Ukrainica Bioorganica Acta
www.bioorganica.org.ua

ISSN 1814-9758 (Print), ISSN 1814-9766 (Online)
Ukrainica Bioorganica Acta 1 (2016) 24—34

УДК 577.322

The development of algorithm for pharmacophore model optimization and rescoring of pharmacophore screening results

S.A. Starosyla, G.P. Volynets, M.V. Protopopov, V.G. Bdzhola, S.M. Yarmoluk*

Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine
150, Zabolotnoho Str., Kyiv, 03680, Ukraine

Summary. Pharmacophore modeling is modern efficient approach which is widely used in drug discovery. This method has several drawbacks. First of all, pharmacophore models are not selective to structurally similar compounds that belong to one chemical class but possess different activity. Another drawback is that during pharmacophore screening a number of non-active compounds can be selected in the top of results. We have developed novel algorithm for pharmacophore model optimization and for rescoring of pharmacophore screening results. The advantages of our algorithm are optimization of pharmacophore feature radii, application of the concept of pharmacophore feature weights and molecular descriptors. The algorithm was validated based on pharmacophore models of FGFR1 and CK2 inhibitors. In the case of FGFR1 pharmacophore model, the application of algorithm allowed to increase the number of active compounds identified during screening of validation set from 21 to 43 and percent of correctly found active compounds in the top was increased from 38 to 60 %. In the case of CK2 pharmacophore model, the number of active compounds identified during screening of validation set increased from 12 to 39 and percent of correctly found active compounds in the top increased from 58 to 74 %.

Keywords: pharmacophore model, pharmacophore screening, inhibitor, protein kinase FGFR1, protein kinase CK2, optimization algorithm.

Introduction. Today pharmacophore modeling is a powerful tool that is widely used for developing active compounds for different molecular targets [1]. It has several advantages in comparison to molecular docking [2]. First of all, pharmacophore modeling allows design of biologically active compounds for molecular targets with unknown crystal structures. Another advantage is higher speed of virtual screening of large compound libraries [3]. It should be noted that pharmacophore modeling also has some drawbacks. Pharmacophore screening even after validation can give false positive and false negative results and as a result only a few active compounds can be obtained in the top of results [4]. Probable reason for this drawback is pharmacophore simplicity. Pharmacophore model is a set of several pharmacophore features which can not be very selective for similar compounds belonging to one chemical class.

Biological experiments often demonstrate, that small change in compound structure can cause significant decrease in activity, but at the same time pharmacophore model can estimate these compounds equally or even overestimate ligand with worse activity [5]. Such phenomena can be overcome by various algorithms using

*** Corresponding author.**
Tel.: +38044-5260759
E-mail address: yarmolyuksm@gmail.com

different approaches, like application of excluded volumes, volumes of pivot ligands, weights of pharmacophore features, QSAR etc.

We have developed novel algorithm for pharmacophore model optimization and for rescoring of pharmacophore screening results.

The algorithm was written in the Java programming language [6], has a graphical interface and a system of parallelization. This approach takes into account pharmacophore feature weights, pharmacophore feature radii and QSAR based on molecular descriptors. As a screening tool was used Pharmer Program [7]. Obabel module from OpenBabel was used to protonate ligands and to generate molecular descriptors [8].

Method. Earlier, we have developed pharmacophore model with Pharmer and using this hypothesis we have found inhibitors of FGFR1 with the IC_{50} values in the range from 1.4 to 15 μM [9]. The pharmacophore model was built based on three complexes of FGFR1 with inhibitors (PDB ID: 4WUN [10], 3RHX [11], 4F64 [12]). Using this model we have developed and tested our algorithm.

To validate algorithm at different stages of development we have taken known FGFR1 inhibitors from Chembl database [13] — totally 936 compounds, which contained 312 active compounds with $IC_{50} \leq 200$ nM and 624 not active with $IC_{50} > 200$ nM (in our work we used 200 nM as the activity cutoff to divide ligands into active and inactive; this value can be changed accordingly to specific task). We have selected three criteria for determination of pharmacophore model quality: the number of active compounds found in the screening, the number of correctly identified active compounds in the top and the ratio of the number of active compounds to the total number of compounds.

The algorithm is based on the values of pharmacophore features weights. Such approach is used in the programs Accelrys Catalyst [14] and Pharmagist [15]. Pharmacophore feature weights can be calculated based on the set of active compounds to determine how many compounds contribute to one feature. Our initial model was based on only three receptor-ligand complexes. It is not enough and we have extended a set of complexes of FGFR1 with inhibitors ($IC_{50} \leq 100$ nM) to eight (PDB ID: 4WUN, 4F64, 4F65 [12], 2FGI [16], 3TT0 [17], 3WJ6 [18], 4NK9 [19], 4NKS [19]) in order to identify more important features for ligand binding to the receptor. We have made the superposition of complexes using Accelrys Discovery Studio Visualizer 4.0 [14]. The water molecules from the complexes were removed. Then, the inhibitors and their receptors were saved in separate files. Ligands were saved in mol2 format and receptors were saved in pdb format. Initial models for each complex were generated with Pharmer.

Development of module for pharmacophore feature averaging. The next step included the development of mathematical tools and code for module of averaging of pharmacophore features after superposition and for calculation of pharmacophore feature weights.

The principle of work for this module unit is similar to that in the program Pharmagist. If pharmacophore features of one type are at a distance less than certain given threshold, then these features form one cluster. After that, clusters are collected at one point if they are also closer than a certain threshold distance. The process is iterative and the clusters are merged into larger ones after each iteration until the final pharmacophore features will be formed.

The pharmacophore feature weights are calculated as follows:

$$w(f_i) = \sum_i (m(f_i)) \quad (1),$$

where $w(f_i)$ — pharmacophore feature weight, $m(f_i)$ — the number of compounds from superposition which contribute to the feature.

We have proposed the following parameters for averaging: four threshold distances between different pharmacophore features in Å (between aromatic features (1.2), between vector heads donor/acceptor (0.8), between vector tails donor/acceptor/aromatic (0.8), between other features (1.2)); three threshold distances for clustering (between clusters of aromatic features (0.9), vector clusters (0.9), clusters of other features (0.9)). Such parameters by default were set based on our investigation on development of pharmacophore model for ASK1 inhibitors [20]. Other important parameter is the minimum number of pharmacophore features which can be averaged (by default = 2). All these parameters can be modified for specific tasks.

Table 1

Weights and radii of pharmacophore features of initial model

Pharmacophore feature	weight	radius, Å
Don	5	0.5
Acc1	4	0.5
Acc2	7	0.5
Ar1	7	1.1
Ar2	7	1.1
Hyd1	3	1

As a result of the work of module for pharmacophore feature averaging, we have obtained initial pharmacophore model, which comprised two aromatic features, one hydrophobic feature, two hydrogen bond acceptors and one hydrogen bond donor. The weights and radii of pharmacophore features are presented in Table 1.

Based on this model, by excluding some features, we have generated six additional hypotheses (including models from our previous study). Using a set of models instead of one model, allows performing quality parameterization of algorithm. Therefore, all further calculations were carried out taking into account all models.

Different sets of pharmacophore feature radii were used for pharmacophore models generation (Table 2). For all models with different sets of radii was performed primary screening using Pharmer and calculation of parameters for models quality (Table 3). Accordingly to the results, we have narrowed the range of models which were used for further investigations. We have taken only models which identified at least 50 active compounds.

Development of module for pharmacophore feature weight rescoring. Then, we have developed the software unit (module) for weight rescoring. During screening with Pharmer, ligands are represented as a set of pharmacophore features and then RMSD values between compound pharmacophore features and model are calculated.

The module for weight rescoring automatically performs pharmacophore screening using Pharmer, selects compounds conformers with the lowest RMSD values and represents each ligand as a set of pharmacophore features. Then,

Table 2

The sets of pharmacophore features radii which were used during screening

	R		
	D/A	Ar	H
1	0.5	1.1	1
2	0.7	1.3	1.2
3	0.9	1.5	1.4
4	1.1	1.6	1.6
5	1.2	1.6	1.6
6	1.3	1.6	1.6
7	1.4	1.6	1.6

the complexes of ligand pharmacophore features with model are analyzed.

For each pair of matched ligand pharmacophore features and model was calculated partial score, which is product of default score (each type of feature has its own score) and feature weight.

The value of partial score depends on the degree of partial overlap between ligand pharmacophore features and model — the more overlap, the higher score. The sum of partial scores forms final score. Therefore, formula for final score calculation is following:

$$S = \sum_i w(f_i) * S_{def}(f_i) * \frac{(t-d)}{t} \quad (2),$$

where S — the final score, $w(f_i)$ — pharmacophore feature weight, $S_{def}(f_i)$ — default score of pharmacophore features, t — pharmacophore feature radius, d — the distance between matched ligand pharmacophore feature and model.

To parameterize (configure) the module of weight rescoring we have performed a series of screenings against selected models, including pharmacophore hypotheses with excluded volumes, which were generated similarly to usual features on hydrogen atoms of amino acid residues in the active site of protein kinase FGFR1 of the studied crystal complexes within 5Å of ligands. To simplify the model, heavy atoms were not involved.

Additionally, we have used pharmacophore feature weights which were calculated accordingly to formula:

$$w(f_i) = \sum_i \sqrt{m(f_i)} \quad (3).$$

The final results of weight rescoring are

Table 3

The results of the primary screening for different pharmacophore models of FGFR1 inhibitors using different sets of radii (numbering according to Table 2). The models which identified at least 50 active compounds are labeled by yellow color

	ArArHDAA			**ArArHAA**			**ArArHDA**			**ArHDAA**		
	Total number	Number of active compounds	% correctly predicted active compounds	Total number	Number of active compounds	% correctly predicted active compounds	Total number	Number of active compounds	% correctly predicted active compounds	Total number	Number of active compounds	% correctly predicted active compounds
1	1	1	100	9	7	71	2	2	100	1	1	100
2	4	4	100	71	30	73	5	4	100	3	3	100
3	7	6	83	126	51	43	14	12	83	7	5	80
4	12	9	89	166	71	45	29	18	67	11	8	75
5	11	8	88	163	71	45	33	21	62	19	11	64
6	15	11	73	162	70	46	34	21	62	25	13	69
7	18	12	75	199	90	41	36	22	59	39	15	47
	ArHAA			**ArHDA**			**ArHDA**					
	Total number	Number of active compounds	% correctly predicted active compounds	Total number	Number of active compounds	% correctly predicted active compounds	Total number	Number of active compounds	% correctly predicted active compounds			
1	76	29	31	12	7	43	100	42	49			
2	143	57	46	39	21	52	133	55	44			
3	211	85	47	75	35	51	156	69	48			
4	256	107	38	105	50	50	199	85	39			
5	283	120	37	115	53	47	231	89	38			
6	315	128	38	136	58	47	264	98	40			
7	341	135	37	145	63	46	280	105	40			

shown in Table 4. In some cases the percent of correctly found active compounds was increased by 7-8 %. At this step we have chosen default scores for different types of pharmacophore features: aromatic feature — 3.5, hydrophobic feature — 3, hydrogen bond donor/acceptor — 2, ion — 1, aromatic → hydrophobic feature — 0.5, excluded volume — 1.

Development of rescoring module based on molecular descriptors. Also, we have developed rescoring module based on molecular descriptors, which were calculated for ligands. This module calculates coefficient which multiplies the final score.

Descriptors (the number of atoms, the number of bonds of different types, molecular we-

Table 4

The weight rescoring of the results of primary screening with or without molecular descriptors for different pharmacophore models of FGFR1 inhibitors (numbering models accordingly to the sets of radii of pharmacophore features are given in Table 2). The percentages of correctly found active compounds from validation set are represented. The best results of weight rescoring are labeled by light green color, the best results of weight rescoring using molecular descriptors are shown by dark green color

Model	Weight rescoring		Weight rescoring with descriptors	
	$\sqrt{w}$	w	$\sqrt{w}$	w
ArArHAA-3	49	49	**53**	**53**
ArArHAA-4	46	44	45	48
ArArHAA-5	**51**	49	48	49
ArArHAA-6	49	49	49	50
ArArHAA-7	48	49	50	50
ArHAA-2	**49**	**49**	51	51
ArHAA-3	38	39	51	49
ArHAA-4	44	42	52	51
ArHAA-5	44	43	53	51
ArHAA-6	42	39	**54**	50
ArHAA-7	41	37	**54**	50
ArHDA-4	**58**	**58**	**60**	58
ArHDA-5	53	53	55	53
ArHDA-6	50	50	50	52
ArHDA-7	49	52	51	51
ArHDA-2	**45**	**45**	55	55
ArHDA-3	39	39	57	54
ArHDA-4	38	38	51	**58**
ArHDA-5	39	39	52	54
ArHDA-6	40	41	45	47
ArHDA-7	39	43	50	50

ight, the number of hydrogen bond donors and acceptors, logP, the number of halogens, topological polar surface area (TPSA), molecular refraction) were calculated automatically by specific module which preprocesses validation set or ligand database for screening.

For further work, algorithm does not use descriptors for which the ratio between the mean values of the descriptor of active and inactive ligands from validation set was more than 0.9.

Independently, for groups of active and inactive ligands we have determined ratios of the average deviations to the average values of each descriptor. If this ratio for descriptor is bigger for active ligands, their parameters, such as average deviation, average value and positive adjustment coefficient will be used. When this ratio is bigger for inactive ligands then the parameters with negative adjustment coefficient will be taken.

This coefficient is calculated accordingly to formula:

$$k(d) = \pm\frac{1 - dev(d)/aver(d)}{\sum_i 1 - \frac{dev(i)}{aver(i)}} \qquad (4),$$

where *k(d)* — adjustment coefficient of descriptor, *dev(d)* — average deviation, *aver(d)* — average value, *dev(i)* and *aver(i)* — average deviation and average value of ith descriptor.

Descriptor parameters such as average deviation, average value and adjustment coefficient were added to table of parameters.

The scheme for calculating the coefficient, which will multiply the weight score is following:

1. *K=1*;

2. $descr(i) > aver(i) - dev(i) \wedge descr(i) < aver(i) + dev(i) \Rightarrow K := K + k(i) * 3$ (5);

3. $descr(i) \leq aver(i) - dev(i) \vee descr(i) \geq aver(i) + dev(i) \Rightarrow K := K - k(i)$ (6), where K is a coefficient, which will multiply the weight score, *descr(i)* — the value of i^{th} descriptor of studied ligand, *aver(i)* — the mean value of descriptor, which was taken from the table of parameters, *k(i)* — adjustment coefficient of descriptor, which was taken from the table of parameters.

The results of rescoring using molecular descriptors are shown in Table 4.

Development of modules for optimization of pharmacophore feature radii, weights and default scores. From the results presented in Table 3 and 4 is clearly that the radii of pharmacophore features and their weights greatly affect the quality parameters of models without any relationship. Best-quality model can be determined by checking the whole set of models with individual radii for each pharmacophore feature and with different weights. Therefore, we have developed optimizer of pharmacophore feature radii and optimizer of pharmacophore feature weights and default score for these features.

In radii optimizer based on the radii of primary pharmacophore model was created matrix $m \ \ n$, where m — the number of generated models, n — radii of pharmacophore features. To generate the matrix the number of steps of change for each radius and the value range which varies per one step were set up.

The number of generated models is $m=f^n$ (7), where f — the number of steps, n — the number of radii. For example, the matrix generated for pharmacophore model consisting of three features, with the number of steps — 2 and the change Δ, will be as follows:

	r1	r2	r3
m1	$r1_0$	$r2_0$	$r3_0$
m2	$r1_0$	$r2_0$	$r3_0 + \Delta$
m3	$r1_0$	$r2_0 + \Delta$	$r3_0$
m4	$r1_0$	$r2_0 + \Delta$	$r3_0 + \Delta$
m5	$r1_0 + \Delta$	$r2_0$	$r3_0$
m6	$r1_0 + \Delta$	$r2_0$	$r3_0 + \Delta$
m7	$r1_0 + \Delta$	$r2_0 + \Delta$	$r3_0$
m8	$r1_0 + \Delta$	$r2_0 + \Delta$	$r3_0 + \Delta$

Then, the models based on the generated sets of radii of pharmacophore features were generated and pharmacophore screening of validation compound set against these models was performed and quality parameters for their evaluation were proposed. Based on these parameters total score was estimated accordingly to the following formula:

$$q = \frac{a}{a_{max}} + \frac{a/all}{(a/all)_{max}} + \frac{a_r/a}{(a_r/a)_{max}} \quad (8),$$

where a — the number of active compounds from validation set, which were identified during screening against selected model, a_{max} — the maximum number of active compounds from the validation set found during screening of all generated models, *all* — the number of all (active and inactive) compounds from validation set, identified during screening against selected model, $(a/all)_{max}$ — maximum ratio of active compounds from validation set to the total number of compounds identified during screening among all generated models, a_r — the number of correctly found active compounds from validation set against selected model, $(a_r/a)_{max}$ — maximum ratio of correctly identified active compounds from validation set to the total number of compounds identified during screening among all generated models. Accordingly to the score the models were sorted and the best-scored one was selected.

The optimizer of pharmacophore feature weights and their default scores works similarly to radii optimizer, but in this case the matrix of default scores for pharmacophore features was generated. These default score sets were used for the best models with optimized radii of pharmacophore features (the number of models can be selected manually). In addition, different pharmacophore feature weights were used accordingly to modified formula (3):

$$w(f_i) = \sum_i \sqrt[n]{m(f_i)} \quad (9),$$

where n — is value from 1 to 6.

The screening was performed against the models with all possible combinations of pharmacophore feature weights and default scores. As a result, the model which found the most active compounds from validation set in the top of weight rescoring was determined.

The general scheme of the algorithm is pre-

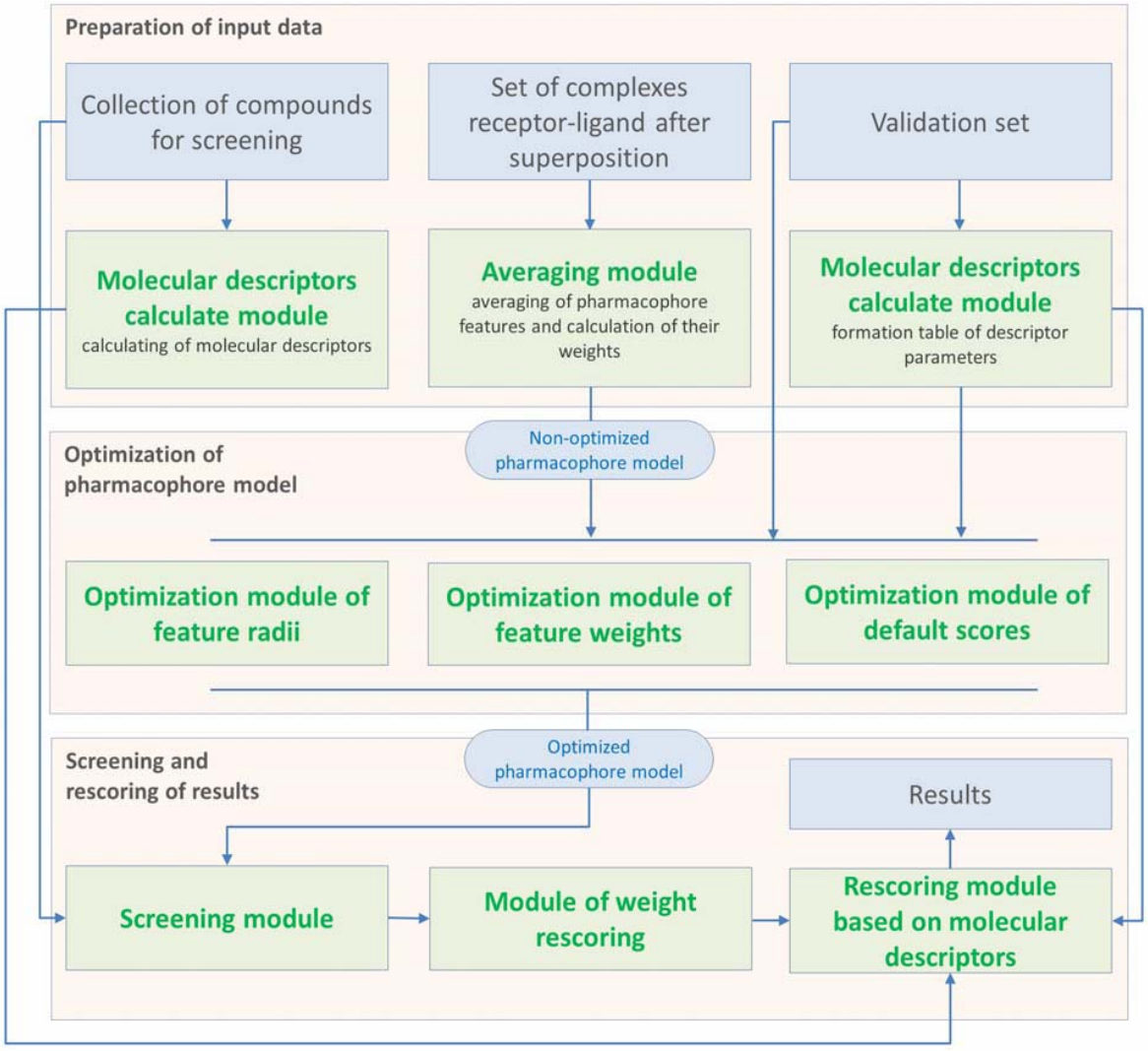

Figure 1. The general scheme of the algorithm.

sented in Figure 1. The algorithm consists of three main stages. Preparation of input data includes development of conformers for validation set and database for screening, converting these sets and generating of molecular descriptors for them, formation of the table with descriptor parameters, and generation of the non-optimized pharmacophore model. Optimization of pharmacophore model includes optimization of pharmacophore features radii, weight and default scores. Screening and rescoring of results at the third stage is performed by three respective modules.

Validation of the algorithm. As a control experiment we have performed full optimization cycles for our pharmacophore models of FGFR1 and CK2 inhibitors, screening of validation set against these models and rescoring of obtained results.

Validation of the algorithm based on FGFR1 pharmacophore model. The validation set for FGFR1, which included 936 compounds, was divided randomly into two subsets. The amount of active compounds in both subsets was equal. One subset was used as training set, another one as test set. Based on the training set we have performed pharmacophore model optimization. Then, we carried out screening of test set against optimized and non-optimized pharmacophore hypotheses.

Original, non-optimized pharmacophore model of FGFR1 inhibitors identified 50 compounds from validation set, among them were 21 active compounds and percent of correctly found active compounds was 38 %. Optimized pharmacophore hypothesis identified 98 compounds, among them were 43 active compounds and percent of correctly found active compounds was

60 %. Optimized pharmacophore feature weights and radii are presented in Table 5. Optimized default scores for pharmacophore features are following: for aromatic features — 3.7, for hydrophobic — 3.0, for hydrogen bond donor/acceptor — 2.0, aromatic → hydrophobic feature — 0.7.

Superposition of the complexes of FGFR1 with inhibitors, non-optimized pharmacophore model of FGFR1 inhibitors and optimized model are shown in Figure 2.

Based on the results of both screenings we have calculated Mathew's correlation coefficients. Mathew's coefficient for non-optimized pharmacophore model was 0.47, and for optimized — 0.55. This indicates that the accuracy of pharmacophore model after optimization was increased.

Also, we have calculated the values of Z-scores. The screening results are presented in different units: for non-optimized model in RMSD, for optimized — in its own algorithm score. In order to compare data, Z-scores were presented in percents. In case of RMSD, minimal value was taken as 100 %, other values were taken as percentages from it. In case of algorithm score, maximal value was taken as 100 %, other values were taken as percentages from it.

Table 5

Optimized pharmacophore feature weights and radii of FGFR1 inhibitors pharmacophore model

Pharmacophore model	Weight	Radius, Å
Don	2.2361	0.9
Acc	2.6458	0.9
Ar	2.6458	1.7
Hyd	1.7321	1.5

Therefore, during screening against non-optimized pharmacophore model, Z-score values for active compounds were in the range from 100 to -3.31 and mean value was 1.21, for non-active compounds Z-score values were in the range from 65.4 to -3.31 and mean value was -0.61.

During screening against optimized pharmacophore model, Z-score values for active compounds were in the range from 100 to -8.49 and

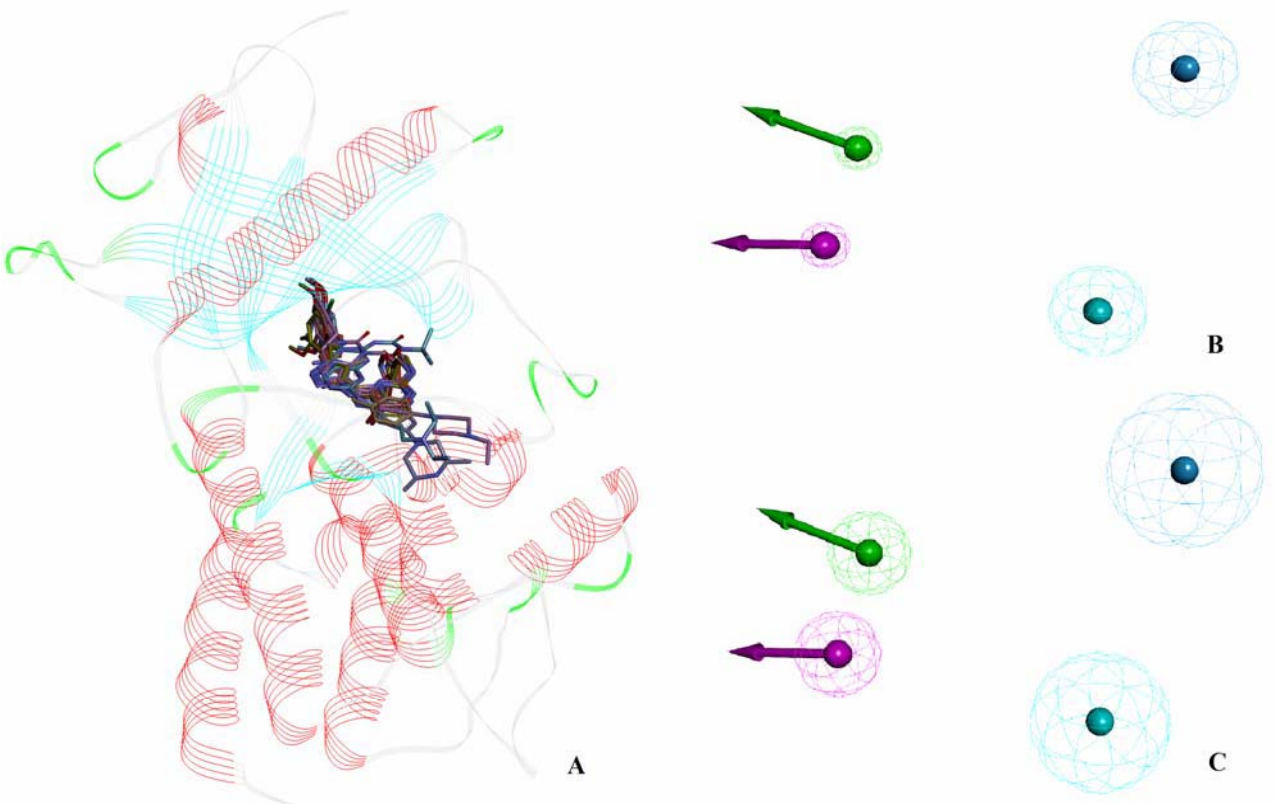

Figure 2. Superposition of the complexes of FGFR1 with inhibitors (A), non-optimized pharmacophore model of FGFR1 inhibitors (B), and optimized model (C). Hydrogen bond donors are indicated by purple arrows, hydrogen bond acceptors — green arrows, aromatic features — blue poins, hydrophobic features — cyan points.

Table 6

Optimized and non-optimized weights and radii of pharmacophore features of the pharmacophore model of CK2 inhibitors

Feature	Weight	Radius (non-optimized), Å	Radius (optimized), Å
Don	11	0.5	1
Acc	13	0.5	0.5
Ar	14	1.1	1.6
Hyd	13	1	1

mean value was 4.43, for non-active compounds Z-score values were in the range from 83.44 to -8.49 and mean value was -2.22.

These results indicate that score values of active compounds during screening against optimized pharmacophore hypothesis have more positive deviation from mean value and scores of non-active compounds have more negative deviation from mean value therefore the optimized model has better predictive ability.

Validation of the algorithm based on CK2 pharmacophore model. To form the validation set of protein kinase CK2 inhibitors we have taken known CK2 inhibitors from Chembl database — 1091 compounds, which included 194 active compounds (IC_{50}≤200 nM) and 897 not active (IC_{50}>200 nM). The validation set of protein kinase CK2 was divided into two subsets. One subset was used as training set, another one as test set.

The original model of CK2 inhibitors contained 6 pharmacophore features [21]. To calculate pharmacophore feature weights we have taken 14 complexes of protein kinase CK2 with inhibitors (IC_{50}≤100 nM) (PDB ID: 3AT4 [22], 3E3B [23], 3MB6 [24], 3MB7 [24], 3PE1 [25], 3PE2 [25], 3R0T [25], 3U4U [26], 4GRB [27], 4KWP [28], 4UB7 [29], 2ZJW [30], 3AT3 [22], 4ANM [31]). The weights and radii are shown in Table 6. The screening of validation set against this model didn't give positive results, therefore we have deleted two hydrogen bond donors located in hydrophobic pocket II, because these pharmacophore features possessed the lowest weights. The new screening allowed us to find 22 compounds from validation set, among them were 12 active compounds and percent of correctly found active compounds was 58 %. After optimization, the model identified 59 compounds, among them were 39 active compounds and percent of correctly found active compounds was 74 %. During optimization the weights of pharmacophore features were not changed. Optimized radii of pharmacophore model are represented in Table 6. Optimized default scores for pharmacophore features are following: for aromatic features — 2.9, for hydrophobic features — 2.4, for hydrogen bond donors/acceptors — 1.4, aromatic → hydrophobic feature — 1.1.

Superposition of the complexes of CK2 with inhibitors, non-optimized pharmacophore model of CK2 inhibitors and optimized model are shown in Figure 3.

Based on the results of both screenings we have calculated Mathew's correlation coefficients. Mathew's coefficient for non-optimized pharmacophore model was 0.57, and for optimized — 0.72. This indicates that the accuracy of pharmacophore model after optimization was significantly increased.

Also, we have calculated the values of Z-scores. During screening against non-optimized CK2 pharmacophore model, Z-score values for active compounds were in the range from 100 to -1.83 and mean value was 4.34, for non-active compounds Z-score values were in the range from 77.06 to -1.83 and mean value was -0.94.

During screening against optimized pharmacophore model, Z-score values for active compounds were in the range from 65.36 to -3.57 and mean value was 9.85, for non-active compounds Z-score values were in the range from 100 to -3.57 and mean value was -2.13.

These results indicate that score values of active compounds during screening against optimized pharmacophore hypothesis have more positive deviation from mean value and scores of non-active compounds have more negative deviation from mean value therefore the optimized model has better predictive ability.

Надійшла в редакцію 14.04.2016 р.

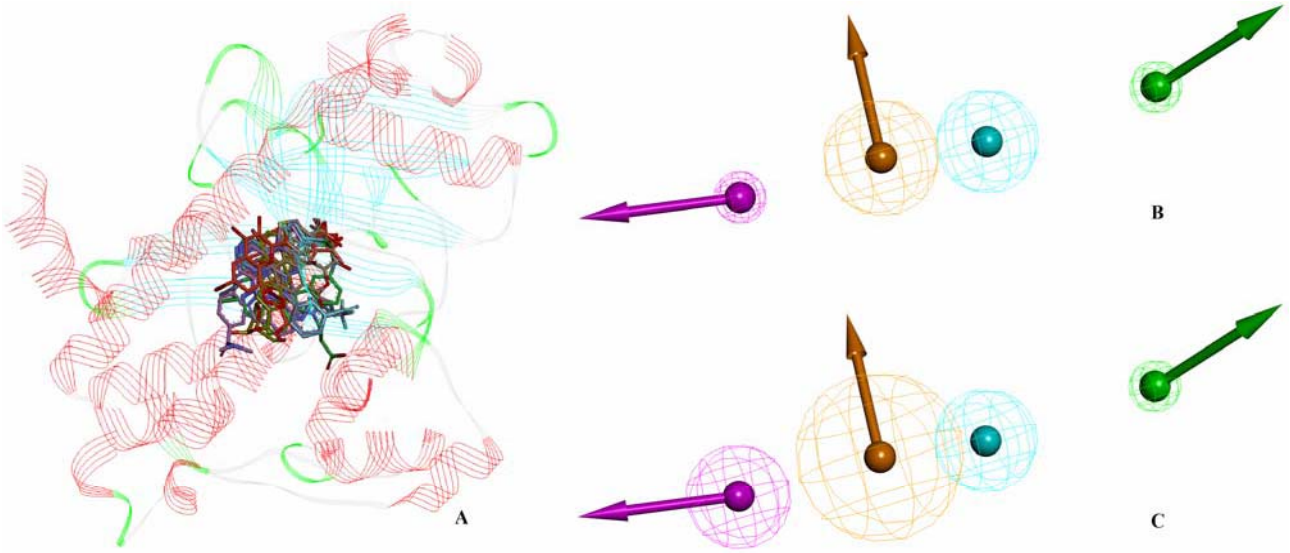

Figure 3. Superposition of the complexes of CK2 with inhibitors (A), non-optimized pharmacophore model of CK2 inhibitors (B), and optimized model (C). Hydrogen bond donors are indicated by purple arrows, hydrogen bond acceptors — green arrows, aromatic features — orange arrows, hydrophobic features — cyan points.

Розробка алгоритму оптимізації фармакофорних моделей і рескорингу результатів фармакофорного скринінгу

С.А. Старосила, Г.П. Волинець, М.В. Протопопов, В.Г. Бджола, С.М. Ярмолюк

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

Резюме. Фармакофорне моделювання — сучасний ефективний підхід, що широко використовується в розробці лікарських препаратів. Цей метод має такі недоліки: 1) фармакофорні моделі часто є неселективними для структурно подібних сполук, які належать до одного хімічного класу, але мають різну активність; 2) під час фармакофорного скринінгу багато неактивних сполук потрапляють у топ результатів. Нами розроблено новий алгоритм оптимізації фармакофорних моделей та рескорингу результатів фармакофорного скринінгу. Перевагою цього алгоритму є поєднання оптимізації радіусів фармакофорних точок моделі і використання концепції вагів фармокофорних точок та молекулярних дескрипторів. Алгоритм перевірено на фармакофорних моделях інгібіторів FGFR1 і CK2. Оптимізація моделі інгібіторів FGFR1 за допомогою алгоритму дала змогу збільшити кількість знаходження активних сполук при скринінгу валідаційної вибірки з 21 до 43 і покращити відсоток правильно знайдених активних сполук із 38 до 60 %. Оптимізація моделі інгібіторів CK2 дала змогу збільшити кількість знаходження активних сполук при скринінгу валідаційної вибірки з 12 до 39 і покращити відсоток правильно знайдених активних сполук з 58 до 74 %.

Ключові слова: фармакофорна модель, фармакофорний скринінг, інгібітор, протеїнкіназа FGFR1, протеїнкіназа CK2, алгоритм оптимізації.

References

1. *Starosyla S.A., Volynets G.P., Bdzhola V.G., Golub A.G., Yarmoluk S.M.* Pharmacophore approaches in protein kinase inhibitors design // World J Pharmacol. — 2014. — 3(4). — P. 162-173.

2. *Chen Z., Li H.L., Zhang Q.J., Bao X.G., Yu K.Q., Luo X.M., Zhu W.L., Jiang H.L.* Pharmacophore-based virtual screening versus docking-based virtual screening: a benchmark comparison against eight targets // Acta Pharmacol Sin. — 2009. — 30 (12). — P. 1694-708.

3. *Koes D.R.* Pharmacophore modeling: methods and applications // Methods in Pharmacology and Toxicology. — 2015. — DOI 10.1007/7653_2015_46

4. *Yang S.Y.* Pharmacophore modeling and applications in drug discovery: challenges and recent advances // Drug Discov Today. — 2010. — 15 (11-12). — P. 444-50.

5. *Langer T., Hoffmann R.D.* Pharmacophores and pharmacophore searches. Wiley-VCH; [Chichester: John Wiley, distributor]: Weinheim, 2006.

6. https://java.com/en/

7. *Koes D.R., Camacho C.J.* Pharmer: efficient and exact pharmacophore search // J Chem Inf Model. — 2011. — 51 (6). — P. 1307-14.

8. *O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R.* Open Babel: An open chemical toolbox // J Cheminform. — 2011. — 3, 33. doi:10.1186/1758-2946-3-33

9. *Starosyla S.A., Volynets G.P., Protopopov M.V., Bdzhola V.G., Dzyadevych S.V., Borovykov O.V., Pletnyova L.V., Yarmoluk S.M.* Discovery of novel protein kinase FGFR1 inhibitors using pharmacophore mode-

ling // Ukrainica Bioorganica Acta. — 2015. — 13 (1). — P. 13-20.

10. *Yosaatmadja Y., Patterson A.V., Smaill J.B., Squire C.J.* The 1.65 Å resolution structure of the complex of AZD4547 with the kinase domain of FGFR1 displays exquisite molecular recognition // Acta Crystallogr D Biol Crystallogr. — 2015. — 71 (Pt 3). — P. 525-33.

11. *Eathiraj S., Palma R., Hirschi M., Volckova E., Nakuci E., Castro J., Chen C.R., Chan T.C., France D.S., Ashwell M.A.* A novel mode of protein kinase inhibition exploiting hydrophobic motifs of autoinhibited kinases: discovery of ATP-independent inhibitors of fibroblast growth factor receptor // J Biol Chem. — 2011. — 286 (23). — P. 20677-87.

12. *Norman R.A., Schott A.K., Andrews D.M., Breed J., Foote K.M., Garner A.P., Ogg D., Orme J.P., Pink J.H., Roberts K., Rudge D.A., Thomas A.P., Leach A.G.* Protein-ligand crystal structures can guide the design of selective inhibitors of the FGFR tyrosine kinase // J Med Chem. — 2012. — 55 (11). — P. 5003-12.

13. https://www.ebi.ac.uk/chembl/

14. http://accelrys.com/

15. *Dror O., Schneidman-Duhovny D., Inbar Y., Nussinov R., Wolfson H.J.* Novel approach for efficient pharmacophore-based virtual screening: method and applications // J Chem Inf Model. — 2009. — 49 (10). — P. 2333-43.

16. *Mohammadi M., Froum S., Hamby J.M., Schroeder M.C., Panek R.L., Lu G.H., Eliseenkova A.V., Green D., Schlessinger J., Hubbard S.R.* Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain // EMBO J. — 1998. — 17 (20). — P. 5896-904.

17. *Guagnano V., Furet P., Spanka C., Bordas V., Le Douget M., Stamm C., Brueggen J., Jensen M.R., Schnell C., Schmid H., Wartmann M., Berghausen, J., Drueckes P., Zimmerlin A., Bussiere D., Murray J., Graus Porta D.* Discovery of 3-(2,6-dichloro-3,5-dimethoxy-phenyl)-1-{6-[4-(4-ethyl-piperazin-1-yl)-phenylamino]-pyrimidin-4-yl}-1-methyl-urea (NVP-BGJ398), a potent and selective inhibitor of the fibroblast growth factor receptor family of receptor tyrosine kinase // J Med Chem. — 2011. — 54 (20). — P. 7066-83.

18. *Nakanishi Y., Akiyama N., Tsukaguchi T., Fujii T., Sakata K., Sase H., Isobe T., Morikami K., Shindoh H., Mio T., Ebiike H., Taka N., Aoki Y., Ishii N.* The fibroblast growth factor receptor genetic status as a potential predictor of the sensitivity to CH5183284/Debio 1347, a novel selective FGFR inhibitor // Mol Cancer Ther. — 2014. — 13 (11). — P. 2547-58.

19. *Klein T., Tucker J., Holdgate G.A., Norman R.A., Breeze A.L.* FGFR1 kinase inhibitors: close regioisomers adopt divergent binding modes and display distinct biophysical signatures // ACS Med Chem Lett. — 2014. — 5 (2). — P. 166-71.

20. *Starosyla S.A., Volynets G.P., Bdzhola V.G., Golub A.G., Protopopov M.V., Yarmoluk S.M.* ASK1 pharmacophore model derived from diverse classes of inhibitors // Bioorg Med Chem Lett. — 2014. — 24 (18). — P. 4418-23.

21. *Starosyla S.A., Volynets G.P., Protopopov M.V., Bdzhola V.G., Dzyadevych S.V., Borovykov O.V., Matyushok V.I., Yarmoluk S.M.* Discovery of novel protein kinase CK2 inhibitors using pharmacophore modeling // Ukrainica Bioorganica Acta. — 2014. — 12 (2). — P. 23-34.

22. *Kinoshita T., Sekiguchi Y., Fukada H., Nakaniwa T., Tada T., Nakamura S., Kitaura K., Ohno H., Suzuki Y., Hirasawa A., Nakanishi I., Tsujimoto G.* A detailed thermodynamic profile of cyclopentyl and isopropyl derivatives binding to CK2 kinase // Mol Cell Biochem. — 2011. — 356 (1-2). — P. 97-105.

23. *Nakaniwa T., Kinoshita T., Sekiguchi Y., Tada T., Nakanishi I., Kitaura K., Suzuki Y., Ohno H., Hirasawa A., Tsujimoto G.* Structure of human protein kinase CK2 alpha 2 with a potent indazole-derivative inhibitor // Acta Crystallogr Sect F Struct Biol Cryst Commun. — 2009. — 65 (Pt 2). — P. 75-9.

24. *Lopez-Ramos M., Prudent R., Moucadel V., Sautel C.F., Barette C., Lafanechere L., Mouawad L., Grierson D., Schmidt F., Florent J.C., Filippakopoulos P., Bullock, A.N., Knapp S., Reiser J.B., Cochet C.* New potent dual inhibitors of CK2 and Pim kinases: discovery and structural insights // FASEB J. — 2010. — 24 (9). — P. 3171-85.

25. *Battistutta R., Cozza G., Pierre F., Papinutto E., Lolli G., Sarno S., O'Brien S.E., Siddiqui-Jain A., Haddach M., Anderes K., Ryckman D.M., Meggio F., Pinna L.A.* Unprecedented selectivity and structural determinants of a new class of protein kinase CK2 inhibitors in clinical trials for the treatment of cancer // Biochemistry. — 2011. — 50 (39). — P. 8478-88.

26. *Dowling J.E., Chuaqui C., Pontz T.W., Lyne P.D., Larsen N.A., Block M.H., Chen H., Su N., Wu A., Russell D., Pollard H., Lee J.W., Peng B., Thakur K., Ye Q., Zhang T., Brassil P., Racicot V., Bao L., Denz C.R., Cooke E.* Potent and selective inhibitors of CK2 kinase identified through structure-guided hybridization // ACS Med Chem Lett. — 2012. — 3 (4). — P. 278-83.

27. *Dowling J.E., Alimzhanov M., Bao L., Block M.H., Chuaqui C., Cooke E.L., Denz C.R., Hird A., Huang S., Larsen N.A., Peng B., Pontz T.W., Rivard-Costa C., Saeh J.C., Thakur K., Ye Q., Zhang T., Lyne P.D.* Structure and property based design of pyrazolo[1,5-a]pyrimidine Inhibitors of CK2 kinase with activity *in vivo* // ACS Med Chem Lett. — 2013. — 4 (8). — P. 800-5.

28. *Cozza G., Girardi C., Ranchio A., Lolli G., Sarno S., Orzeszko A., Kazimierczuk Z., Battistutta R., Ruzzene M., Pinna L.A.* Cell-permeable dual inhibitors of protein kinases CK2 and PIM-1: structural features and pharmacological potential // Cell Mol Life Sci. — 2014. — 71 (16). — P. 3173-85.

29. *Guerra B., Bischoff N., Bdzhola V.G., Yarmoluk S.M., Issinger, O.G., Golub A.G., Niefind K.* A note of caution on the role of halogen bonds for protein kinase/inhibitor recognition suggested by high- and low-salt CK2α complex structures // ACS Chem Biol. — 2015. — 10 (7). — P. 1654-60.

30. *Sekiguchi Y., Nakaniwa T., Kinoshita T., Nakanishi I., Kitaura K., Hirasawa A., Tsujimoto G., Tada T.* Structural insight into human CK2alpha in complex with the potent inhibitor ellagic acid // Bioorg Med Chem Lett. — 2009. — 19 (11). — P. 2920-3.

31. *Koltun E.S., Tsuhako A.L., Brown D.S., Aay N., Arcalas A., Chan V., Du H., Engst S., Ferguson K., Franzini M., Galan A., Holst C.R., Huang P., Kane B., Kim M.H., Li J., Markby D., Mohan M., Noson K., Plonowski A., Richards S., Robertson S., Shaw K., Stott G., Stout T.J., Young J., Yu P., Zaharia C.A., Zhang W., Zhou P., Nuss J.M., Xu W., Kearney P.C.* Discovery of XL413, a potent and selective CDC7 inhibitor // Bioorg Med Chem Lett. — 2012. — 22 (11). — P. 3727-31.