

UDC 577.322.23

Design of potentially active ligands for SH2 domains by molecular modeling methods

V. V. Hurmach¹, O. M. Balinskyi¹, M. O. Platonov², O. M. Boyko¹,
P. O. Borysko¹, Yu. I. Prylutskyi¹

¹Taras Shevchenko National University of Kyiv
64, Volodymyrska Str., Kyiv, Ukraine, 01601

²Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03680

gymach@gmail.com

*Search for new chemical structures possessing specific biological activity is a complex problem that needs the use of the latest achievements of molecular modeling technologies. It is well known that SH2 domains play a major role in ontogenesis as intermediaries of specific protein-protein interactions. **Aim.** Developing an algorithm to investigate the properties of SH2 domain binding, search for new potential active compounds for the whole SH2 domains class. **Methods.** In this paper, we utilize a complex of computer modeling methods to create a generic set of potentially active compounds targeting universally at the whole class of SH2 domains. A cluster analysis of all available three-dimensional structures of SH2 domains was performed and general pharmacophore models were formulated. The models were used for virtual screening of collection of drug-like compounds provided by Enamine Ltd. **Results.** The design technique for library of potentially active compounds for SH2 domains class was proposed. **Conclusions.** The original algorithm of SH2 domains research with molecular docking method was developed. Using our algorithm, the active compounds for SH2 domains were found.*

Keywords: molecular modeling, SH2 domains, pharmacophore models, virtual screening.

Introduction. Biological processes in the living system involve a wide variety of protein molecules interacting with each other in stable or dynamic protein complexes. The number and variety of protein interactions are so large and complicated, it makes an extremely complex and intricate network [1]. Hence, knowledge of the spatial structure of cellular protein complexes and their ligands is an important step forward to understanding the mechanisms of cell functioning.

SH2 domains are mostly contained in the cancer-related proteins (*e. g.* Src oncoprotein) and the proteins which are parts of cell signaling pathways. Human genome encodes about 120 SH2 domains which are included in 110 different proteins. They are present in various protein classes including protein kinases (Src, Lck), phos-

phatases (SHP2, SHIP2), phospholipases (PLC γ 1), transcription factors (STAT), regulatory proteins (SOCS), adapter proteins (Grb2), structural proteins (SHC) and others. The widespread presence of SH2 domains in animals and almost complete absence of them in microorganisms (*e. g.* primitive SH2 fragments in yeast) may testify that their appearance is related to complication of the signal transduction mechanisms in multicellular organisms [2]. SH2 domain is a compact globular domain which is actively involved in intracellular signaling pathways. It plays an important role in mediating specific protein-protein interactions [3, 4]. It consists of about 100 amino acids and includes seven β -sheets and two α -helices. Its binding site is characterized by a well-defined placement of Arg β B5 which makes a strong hydrogen bond with two oxygen atoms of phosphotyrosine [5]. Also, the binding pocket is characterized by a large

hydrophobic part, which creates prerequisites for seeking various selective ligands for this class of domains. The SH2 domain is the largest class of pTyr-distinctive domains [6].

Recent research has shown that it is possible to divide SH2 domains according to the specificity of recognition of pTyr residue with C-terminus. Such recognition may occur in position +1, +2, and +3 residues in relation to pTyr [7]. Hence, 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 main goal of this work is to develop an appropriate algorithm to investigate the SH2 domain binding properties and to make a universal library which potentially could be active to the whole SH2 domain class. A generic library of ligands for the whole class of SH2 domains may be created by the molecular modeling methods. Such a library may be used for prediction of the intracellular disturbances, related to the functioning of SH2 domains and its therapeutic utilization.

Materials and methods. *Selection of protein structures.* We have reported the generic SH2 domain clusterization algorithm in detail previously [8]. 1129 three-dimensional SH2 binding site structures retrieved from a Protein Data Bank (PDB) were numerically described based on the geometric shape of their molecular surface. Using cluster analysis of the resulting descriptors, 8 diverse structures were selected for the virtual screening study.

Preparation of small molecule compound collection. A collection of drug-like compounds provided by Enamine Ltd. [9] and containing 1.2 million compounds was used as an original source of the molecular data. Taking into account that docking requires a lot of resources, the Enamine base of chemicals was reduced step by step. Firstly, the Enamine database was reduced to 500 thousand compounds according to the criterion of two-dimensional chemical structures. The selected compounds were used at the second stage for creations of all possible compound conformations (maximum relative energy 50 kJ/mol and minimal RMSD (root-mean-square deviation) of atoms of two conformations 0.4 nm). The

amount of the compounds' conformation was about 3.5 million. Then, 200 thousand compounds were selected from them using the pharmacophore properties diversity represented by piDAPH3 fingerprints (such as charge, donor or acceptors of Hydrogen bond, π -bonds). Finally, using three-dimensional shape of the molecule, 50 thousand compounds were selected.

Virtual screening. Molecular docking (MD) was performed using a flexible ligand and a fixed receptor model. We used an algorithm of systematic docking (SDOCK+) implemented in QXP docking software, which had shown high reproducing ability of ligand conformation with minimum RMSD in comparison to the crystallographic data [10]. The maximum number of SDOCK+ routine steps was set to 100, and the 10 best structures (based on built-in QXP scoring function [11]) were retained for each compound. In accordance with the defined pharmacophore models, the resulting protein-ligand complex structures had been filtered by intrinsic Flo+ filters and multiRMSD software package [12]. Filtering was based on such criteria as the built-in QXP scoring function, the number of hydrogen bonds, the protein-ligand contact surface area and the distance from ligand to key points of the corresponding pharmacophore model.

Results and discussion. *Selection of SH2 domain structures.* Based on the SH2 domain-ligand binding analysis the main points of pharmacophore models were found. So, the main points of those models are the binding site of the pocket (characterized by the large number of hydrogen bond donors (Fig. 1, A, B, C (ARG26, ARG42) and D (ARG524, ARG612)), the center of the pocket (determined by the oxygen position of the second amino acid residue (Fig. 1, A, B, C (His63) and D (His636)) and one or two hydrophobic pockets, which are located behind the pocket center and generally filled with the hydrophobic part of ligand. Then, using the cluster analysis of binding site, the shapes of 8 diverse SH2 domain structures were selected (Table) [8].

Molecular docking. Prior to the docking routine, water molecules were removed from the PDB structures. All Arg and Lys residues were protonated in order to be capable of forming a large number of hydrogen bonds. Regardless of the fact that all binding site clusters were formed within the common framework, particular sets of rules for ligand selection were defined for each cluster,

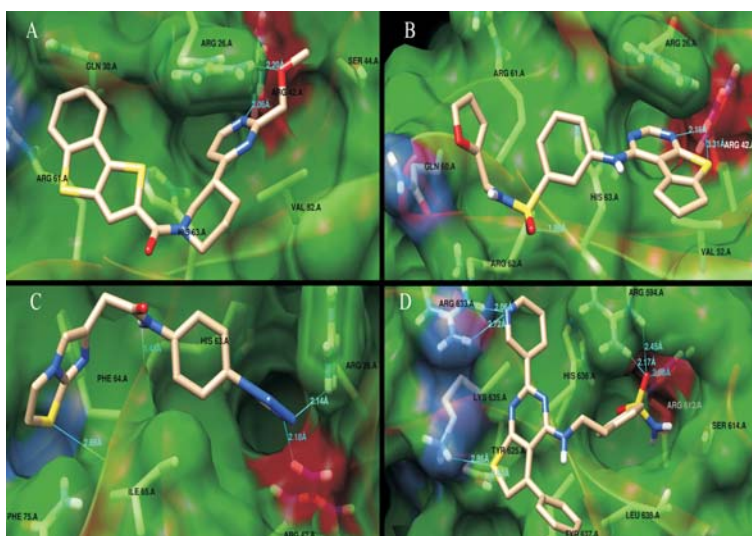


Fig. 1. Models of SH2 binding based on the example of centroids 2KK6 (*A*, *B*, *C*) and 1UUS (*D*) (ribbon and surface diagram). Centroid surface is shown as a translucent green skin, the secondary structural elements (α A (right) and α B (left) helices) of the SH2 domain in red and the binding site in red. Crucial points which deter-

determined by the pTyr-binding site and hydrophobic pocket orientation.

A set of 50000 small molecule structures was docked into the selected SH2 domain binding sites. All possible stereoisomers were generated resulting in about 125000 three-dimensional structures, subsequently docked into the binding sites using QXP/Flo software with 100 steps of SDOCK+ routine. The 10 lowest energy complex structures were selected for each small molecule, resulting in a total of about 10 millions of protein-ligand complexes, which were then filtered by the internal Flo+ filters and the geometry-based filters [12]. In all cases, we had followed the conditions of maximum possible imitation of phosphotyrosine: the ligand had to fill compactly a volume of phosphotyrosine-binding pocket by creating strong hydrogen bonds and cation- π interaction with the main amino acid residue. In all cases it was one or two ARG. So, the general filtering criterion was similar. Exactly this specific binding site was considered to be the key one for the target inhibition. 30000 complexes retained after filtering were the subject for visual analysis.

Chemical groups which can mimic phosphotyrosine and the four main binding models (*A*, *B*, *C*, and *D*) are depicted in Fig. 1. Description of the models follows:

A – the model is characterized by the presence of additional hydrogen bonds involving polar radicals of a phosphotyrosine simulation fragment: hydrogen bonds formed by Arg26 and Arg46. Additionally, strong binding to Arg26 occurs through the cation- π interaction with the pyrimidine heterocycle. The ligand also par-

tially fills the Design of potentially active ligands for SH2 domains by modeling methods the other part of the subpocket and has steric interactions with Gln30, Val52 and His63. In this part, the polycycle is spatially close to Arg61 and the possibility of creation of cation- π interaction exists.

B – the model is characterized by binding the massive heterocycle core, the volume of which exceeds that of the phosphotyrosine fragment. The stabilization occurs through the cation- π interaction with Arg26 and hydrogen bonds with Arg42. Almost whole surface of the tricycle is involved in this process. Formation of the hydrogen bond between His63 and sulfur dioxide is important as well, as this bond causes more rigidity in the protein-ligand binding. Furan fragment adds stabilization to the protein-ligand complex due to Van der Waals force.

C – the model is characterized by partial filling of the phosphotyrosine site. Nitrogen-containing heterocycle creates the hydrogen bonds with Arg42 and Arg26. In this case, the tetrazole fragment, which is a classical acid group electrostatic simulator, plays the role of phospho-tyrosine mimetic. The model also anticipates the hardening of peptide binding due to the formation of another cluster of hydrogen bonds between the NH group and His63, which is located in the other part of the binding pocket but plays an important role in the binding of peptide substrates. Additionally, the formation of the steric interactions with Gln67, Ile65, His63, Phe75 and Phe64 occurs. Owing to the

SH2 domain structures selected for MD

Centroid: structure (PDB code, chain, model)	Organism, protein and domain	Number of structures
1o49, chain A	<i>Homo sapiens</i> SRC SH2	223
2fci, chain A, model 6	<i>Bos taurus</i> PLCG1 SH2-2	48
2ge9, chain A, model 15	<i>Homo sapiens</i> BTK SH2	72
3in7, chain A	<i>Homo sapiens</i> GRB2 SH2	183
2jyq, chain A, model 9	<i>Homo sapiens</i> GRB2 SH2	114
2k7a, chain B, model 5	<i>Mus musculus</i> ITK SH2	218
2kk6, chain A, model 14	<i>Homo sapiens</i> FER SH2	149
1uus, chain A	<i>Dictyostelium discoideum</i> DSTA SH2	129

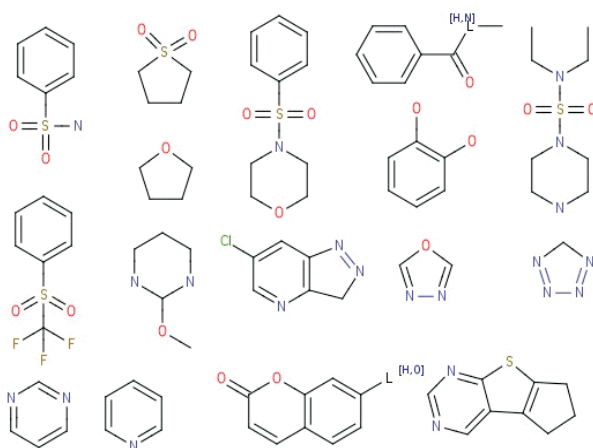


Fig. 2. The fragments, which can imitate phosphotyrosine

fact that ligand has flat spatial conformation, it is almost completely distributed in the pocket and in both parts of the site the binding is not complete. Nevertheless, the compounds which have been selected from this model exhibited good inhibitory effect.

D – the model is represented by a centroid, which is based on PDB record 1UUS. This model has a fundamentally different binding mechanism with the amino acid environment. In this case ligand fills the phosphotyrosine binding site and forms strong hydrogen bonds between Arg594, Arg612 and benzenesulfonamide part of the ligand. The model involves the second cluster of hydrogen bonds between pyridine and Arg633, which is spatially close to the center of the pocket. Moreover, the significant part of the ligand creates stacking and cation- π interaction with Lys635, His636 and Tyr637. Thus, the ligand is firmly fixed in the binding site and shows a significant inhibition effect.

Based on the fact that in all cases the total estimated free energy of complexation is close to the possible minimum value (*A*: -16.0 ; *B*: -16.6 ; *C*: -19.5 and *D*: -23.5 kJ/mol), it is too difficult without biological testing to predict which model is better. Furthermore, in various tests different model may appear to be better even in case if biological tests are used (it depends on the SH2 domain used in the test). Thus, our screening has shown 13 active compounds from *A*, 56 from *C* and 14 from *D* model. Model *B* did not check out at all. It might be connected with the binding complication between massive heterocyclic core and small pTyr binding site of the domain. The resulting library included 3018 structures, each containing 20 to 40 atoms. Computer simulations show that the selected structures form a complex network of chemical interactions in phosphotyrosine binding pocket and reflect all possible modes of the SH2 domain blocking. This confirms a significant chemical diversity of the library and high flexibility of the phosphotyrosine pocket. Moreover, new specific groups able to imitate pTyr have been detected (Fig. 2). Comparison of ADME (absorption, distribution, metabolism and excretion) (LogP, LogS, PSA, Molweight, HbAcc, HbDon, RotationBonds) parameters in the initial and final databases had been carried out to check representativeness of the obtained targeted library. As a result, the difference appeared to be minor. That means the library possesses full chemical diversity of the initial database of compounds within the scope of drug-likeness rules. Nevertheless, a shift of the distribution towards an increase of rotational bonds should be mentioned; this is not surprising because such molecules are able to obtain

a large number of diverse conformations, and thus are more likely to be selected.

Conclusions. This paper presents a few-stage approach to the creation of universal library for the whole SH2 domain class. After docking and analysis of its result, it has been created a general library of the inhibitors mimicking main modes of blocking the binding site of SH2 domain family. Moreover, specific structural fragments (*e. g.* cyclosulfolanic and arylsulfamid groups) have been identified as pTyr substitutes. Almost in all cases the ligand tightly fills the phosphotyrosine binding site and creates hydrogen bonds with the key amino acids Arg and Lys. Besides, there is a steric interaction between ligands and amino acids Glu, Val, Pro, His and Ser.

Дизайн потенційно активних речовин для SH2 доменів методами молекулярного моделювання

В. В. Гурмач, О. М. Балінський, М. О. Платонов, О. М. Бойко, П. О. Бориско, Ю. І. Прилуцький

Резюме

Пошук нових хімічних структур, здатних виявляти специфічну біологічну дію, є комплексною проблемою, що потребує використання сучасних методів молекулярного моделювання. Відомо, що SH2 домени відіграють важливу роль в онтогенезі як посередники специфічних білково-білкових взаємодій. **Мета.** Розробка алгоритму для дослідження властивостей зв'язування SH2 доменів, пошук потенційно активних речовин для всього класу SH2 доменів. **Методи.** Використано комплекс методів комп'ютерного моделювання для створення універсального цільового сету потенційно активних речовин на весь клас SH2 доменів. Проведено кластерний аналіз наявних тривимірних структур SH2 доменів та визначено загальні фармакофорні моделі, які застосовано для віртуального скринінгу колекції лікоподібних речовин підприємства Enatipine. **Результати.** Запропоновано техніку проектування бібліотеки потенційно активних речовин для класу SH2 доменів. **Висновки.** Розроблено оригінальний алгоритм дослідження SH2 доменів із застосуванням методу молекулярного докінгу. Використовуючи цей алгоритм, знайдено активні речовини для SH2 доменів.

Ключові слова: молекулярне моделювання, SH2 домени, фармакофорні моделі, віртуальний скринінг.

Дизайн потенціально активних речовин для SH2 доменів методами молекулярного моделювання

В. В. Гурмач, А. М. Балинский, М. О. Платонов, А. Н. Бойко, П. О. Бориско, Ю. И. Прилуцкий

Резюме

Поиск новых химических структур, способных проявлять специфическое биологическое действие, является комплексной проблемой, требующей использования современных методов молекулярного моделирования. Известно, что SH2 домены играют важную роль в онтогенезе как посредники специфических белково-белковых взаимодействий. **Цель.** Разработка алгоритма для исследо-

вания свойств связывания SH2 доменов, поиск потенциально активных веществ для всего класса SH2 доменов. **Методы.** Применен комплекс методов компьютерного моделирования для создания универсальной целевой выборки потенциально активных веществ на весь класс SH2 доменов. Проведен кластерный анализ имеющихся трехмерных структур SH2 доменов и выделены общие фармакофорные модели, использованные для проведения виртуального скрининга коллекции лекарственных подобных веществ предприятия Enatipine. **Результаты.** Предложена техника проектирования библиотеки потенциально активных веществ для класса SH2 доменов. **Выводы.** Разработан оригинальный алгоритм исследования SH2 доменов с использованием метода молекулярного докинга. Применяя этот алгоритм, найдены активные вещества для SH2 доменов.

Ключевые слова: молекулярное моделирование, SH2 домены, фармакофорные модели, виртуальный скрининг.

REFERENCES

1. Alm E, Arkin AP. Biological networks. *Curr Opin Struct Biol*. 2003;**13**(2):193–202.
2. Silva CM. Role of STATs as downstream signal transducers in Src family kinase-mediated tumorigenesis. *Oncogene*. 2004;**23**(48):8017–23.
3. Koch CA, Anderson D, Moran MF, Ellis C, Pawson T. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science*. 1991;**252**(5006):668–74.
4. Haiyun L, Shamima Banu Bte Sm Rashid, Hao Li, Wee KL, Yih-Cherng L. Knowledge-guided docking of flexible ligands to SH2 domain proteins. IEEE International Conference on Bioinformatics and Bioengineering (Philadelphia, Pennsylvania, USA, May 31–June 3 2010). Philadelphia, 2010:185–190.
5. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc*. 2003;**125**(7):1731–7.
6. Pawson T, Gish GD, Nash P. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol*. 2001;**11**(12):504–11.
7. Liu BA, Jablonowski K, Raina M, Arce M, Pawson T, Nash PD. The human and mouse complement of SH2 domain proteins: establishing the boundaries of phosphotyrosine signaling. *Mol Cell*. 2006;**22**(6):851–68.
8. Balinskyi OM, Sudakov OO, Platonov MO. Small-molecule ligand search for protein interaction domains involving shape-based molecular modeling methods. *Bulletin of Taras Shevchenko National University of Kyiv. Series Physics & Mathematics*. 2012; **3**:287–91.
9. Chuprina A, Lukin O, Demoiseaux R, Buzko A, Shivanyuk A. Drug- and lead-likeness, target class, and molecular diversity analysis of 7.9 million commercially available organic compounds provided by 29 suppliers. *J Chem Inf Model*. 2010;**50**(4):470–9.
10. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS. A critical assessment of docking programs and scoring functions. *J Med Chem*. 2006;**49**(20):5912–31.
11. McMartin C, Bohacek RS. QXP: powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des*. 1997;**11**(4):333–44.
12. Sudakov OO, Balinskyi OM, Platonov MO, Kovalskyy DB. Geometric filters for protein–ligand complexes based on phenomenological molecular models. *Biopolym Cell*. 2013; **29**(5):418–23.

Received 15.05.14