; 6: 358–351.
7. *Hurmach V.V., Balinskyi O.M., Platonov M.O.* et al. Design of potentially active ligands for SH2 domains by molecular modeling methods. **Biopolym. Cell,** 2014; 30: 321–325.
8. *Karlin S., Brocchieri L.* Evolutionary conservation of RecA genes in relation to protein structure and function. **J. Bacteriol,** 1996; 178: 1881–1894.
9. *Lee B., Richards M*. The Interpretation of Protein Structures: Estimation of Static Accessibility. **J. Mol. Biol,** 1971; 55: 379–400.
10. *Liu B.A., Jablonowski K., Raina M.* et al. The human and mouse complement of SH2 domain proteinsestablishing the boundaries of phosphotyrosine signaling. **Mol. Cell,** 2006; 22: 851–868.
11. *Needleman S. B.; Christian D. W.* A general method applicable to the search for similarities in the amino acid sequence of two proteins. **J. Mol. Biol,** 1970; 48: 443–53.
12. *Pearl F., Todd J. E., Bray A. C.* et al. Using the CATH domain database to assign structures and functions to the genome sequences. **Biochem. Soc. Trans,** 2000; 28: 269–275.
13. *Pirovano W., Feenstra K.A., Heringa J.* Homology-extended alignment strategy. **Bioinformatics,** 2009; 24: 492–497.
14. *Silva CM.* Role of STATs as downstream signal transducers in Src family kinase-mediated tumorigenesis. **Oncogene,** 2004; 23: 8017–8023.
15. *Thompson J.D., Gibson T.J., Plewniak F.* et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Res,** 1997; 25: 4876–4882.

ПОРІВНЯЛЬНИЙ АНАЛІЗ СТРУКТУР ДОМЕНІВ SH2

***В. В. Гурмач*[1], *М. О. Платонов*[2], *О. М. Бойко*[1], *Ю. І. Прилуцький*[1]**

[1]*Київський національний університет ім. Тараса Шевченка*
вул. Володимирська, 64, Київ 01601, Україна
e-mail: gyrmach@gmail.com

[2]*Інститут молекулярної біології і генетики НАН України*
вул. Академіка Заболотного, 150, Київ 03680, Україна

Src Homology 2 (SH2) – компактні глобулярні домени, які беруть участь у міжклітинній сигналізації та відіграють важливу роль як посередники специфічних протеїн-протеїнових взаємодій. Вони складаються зі 100 амінокислот, які формують сім β-складок і дві α-спіралі. Домени SH2 містять дві висококонсервативні ділянки кишені зв'язування – pTyr і pTyr+3. Знання структури білкових комплексів є важливим кроком у розумінні механізмів їхнього функціонування. З використанням *in silico* методів проаналізовано амінокислотні послідовності сайту зв'язування доменів SH2 і ділянки, прилеглі до нього. Усі домени SH2 були поділені на групи за

подібністю сиквенсів. Аналіз консервативності й ідентичності доменів SH2 виявив ділянки послідовностей, які притаманні всім доменам, і ділянки, притаманні лише деяким доменам. Окрім того, дослідження їхньої поверхні засвідчило, що найбільш консервативні ділянки займають найменшу площу. Одержані результати вказують на здатність доменів SH2 розпізнавати не тільки лінійні фосфопептидні послідовності, що відкриває нові уявлення стосовно інтерпретації можливих механізмів взаємодії доменів SH2 з лігандами (наприклад, можливість зв'язування з протеїнами або лігандами не тільки в межах сайту зв'язування домену).

Ключові слова: домен SH2, сайт зв'язування, консервативність.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СТРУКТУР ДОМЕНОВ SH2

В. В. Гурмач[1]***, О. М. Платонов***[2]***, А. Н. Бойко***[1]***, Ю. И. Прилуцкий***[1]

[1]Киевский национальный университет им. Тараса Шевченко,
ул. Владимирская, 64, Киев 01601, Украина
e-mail: gyrmach@gmail.com

[2]Институт молекулярной биологии и генетики НАН Украины
ул. Академика Заболотного, 150, Киев 01601, Украина

Src Homology 2 (SH2) – компактные глобулярные домены, которые участвуют в межклеточной сигнализации и играют важную роль как посредники специфических белок-белковых взаимодействий. Они состоят из 100 аминокислот, образующих семь β-складок и две α-спирали. Домены SH2 содержат два высококонсервативных участка кармана связывания – pTyr и pTyr+3. Знание структуры белковых комплексов является важным шагом в понимании механизмов их функционирования. С использованием *in silico* методов проанализированы аминокислотные последовательности сайта связывания доменов SH2 и последовательности, прилегающие к нему. Все домены SH2 разбивали на группы по сходству сиквенсов. Анализ консервативности и идентичности доменов SH2 обнаружил участки последовательностей, присущие всем доменам, и участки, присущие только некоторым доменам. Кроме того, исследование их поверхности показало, что наиболее консервативные участки занимают наименьшую площадь. Полученные результаты указывают на способность доменов SH2 распознавать не только линейные фосфопептидные последовательности, что открывает новые представления относительно интерпретации возможных механизмов взаимодействия доменов SH2 с лигандами (например, возможность связывания с белками или лигандами не только в пределах сайта связывания домена).

Ключевые слова: домен SH2, сайт связывания, консервативность.

Одержано: 06.03.2015