Russian Chemical Reviews **81** (2) 158 – 174 (2012) © 2012 Russian Academy of Sciences and Turpion Ltd

DOI 10.1070/RC2012v081n02ABEH004222

Fragment-based lead design

O A Filz, V V Poroikov

Contents

I.	Introduction	158
II.	Fragment library design	159
III.	Selection of fragments based on the estimation of their contribution to biological activity	162
IV.	Methods of fragment-based generation of molecular structures	167
V.	Prospects of the development of methods for the fragment-based lead design	171

Abstract. State-of-the-art approaches to the fragment-based design of organic compounds with desired properties are considered. The review covers methods, which are used in different steps of the design, such as computational methods for fragment library design, experimental and computational methods for fragment discovery and methods for the generation of structures of organic compounds. Examples are given of drug candidates, which were constructed using the fragment-based approach. The bibliography includes 156 references.

I. Introduction

One of the main goals of medicinal chemistry is to discover novel high-affinity ligands, *viz*., organic compounds capable of binding to pharmacological targets, which can be used for drug design. Biological screening is widely used to assess interactions between organic compounds and various proteins responsible for the pharmacotherapeutic effect. The total screening involves the examination of interactions between all possible ligands and all known pharmacological targets. However, the chemical space (all available and theoretically possible structures) of drug-like compounds includes up to 10^{180} structures.[1] According to Fink and Reymond,[2] the total number of possible unique structures with a molecular weight < 500 Da, which are of interest for researchers in the field of drug design, is at least 10^{60} (Ref. 3). Evidently, it is impossible to perform the chemical synthesis of such a huge number of compounds and to carry out *in vitro* assays of their biological activity. Even computer-aided virtual screening, *e.g*., the *in silico* analysis of interactions between all possible compounds and all target proteins, is nowadays practically impossible. The development of new methods for the synthesis of complex organic compounds (antibiotics, peptides, oligonucleotides and natural compounds) and the emergence of new areas of synthetic chemistry, in particular, combinatorial synthesis,[4–10] have resulted in a great increase in the number of synthetically feasible organic compounds, which can be tested for particular types of biological activity.

The design of novel ligands from fragments, which are selected based on the assessement of their contributions to particular biological activities, is employed for the discovery and design of drug chemotypes. Besides, this method allows one to avoid the rapid uncontrolled growth of the number of *in silico* generated compounds (the combinatorial explosion).

Fragment-based ligand design approaches provide possibilities to design new biologically active molecules containing building blocks, such as atoms, fragments or functional groups which are involved in interactions with a target protein or contribute to particular biological activity.

In pioneering studies, the generation of new molecules was performed using atoms as the building blocks.[11–18] An advantage of such approaches is that the exhaustive screening of the total chemical space can be carried out. In addition, in this case it is possible to predict changes in the properties upon even slight variations in the structure, for example, upon the addition of individual atoms in the course of the generation. Therefore, the optimization of properties is substantially simplified compared to the optimization of structures generated from fragments.[19] The main problem encountered when using atoms is that it is difficult to construct molecules of synthetically feasible stable compounds, which can be then tested for biological activity.[19,20]

According to the estimates,[21] the probability of the successful synthesis of novel structures constructed from

O A Filz, V V Poroikov V N Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, ul. Pogodinskaya 10, 119121 Moscow, Russian Federation. Fax (7-499) 245 08 57, tel (7-499) 255 30 29, e-mail: fioland@yandex.ru (O A Filz), tel. (7-499) 246 09 20, e-mail: vladimir.poroikov@ibmc.msk.ru (V V Poroikov)

Received 19 April 2011
Uspekhi Khimii **81** (2) 158 – 174 (2012); translated by T N Safonova

fragments is higher compared to the synthesis of structures generated from atoms. In addition, specially designed algorithms allow researchers to select fragments, which will make a particular (positive and negative) contribution to the desired properties of organic compounds. Consequently, an advantage of the use of fragments compared to atoms is that it allows a researcher to design novel structures having the desired set of features. It should be noted that if fragments are small, it is still possible to design novel chemotypes with desired properties.

There are the following advantages of the fragment-based design of novel molecules compared to conventional methods of biological screening and rational drug design (the selection of the most promising structures based on the results of molecular modelling, quantitative structure–property relationship analysis, *etc.*)

— in the step of the structure design, one can add chemical groups which enhance solubility and bioavailability of new compounds and exclude potentially toxic groups;

— a greater chemical diversity of the structures is at hand, resulting in the design of novel drug chemotypes;

— novel lead compounds can be discovered by combining fragments that are present in known pharmaceutical substances having the desired biological activity and that have weak binding affinity for different regions in the active site of the target protein;[22] therefore, high-affinity ligands can be designed for targets containing several shallow binding pockets by analogy with the assembly of large components from small units;

— multitarget molecules can be constructed;

— a larger part of the chemical space can be screened for molecules having particular biological activity.

A characteristics referred to as ligand efficiency (L) was proposed[23] for representing the advantages of fragment-based lead design.

$$L = \frac{-\Delta G}{H} = -\frac{RT \ln K_d}{H} \approx \frac{-RT \ln(IC_{50})}{H},$$

where ΔG is the Gibbs free energy difference between the free protein and the protein–ligand complex, H is the number of nonhydrogen atoms in the fragment, K_d is the dissociation constant of the protein–ligand complex, R is the universal gas constant, T is the temperature and IC_{50} is the half maximal inhibitory concentration.

As can be seen from this formula, the ligand efficiency is inversely proportional to the number of nonhydrogen atoms in the structure. Consequently, the concept of ligand efficiency relates the binding characteristics per nonhydrogen atom in the structures of the fragments to those of the drug-like compounds (compounds whose physicochemical properties are similar to the properties of known pharmaceuticals) generated from these fragments.

The ligand efficiency was calculated for the first time by Kuntz *et al.*[24] for a group comprising 60 proteins and their ligands. It was expected that the L values for high-affinity ligands would be not higher than 1.5 kcal mol^{-1} atom^{-1}. An analysis of L for fragments which were used for the design of novel drugs such as inhibitors of DNA gyrase, p38 MAP kinase, the blood coagulation factor Xa, phosphodiesterase IV, stromelysin, caspase and neuraminidase,[21] showed that $L < 0.5$ kcal mol^{-1} atom^{-1}. Therefore, the relative affinity for the target protein can be estimated using L of the fragments under consideration (provided that the IC_{50} values are of the same order of magnitude).

However, the ligand efficiency is calculated taking into account the binding energy of the whole molecule (fragment) and, consequently, it characterizes the molecule as a whole. The group efficiency (G), which accounts for changes in the binding energy ($\Delta\Delta G$) upon the addition of the functional group containing a specified number of nonhydrogen atoms, was proposed[25] as a measure of the influence of individual functional groups on the binding to the target protein

$$G = \frac{-\Delta\Delta G}{\Delta H},$$

where ΔH is the change in the number of nonhydrogen atoms.

It was shown[25] that as substituents are linked to the ligand its affinity gradually increases with an increase in the weight of the attached group and the number of nonhydrogen atoms. Consequently, the parameter G can be used to assess the contribution of a particular group only taking into account its size and the number of nonhydrogen atoms in this group.

In the general case, the fragment-based lead design includes the following steps:

— the selection of fragments from a library or the dissection of the molecule into fragments to design a fragment library;

— the selection of fragments based on the estimation of their contribution to desired properties;

— the generation of molecular structures;

— the selection of structures meeting the criteria specified by a researcher;

— the synthesis and testing of the candidates.

Depending on a particular goal, some of the above-mentioned steps can be bypassed.

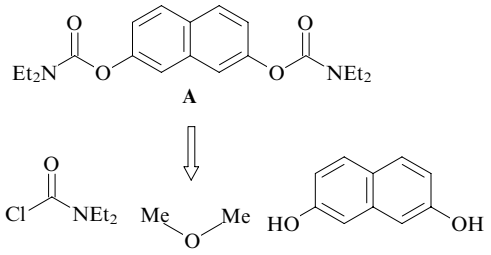
The present review considers computer-aided fragment library design, methods for the selection of fragments based on the estimation of their contributions to the required biological activity and strategies for *de novo* molecule design.

II. Fragment library design

Chemical fragment libraries are constructed mainly with the use of the following methods: retrosynthesis, pseudo-retrosynthesis and screening (relying on expert judgment) for molecules with molecular weights lower than 200–300 Da in database of commercially available compounds (Table 1). Several original approaches, which do not involve the retrosynthetic analysis, are also available (see below for details).

Retrosynthetic methods are based on the screening for known substructures (so-called synthons) in databases of chemical reactions and the decomposition of molecules into

fragments.[26] For example, three reactants can be used to obtain compound **A**



The retrosynthetic analysis of chemical structures can be compared with the game of chess in respect of the number of possible combinations, from which it is necessary to choose the correct solution.

According to Todd,[27] the retrosynthetic analysis can be classified into three approaches:

— direct associative approach, where the 'product' is represented as a simple collection of undisguised subunits and where it is necessary to merely reveal these subunits for the determination of the starting structures of the 'reactants';

— intermediate approach, where a complex structure of the 'product' differs from other synthetically accessible structures and the problem is to find the appropriate sequence of reactions for their interconversion; this approach is implemented, for example, in the WODCA, SYNOPSIS, EROS and TOSCA programmes;[28]

— logic-centred approach, where a logical analysis generates a 'synthetic tree' without any assumptions as to the starting structures,[27, 29] it is assumed that a computer can reproduce the major part of logical reasoning necessary for the discovery of chemical synthesis pathways only using a system of rules developed by a researcher; it is clear that limitations of this approach depend on the completeness of a set of these rules; this approach is implemented in the LHASA programme.[30, 31]

According to the classification by Zefirov and Gordeeva,[29] the development of programmes for retrosynthetic analysis is based on two approaches (empirical and non-empirical). In the empirical approach, it is assumed that transformations of the system under consideration are performed based on the data on known organic reactions. In the non-empirical approach, transformations are generated without any such knowledge.[29] In the former case, a system of rules developed by an expert chemist is used. These rules are written either in a special language [31] or in one of the available languages (for example, XML). In the latter case, a combinatorial algorithm (a set of combinatorial logic instructions) is used to find transformations. An advantage of the empirical approach is that the programme usually predicts plausible synthetic pathways, whereas a drawback is that it does not allow one to suggest a radically new pathway or to discover a novel reaction, because of the limitations of a particular transformation library. For programmes implementing the non-empirical approach, precise criteria for the selection are required to avoid obtaining inadequate results.

Programmes implementing the empirical approach employ transformation libraries sharing common features. The transformations are described by encoded structural fragments, which should be present in the system for this transformation to take place. In addition, the description of the transformation includes a set of structural changes required for the generation of structures corresponding to a particular transformation. Additional information can be encoded in the description of the transformation depending on its particular realization. For example, data on ~ 70 organic reactions which can be used for transformations of molecular structures are encoded in the SYNOPSIS programme.[32]

Table 1. Characteristics of main commercially available fragment libraries.

Library name (country)	Characteristics of the library	Number of building blocks
ACBBlock (Russia)	compounds belonging to different chemical classes	1280
	special NMR-directed fragment library	760
ADESIS (USA)	libraries of heterocyclic compounds and functional groups	1361
ALINDA (Russia)	fragment library of different chemical classes of compounds (M_w from 50 to 700 Da)	14 759
Activate-scientific GmbH (Germany)	fragment library of different chemical classes of compounds, synthons	7966
APICHEM (China)	different chemical classes of low-molecular-weight compounds	14 409
ASDI (USA)	different chemical classes of compounds	$\sim$10 000
Asinex (Russia)	building blocks for the *de novo* structure design	6200
	specially selected biological fragments	1700
BioBlocks (USA)	building blocks for different chemical classes of compounds selected based on pharmacophore properties	> 2200
BioNet (USA)	regullary updated library of building blocks for different chemical classes of compounds	7205
Combi-blocks (USA)	libraries for biochemical screening	–
May-Bridge (UK)	fragment library for NMR screening	7951
	fragment library for X-ray crystallographic screening	1500
Zerenex (UK)	screening library of heterocyclic compounds	> 8000

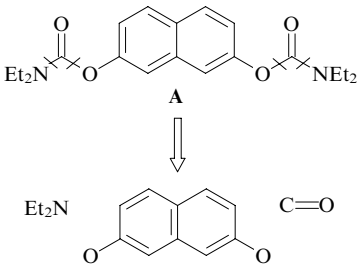
In programmes implementing non-empirical approaches, transformations are performed not based on the screening of libraries but as a result of particular logic operations. For example, the EROS[33] and TOSCA programmes[34] involve series of reaction generators, which describe, in the general form, the changes in the bonds (cleavage or formation of bonds, a decrease or an increase in their order) as a result of chemical reactions. For different reaction centres and different bonds, the same generator will create different chemical transformations.[29]

Approximately 50% of all organic reactions (elimination, aliphatic nucleophilic substitution, aromatic electrophilic substitution) correspond to this scheme. Therefore, the reaction generators can create new transformations for a particular system, which is impossible in empirical approaches. The main problem that is encountered in the programmes based on encoding of different types of reactions is the determination of the reaction centre.

The limitations of the retrosynthetic approach in the fragment library design are associated with the fact that fragments can substantially differ from the starting structures (for example, in reactions giving cyclic molecules starting from acyclic ones). Consequently, the biological properties of the fragments obtained in the framework of this approach may substantially differ from the properties of the starting structures.

In addition, the retrosynthetic analysis does not take into account several pathways of chemical reactions and, consequently, the most probable pathway is not discovered. Besides, it is impossible to take into account the differences in the chemical synthesis of optical *R*- and *S*-enantiomers. It is difficult to separate some stereoisomers, and the reaction predicted for one optical isomer may be impossible in practice.[35] Retrosynthetic methods rely mainly on information on known chemical reactions. Hence, this approach cannot be applied to molecules if chemical reactions for their synthesis are unkown, for example, in the case of natural compounds.

The fragment library design with the use of retrosynthetic methods relies on bond cleavage rules. For example, for compound **A**



A

The rules are based on the analysis of commonly used reactions that can be performed in a laboratory. The RECAP procedure[36] is the most well-known fragmentation algorithm. It includes the following default instructions:

— do not break bonds between nonhydrogen and hydrogen atoms;

— do not break bonds in the ring;

— do not break bonds between an atom in the ring and the attached heteroatom (except for particular cases specified by the user);

— the bond should not be cleaved if the fragment that is formed upon decomposition is in the list of forbidden fragments or the number of atoms or bonds in the fragment is smaller than the pre-specified value.

It is assumed that the cleaved bond depends neither on the type of the reaction nor on the nature of the starting compounds involved in the reaction.

The following advantages of this approach should be noted:

— fragments are generated from structures of drugs, which often display pharmacokinetic and pharmacodynamic properties (which are responsible for high bioavailability of the drug), resulting in the higher probability that novel molecules will exhibit such properties;

— the high probability of the synthesis of compounds, whose structures are made up of fragments.

The designed structures were shown to exhibit a high degree of similarity with known compounds exhibiting this activity.[37] A drawback of this approach is that the developed rules cannot be applied for the decomposition of molecular structures of some chemical classes into fragments. In particular, it was shown that 2682 (36%) of 7395 organic compounds from the COBRA database[38] were not cleaved into fragments with the use of the standard RECAP algorithm, because none of the RECAP rules are met by undivided structures.[39] Consequently, it is necessary to develop new rules, which would allow one to split structures of such chemical compounds into fragments, for which standard retrosynthetic and pseudo-retrosynthetic algorithms do not work.

There are several methods, which include rules similar to the RECAP rules for the generation of fragments from molecular structures, for example, the defragmentation method. This method is based on the disconnection of the molecular structure of the ligand into cyclic substructures and linear substituents.[40–46] In this case, the molecular graph is divided into substructures containing a ring, substituents and so-called linkages (chains of atoms and bonds) between cyclic substructures. Different versions of the pseudo-retrosynthesis used for the dissection of structures into fragments are implemented also in the TOPAS,[47] FlexNovo,[48] MEGA[49] and SkelGen programmes.[50]

In the RACHEL programme,[51] the molecular structures are cleaved into fragments at all rotatable bonds. It should be noted that this approach does not take into account which bonds can be cleaved in the course of the chemical reaction and which bonds, on the contrary, remain intact. When using pseudo-retrosynthetic algorithms, it is impossible to determine the sequence of chemical reactions giving new structures, which is considered as the main drawback of the method.[35]

Another way of designing fragment libraries is to select fragments based on expert judgment and the inclusion of these fragments into databases.[52,53] The main drawback of this approach is that it is time consuming and labour intensive. The selection of fragments and their inclusion into databases are generally based on several criteria, which should be met by new fragments.[54] Thus, by analogy with Lipinski's rule of five for drug-like structures, the rule of three was proposed for fragments.[55] According to this rule, the molecular weight of the fragments should be < 300 Da, the number of hydrogen bond acceptors is < 3, the number of hydrogen bond donors is < 3, $\log P < 3$ (P is the

partition coefficient in a n-octanol–water system). These are the main criteria for the fragment selection. The number of rotatable bonds <3 and the polar surface area of the fragment <60 Å^2 are additional criteria, which can be taken into account when selecting fragments.

In addition to methods for the fragment generation with the use of retrosynthetic and pseudo-retrosynthetic algorithms, several original algorithms were developed for the decomposition of molecular structures into fragments, which do not take into account the possibility of chemical bond cleavage. For example, three-dimensional molecular structures in the SQUIRRELnovo programme are represented by sets of mathematically defined surfaces (hyperbolic paraboloids).[56] Since the dissection is performed with preservation of the three-dimensional structures of the fragments, these libraries can be employed for screening with the use of molecular modeling methods. However, it should be noted that molecules of most of organic compounds can adopt different conformations. The retention of one conformation of the fragments upon decomposition of the molecule may lead to a loss of information in the design of novel molecules.

The selection of fragments based on the estimation of their contribution to particular biological activity is a necessary step, which allows one, first, to decrease the probability of combinatorial explosion in early generations and, second, to design high-affinity ligands.

III. Selection of fragments based on the estimation of their contribution to biological activity

Main difficulties in performing the fragment-based ligand design are associated with limitations of methods for the selection of fragments interacting with a pharmacological target.

Nowadays, fragments that can interact with a particular target protein are selected by estimating the inhibition constants from the results of *in vitro* experiments and with the use of computational methods for the estimation of the probability of interactions of low-molecular-weight compounds with target proteins.

1. Experimental methods

Limitations of experimental methods for the fragment-based molecular design are associated with requirements on an ideal target. The binding site of an ideal target consists of several shallow regions such as, for example, in the protein Bcl-xl (in this case it is necessary to investigate the location of several fragments in the active site) or contains a deep pocket as, for example, in the heat shock protein HSP90 (in this case new structures are formed by growing from one fragment). The ligand–protein binding affinity depends on the presence of the well-recognizable site and its environment, which increases the binding specificity and indirectly characterizes the ability of the protein to bind ligands.

In the study,[57] 23 proteins and 28 binding sites of these proteins were investigated. Some proteins (among which are Bcl-xl, MDM2, protein tyrosine phosphatase 1B, tyrosine protein kinase Lck and peptidyl-prolyl *cis–trans*-isomerase) are characterized by high affinity for drug-like compounds. For most of the binding pockets (71%) of such targets, the percentage of fragments capable of interacting with proteins was 0.1% (determined by NMR spectroscopy). Fragments with a molecular weight of <220 Da were used. For other targets (among which are bacterial transferase MurA, glutamate racemase MurI and aminoacyl tRNA synthetase SARS), the ability to bind low-molecular-weight compounds is unknown. The percentage of fragments which bind in the active sites of these proteins is much smaller (0.1%). It was suggested [57] that the percentage of fragments capable of binding to a particular protein can serve as a criterion of its ability to interact with low-molecular-weight compounds. Note that when using this criterion, it should be taken into account how long a particular protein was studied (for example, some proteins were investigated for tens of years, whereas other proteins were discovered and studied rather recently, and, consequently, the number of

Table 2. Main qualitative characteristics of methods for the selection of lead fragement.

Method	Possibility of distinguishing between binding in the active site and binding in the inactive site	Necessity of the data on the 3D structure of the receptor	Limitations of the method
High-throughput tests	impossible	not necessary	a large number of false-positive results
Surface plasmon resonance	possible in the case of strong immobilization of the target	"	false-positive results are not exlcuded
Affinity chromatography coupled with mass spectrometry	impossible	"	–
NMR spectroscopy (signal of the receptor or a fragment)	possible	necessary	high concentrations of proteins and ligands are required
X-Ray crystallography	"	"	high concentrations of proteins and ligands in solution are required
Docking	possible if the site is known from experimental data	"	false-positive results are not exlcuded
Pharmacophore model design	impossible	not necessary	difficult to achieve chemical novelty; false-positive results are not exlcuded

known ligands for these proteins may be substantially different).

Nowadays, the interactions between proteins and fragments of organic compounds are estimated by experimental methods, such as the high-throughput detection of protein–ligand interactions, surface plasmon resonance, affinity chromatography coupled with mass spectrometry, X-ray crystallography and NMR spectroscopy (Table 2).[58–60]

It should be noted that most experimental methods require that ligands be soluble in water and be present in solution at high concentrations. In this case, interactions between proteins and the fragment under study can be experimentally detected.

The binding of fragments to particular target proteins is experimentally studied with the use of commercially available fragment libraries (see Table 1). The list of companies selling fragment libraries was reported.[61]

a. Studies on panels of target proteins

Mechanisms of drug action can be studied at the molecular (submolecular) level with the use of panels of target proteins for the estimation of protein–ligand interactions. A microchip consists of several thousands of cells, each cell containing an individual target protein conjugated (bound) to the cell by a special chemical linkage. An analyte is added to each cell. Then an interaction between the target protein and the analyte is detected by measuring fluorescence or chemiluminescence.

This high-throughput method allows a researcher to test a large number of compounds belonging to one or several chemical classes and detect changes in the ligand affinity upon small changes in the structure. An additional advantage of this method is that it allow a researcher to study the structure–property relationship for lead compounds with the goal of optimizing their structures.

The main drawback of the method is a large number of false-positive results,[53] which substantially increases in the case of protein–ligand interactions characterized by affinity constants in the range of 10–100 μmol litre^{-1} (Refs 58, 62). This may be a consequence of a larger number of nonspecific interactions compared to interactions of drug-like compounds.

Second, low solubility of the ligand under investigation and its weak own fluorescence or chemiluminescence can cause errors in the detection of fragments that make a contribution to the interaction with the protein.[63] Human errors in the calculations of the concentrations of the fluorescent or chemiluminescent agent also cannot be excluded.

To minimize false-positive results [when fluorescence is observed in the absence of protein–ligand (fragment) interactions], a method was developed based on measurements of the fluorescence signal characterizing the protein–ligand interaction coupled with measurements of the energy released upon changes in the conformation of the protein molecule induced by this interaction.[64] This approach was applied[65] for the discovery of inhibitors of the enzyme BACE-1, which is involved in the synthesis of β-amyloid and plays a substantial role in Alzheimer's disease pathogenesis. It should be noted that even this method does not allow users to differentiate between the nonspecific binding of the enzyme and the binding in the active site based of the detected signal. Apparently, this is associated with the insufficiently complex molecular structures of the compounds under study. In this case, the complete specificity of the binding in the active site of the target protein is not provided.

b. Affinity chromatography and mass spectrometry

Affinity chromatography is based on selective interactions of proteins with affinity labels covalently bound to an inert carrier. Immunoglobulin G (IgG)-binding fragments, which specifically interact with the protein under study,[66] and low-molecular-weight ligands can be used as affinity labels.

Affinity chromatography coupled with mass spectrometry was initially used to study protein–protein complexes.[67] For example, a large German research team[68] studied hundreds of proteins from the yeast *Saccaromyces cerevisae* and investigated their ability to form complexes. More recently, the combination of these methods was applied for the estimation of protein–protein interactions for protiens from other organisms, including mammals.[69]

Ganem *et al.*[70] were the first to use mass spectrometry for the estimation of protein–ligand interactions. In particular, it was shown that molecules that are present in solution form a protein–lligand complex through noncovalent interactions and that this complex can be transferred to the gas phase, where it will be subjected to ionization (for example, to electrospray ionization). The conclusion about the interaction between the target macromolecule and the ligand can be drawn from the data on the mass, the charge and the distribution density of charged particles of the complex. This approach was used for the design of inhibitors of the 23S rRNA subunit of the bacterial ribosome.[71]

The matrix-assisted laser desorption/ionization (MALDI) method, which employs another ionization technique, is not widely used in studies of weak non-covalent interactions between proteins and ligands. The MALDI method can be applied for the estimation of these interactions under certain conditions (in particular, high acidity of the medium and certain types of matrices are excluded).[72]

The possible formation of a disulfide bond between sulfur atoms of the fragment and a cysteine residue at the binding site of the protein is taken into account in a modification of mass-spectrometric analysis referred to as tethering.[73,74] If a cysteine residue is absent at the active site, it can be inserted by conventional mutagenesis methods. The covalent binding of the ligand fragment in the active site of the protein can be determined from the analysis of the electrospray ionization mass spectra.[73,74] The tethering technique was successfully used for the discovery of low-molecular-weight inhibitors of thymidylate synthase[73] and inhibitors of the interaction between interleukin-2 and interleukin-2 receptor α.

Mass spectrometric methods can be applied not only for the detection of changes in the structure (conformation) of proteins induced by their interactions with low-molecular-weight compounds but also for the identification of the resulting particles.

An advantage of affinity chromatography coupled with mass spectrometry is that mixtures of low-molecular-weight ligands and proteins can be separated. In addition, studies of protein–ligand complexes do not require the introduction of labels into ligands or target proteins and, consequently, the risk of false-positive results is reduced.

Limitations of the method are that it is impossible to perform the differentiation of the binding (in the active or

inactive sites) due to low molecular weights of the fragments and, consequently, low affinity constants compared to those of the larger ligand molecules. An evident limitation of the tethering technique is that it can be used only for the screening of libraries of sulfur-containing ligand fragments. In addition, sepharose used as a carrier in affinity chromatography is unstable to chemical compounds and microorganisms.

An analogue of the tethering technique, the so-called extended tethering, is used for the screening of structure-based libraries to study interactions of fragments in the immediate vicinity of the known ligand.[73] In this method, a fragment (known ligand) contains an additional protected thiol group. After the formation of a covalent disulfide bond at the active site of the protein, the thiol group can interact with the disulfide group to form a covalent bond. Then the protein–ligand complex is tested for the ability to interact with thiol-containing fragments, thus providing data on the favourable mutual arrangement of fragments in the active site.

Since the tethering technique is based on the ability to form a covalent bond between atoms of the fragment and the target, versions of this approach can be used to search for fragments that form covalent bonds with the protein in question (for example, with the prosthetic group of haemoglobin). However, to the best of our knowledge, this method was not applied for the screening of fragment libraries.

c. Surface plasmon resonance

Biological target molecules can be deposited onto metal thin films. An interaction with biological objects (DNA, viruses, antibodies) causes changes in the refractive index of the film, resulting in the change in the intensity of the reflected light, the phenomenon referred to as surface plasmon resonance.

Initially, the surface plasmon resonance technique was used to investigate protein–protein or protein–ligand interactions (at submicromolar concentrations).[75] More recently, it has been shown that this method can be used for the detection of fragments characterized by weak interactions with a target protein. For example, the ligand design was performed for the protein BACE-1.[76] Nowadays, the surface plasmon resonance technique is used primarily for the refinement of results of previous experiments and to confirm that the selected fragments are lead compounds for drug discovery.[60]

Difficulties in applying the surface plasmon resonance method are associated with that it is necessary to immobilize targets in such a way that the binding site will be accessible for the ligand. Once a relatively high degree of immobilization is achieved and the target protein is immobilized on the support in an appropriate manner, it is possible to distinguish between specific and nonspecific interactions, calculate the dissociation constant and determine whether the binding is reversible.[60]

Based on the classical version of this method, a technique was developed for the detection of fragments that influence the binding of the ligand to the target protein.[77] Microchips with covalently attached $\sim$9000 fragments were used for this purpose. A solution containing the protein under study was passed through the microchip and its binding to fragments was detected. This method allows one to distinguish between fragments, which interact with the target protein and which are not involved in these interactions. However, this method does not distinguish between specific and nonspecific binding, because fragments can be fixed in such a way that there will be no contact with the active site of the protein.[78]

Surface plasmon resonance can be employed in combination with X-ray crystallography and biochemical screening. This combination was used for the fragment-based molecular design of inhibitors of hepatitis C virus RNA polymerase.[78]

The main advantage of this method is that the concentration of ligands can be varied from micro- to nanomolar values and, consequently, IC_{50} can be evaluated, which is impossible when microchip panels, affinity chromatography and mass spectrometry are used.

As opposed to X-ray crystallography and NMR spectroscopy, the detailed investigation of the location of ligands in the active sites of macromolecules cannot be performed by surface plasmon resonance.[79]

d. X-Ray crystallography

One step in studies of interactions between fragments and target proteins involves the treatment of protein crystals with highly concentrated (25–100 mmol litre^{-1}) ligand solutions or a mixture containing several ligands. To evaluate the parameters of the protein–ligand complex, it is necessary first to obtain crystals and then to study them by X-ray diffraction. Consequently, X-ray crystallography of protein–ligand complexes requires protein crystal (to obtain crystals, highly purified proteins, which should not precipitate, are used). In addition, the targets should be conformationally stable. In spite of the fact that automated methods for the protein crystal growth have been recently developed, including methods for the optimization of physicochemical conditions of the synthesis (pH of the medium, temperature, the protein concentration in solution),[80] it is rather difficult to achieve required concentrations of a highly purified protein.

Besides, in the case of target proteins integrated into cellular membranes, there are additional difficulties with the protein purification. Let us refer to receptors coupled with G-proteins and a series of ion channels as examples.[81]

The following advantages of X-ray crystallography should be noted:

— short time of investigation (provided that a protein crystal is available);

— the possibility of distinguishing between specific and nonspecific interactions;

— the possibility of distinguishing between the binding in the active site and the binding in the allosteric regulation site;

— the possibility of assessing the character of binding of the ligand fragment to the target protein and their conformations;

— high sensitivity, *i.e.*, the possibility of determining even weak interactions between the protein and the ligand (10^{-5}–10^{-6} mol litre^{-1}).

The fragment screening by X-ray crystallography was used to discover fragments interacting at the active site of thrombin followed by the molecular design of thrombin inhibitors.[82] Fragment libraries were employed to discover cyclin-dependent kinase 2 inhibitors necessary for the design of drugs against malignant neoplasms.[83] Fragments that contribute to interactions with cyclin-dependent kinase 2 were found by X-ray crystallography of a series of complexes of the fragments with target proteins. The structure of the fragment was further optimized taking

into account the data on interactions between individual atoms of the target protein and the low-molecular-weight compound also determined by X-ray diffraction.

X-Ray crystallography combined with biochemical screening can be used for the fragment-based lead design. For example, more than 238 lead fragments were selected for the molecular design of novel B-Raf kinase inhibitors, and the three-dimensional structures of ~100 protein–ligand complexes were solved. Based on these data, a novel high-affinity B-Raf kinase inhibitor PLX4720 was discovered.[84] A similar combined approach was applied to the molecular design of phosphodiesterase-4 inhibitors.[85]

e. NMR spectroscopy

To obtain information on the location of the ligand in the binding site, signals of the isotope-labelled protein are recorded. For example, interactions between amide groups and ligands are detected with the use of a ^{15}N label. The methyl groups are labelled with ^{13}C. Interactions between the protein and the ligand can be determined from the chemical shifts in the NMR spectra. The shifts occur due to the changes in the environment of the active site of the protein caused by this interaction. Studies by NMR spectroscopy taking into account the signal of the target made it possible to discover inhibitors of the protein Bcl-2 necessary for the design of drugs against malignant neoplasms.[86] The signal of the ligand is most often recorded with the use of an ^{1}H label.[87]

NMR spectroscopy and X-ray crystallography are most accurate methods, which not only enable the visualization of the fragment binding to the target based on high-precision data on the structures of protein–ligand complexes but also, in most cases, allow one to distinguish between binding in the active site and binding outside the active site.[58,88] These methods can be used to determine the exact location and conformation of the fragment in the binding site.[58] Unlike X-ray crystallography, NMR spectroscopy can be employed to study conformationally flexible protein molecules.[89] In addition, NMR spectroscopy provides reliable information on interactions between targets and low-molecular-mass compounds, whose affinity constants are in the range of 10–1000 μmol litre^{-1}. In the latter cases, some biological screening methods give a large number of false-positive results due to an increase in the percentage of nonspecific interactions.[57] The main drawback of NMR spectroscopy is that there are constraints on the protein size.[60]

2. Computational methods

Nowadays, the following main computational methods are employed to investigate interactions between fragments of organic compounds and macromolecules: the molecular docking, the construction of pharmacophore models and computational methods for the quantification of structure–activity relationships, as well as molecular mechanics methods (MM-PBSA, MM-GBSA, free-energy perturbation method).†

† In the present review, these molecular mechanics methods are not considered because they are discussed in detail in the following publications: J H Lin *Curr. Top. Med. Chem.* **11** 171 (2011); V Zoete, M B Irving, O Michielin *J. Mol. Recognit.* **23** 142 (2010); C Oostenbrink *Methods Mol. Biol.* **819** 487 (2012).

a. Molecular docking

Molecular docking belongs to molecular modelling methods. It considers numerous hypotheses about the mutual arrangement of molecules in complexes and their conformations. The binding energy is evaluated for each conformation and each position of one molecule with respect to another one. This energy is the sum of the energies of hydrophobic, electrostatic and van der Waals interactions and the hydrogen-bond energy. The goal of molecular docking is to find the position of the ligand in a complex corresponding to the global energy minimum. An algorithm for finding the global minimum varies depending on a particular molecular docking method. The genetic algorithm[90,91] and the Monte Carlo method[92] are most often used for this purpose.

An analysis of several local minima of the binding energy followed by the selection of the global minimum is also employed.[93] More than 30 computer programmes are available for molecular docking, among which are AUTODOCK,[90,94] GOLD[95] and FlexX.[96] Unlike the docking of the whole ligand structure, in the case of the stepwise addition of functional groups or fragments to the starting structure it is necessary to evaluate the binding energy in each step. In addition, it is necessary to automatically remove fragments that make a negative contribution to the interaction with the protein of interest and to preserve the best results used in subsequent steps of the design. During the development of a method for the fragment-based lead design, specialized programme packages were created: Skelgen,[51] Lea3D,[43] SYNOPSIS,[32] FLUX[39,97] and TOPAS.[47]

In molecular docking calculations coupled with the *de novo* design, the most active (according to the prediction) molecules are selected (generally, based on the estimation of the binding energy) in each step of the structure transformation.

Molecular docking methods are often used in combination with other experimental and theoretical methods. For example, NMR spectroscopy was applied to search for a compound capable of inhibiting the heat shock protein HSP90 at high millimolar concentrations, which is a promising target for the cancer treatment.[98,99] Then analogues of this compound were docked into the active site of HSP90, and the hits were experimentally tested. This study resulted in the discovery of HSP90 inhibitors active at nanomolar concentrations.

In spite of the wide use and rapid development of the molecular docking, this method has some limitations. The most substantial one is that it requires the knowledge of the three-dimensional structure of the target protein. For instance, the experimental determination of the three-dimensional structures of membrane-bound proteins or proteins with a high molecular weight presents serious problems. In addition, there is no clear correlation between the calculated and experimental binding energies.[100] To evaluate the binding energies and the optimal conformations of the protein and the ligand with high accuracy, it is necessary to know the position of the binding site (or of the allosteric binding site) and the location of the ligand in this site. These data substantially influence the accuracy of the identification of potentially active molecules.[101]

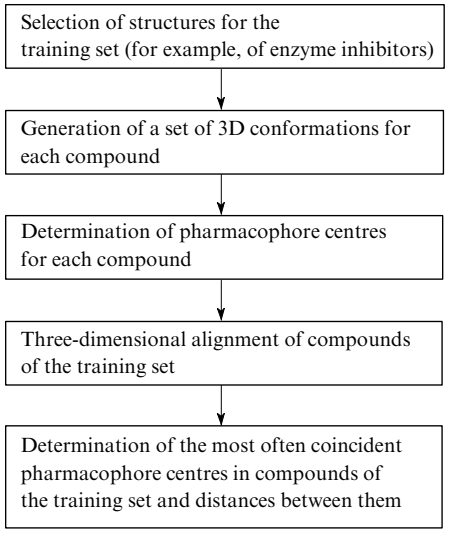
Examples of the successful application of molecular docking for the search for fragments involved in interac-

tions with a protein were reported.[62] Due to a small size of fragments and their considerable conformational flexibility, the accuracy of the determination of the binding energy for fragments is lower than that for drug-like molecules.[101]

b. Pharmacophore models

The purpose of the pharmacophore hypothesis generation is to determine a set of pharmacophore centres and distances between them necessary for intermolecular interactions with the target protein. Hydrophobic regions, aromatic rings, hydrogen bond donors and acceptors and positively and negatively charged regions can serve as pharmacophore centres. A more detailed description of a pharmacophore can include such parameters as excluded volumes, the allowed ranges of the angular orientation of hydrogen bond vectors and the planes of aromatic rings.[102] The construction of a pharmacophore model involves the generation of multiple conformations for each structure from the training set and the determination of the conformational centre for each compound followed by the 3D pharmacophore alignment. The result is a set of pharmacophore centres and distances between them. As a rule, several pharmacophore hypotheses are initially generated, and the final hypothesis is chosen by a researcher. The goal of the substructure search is to find the match between the pharmacophore model and the characteristics of the molecules having allowed conformations and included in databases of chemical compounds. The general algorithm of the pharmacophore search is shown in Scheme 1.

Scheme 1



It should be noted that the pharmacophore search is often performed in combination with other methods of protein–ligand interaction analysis, which require data on the three-dimensional structure of the target protein. For example, inhibitors of prostaglandin D2-synthase were discovered by NMR-based screening of ~2500 fragments,[103] resulting in two dozens of primary hits. They were docked into the active site of the target protein. The conformations of the ligands, which were found by the docking, were used for the construction of the pharmacophore model and the selection of structures for testing. A combination of pharmacophore modelling and biological screening was also applied for the discovery of bacterial DNA gyrase inhibitors.[104] In the latter case, the pharmacophore hypothesis was used for the substructure search in databases. The screening of the hit ligands was performed by NMR spectroscopy and surface plasmon resonance.

If the structures of the test set substantially differ from the structures of the training set, the pharmacophore units after the three-dimensional alignment will differ from those of the training set. This will lead to underestimated values of the scoring function and, as a consequence, to false-positive results. Consequently, the limitations of the pharmacophore-based search in fragment space are associated with the requirement of high similarity between the three-dimensional structures of compounds of the test set and the structures (fragments) of the training set.[105]

c. Computational methods for quantitative structure–activity relationship analysis

The contributions of atoms and/or fragments to particular biological activity can be estimated by (quantitative) structure–activity relationship [(Q)SAR] analysis.

Investigations of structure–activity relationships by (Q)SAR do not require the knowledge of the mechanism of action of chemical compounds; however, it is assumed that this mechanism is the same for all compounds used for the model design. In the simplest case, the common biological activity is determined as a function of the sum of contributions from individual activities of all substituents and the contribution of the scaffold.[106] Therefore, the relationship between the biological activity (A) and the structure of the chemical compound is expressed as the linear regression equation

$$A = a_0 \sum a_i x_j ,$$

where a_0 is the contribution from the molecular structure of the scaffold, which is the average activity of this series of structures, a_i is the contribution from the ith substituent to the total activity, x_j is an independent variable equal to unity or zero if the ith substituent is present or absent in the structure.

The prediction of the values of biological activity from the training set analysis can be performed with the use of multiple and nonlinear regression models[107] and neural network models,[108] which are based on the analysis of relationships between biological activity and various fragment descriptors. Correlations for the prediction of activity of novel analogues of the compounds under study are also available. The ligand binding to a target is determined by electrostatic, hydrophobic and van der Waals interactions, as well as by hydrogen bonding.[109] Hence, the fragment descriptors describe mainly these properties.

Let us give an example of the application of QSAR for the estimation of the contributions of fragments to particular biological activity. Du *et al.*[110] proposed two types of coefficients:

— for the description of physicochemical properties of the molecule;

— for the description of contributions of individual fragments to the activity of the molecule (weight factors).

A model of the inhibition of H5N1 influenza virus neuraminidase was constructed as an example. The method was validated using the random splitting into training and

test sets in a ratio of 80 : 20. The correlation coefficients between the experimental and theoretical values of IC_{50} were calculated for compounds from the training and test sets (the correlation coefficient between the observed and calculated values for the training set is 0.9973; the correlation coefficients between the observed and calculated values for the test set is 0.78). In the authors' opinion, this model can be applied for the optimization of the structures of lead compounds. However, the design of novel drugs using the estimation of fragment contributions was not performed, and, consequently, the predictivity of this approach remains unclear.

The use of the QSAR analysis for the estimation of contributions from individual atoms to the inhibition of angiotensin-converting enzyme (ACE) and dihydrofolate reductase (DHFR) was described.[111] Spowage *et al.* developed descriptors that account for the topological distances between pairs of atoms and their individual properties [topological maximum cross correlation (TMACC) descriptors]. Each pair of atoms in the structural formula of the compound is characterized by a TMACC descriptor. The regression equation was constructed for a combination of TMACC descriptors of each compound; TMACC descriptors were used as independent variables; the contributions from pairs of atoms were dependent variables.

The parameters of the model were determined by the least-squares method. Each physicochemical property was characterized by the positive or negative effect on the activity under study depending on the sign of the calculated coefficient. After the determination of the parameters of the model, the contributions (positive or negative) to the property were evaluated for each property. Five types of the effect on the activity (inhibition of ACE and DHFR) were specified for each atom: strong positive, positive, neutral, negative and strong negative. Based on the estimation of the contributions of individual atoms, the functional groups that can play a role in the activity were predicted. However, the results obtained in this study were not applied to the design of novel ACE and DHFR inhibitors, *i.e.*, the validation of the method was performed only retrospectively with the use of known literature data.

The limitations of the QSAR method as applied to the fragment-based lead design are that the compounds in question should be structurally similar to the compounds from the training test. The main difficulties encountered in designing models for the quantitative structure – activity relationship analysis are associated with the need to collect large biological activity data sets obtained in the same conditions (the same cell lines, proteins, *etc.*).

Once atoms or functional groups making (according to the prediction) a certain contribution to particular biological activity are found, the strategy for the further structural optimization can be chosen. However, it is not always possible to choose the training set for particular proteins such that it will enable the selection of fragment hits for the generation of novel structures.

In the next step following the selection of hits, molecules are assembled from the building blocks. The *de novo* design most often includes the automated generation with a complete screening of all possible molecular structures. It if is necessary to decrease the total number of generated structures, certain constraints are introduced.

IV. Methods of fragment-based generation of molecular structures

There are many approaches to the fragment-based generation of novel structures. The main goal of these approaches is to form synthetically feasible chemically stable molecules, which should exhibit required physicochemical or biological properties. Depending on the specificity of a particular problem, the following seven‡ methods of the fragment-based generation of molecular structures can be used:

— addition of substituents to a scaffold;

— merging of fragments (either random or with application of rules);

— structure transformations using a genetic algorithm;

— template screening;

— structure transformations by chemical reactions;

— structure transformations through biological processes;

— combination of different methods.

The main advantage of the methods based on the addition of all possible substituents to a scaffold[112,113] is that they enable the analysis of all structures of compounds belonging to a particular chemical class. In addition, changes in physicochemical or biological properties of compounds belonging to the same class can be detected even in the case of slight structural changes. A drawback of this approach is a limited chemical novelty of the generated structures.

When merging fragments, all possible ways of their connection at atoms or adjunction points, which are specified by a researcher or a computer programme, are employed.[44,114–117] In some cases, sets of rules are applied to increase the feasibility of the synthesis of the designed molecules. For example, a penalty is given for the formation of bonds between heteroatoms (MCDNLG),[118] bonds between particular pairs of atoms are forbidden (GrowMol and GroupBuild)[119,120] or the formation of bonds only between functional groups and carbon atoms in the sp^3-hybridized state are allowed (BUILDER).[121] In some methods, the RECAP procedure is employed not only for the decomposition of molecules into fragments but also for their subsequent merging into a new structure.[39,98] Additional constraints may be imposed on the assembly of building blocks. For example, every fragment may be connected only to relevant fragments based on the analysis of feasible structures.[122]

The fragment optimized growth algorithm is a special case of fragment merging.[123] In this method, small fragments (atoms, functional group, rings and adjacent rings) are connected by single bonds. In each generation, the probability that two fragments are merged is taken into account. The probability is calculated from the occurrence frequency of substructures containing both fragments linked by a single bond in databases of drug-like compounds.

An advantage of this method is that it generates structures of chemically stable and synthetically feasible compounds. A drawback is a limited structural diversity of generated molecules as a consequence of the constraints on

‡ For the first five algorithms, detailed consideration was reported.[19]

the bond order (only single) and the merging only of fragments that are present in the training set (database). It was supposed [123] that the use of fragments from databases of drugs increases the probability that novel compounds will be synthetically accessible and that their physicochemical properties will be similar to those of the already available drugs.

Approaches, in which several fragments are merged into a structure, make it possible to cover a larger part of chemical space as compared to methods, in which all possible substituents are added to a chemical scaffold.

Fragment-based design methods relying on a genetic algorithm also have gained wide acceptance.[43, 124–129] These methods are based on Darwinian principles of natural selection, where fragments are represented as genes. Mutations, deletions and insertions are correspondingly considered as changes in a particular fragment and its addition to or deletion from the structure. A crossover is a merging of two molecules containing two or several fragments. The estimation of the affinity for a particular target [98, 128] or the estimation of the feasibility of the synthesis of these compounds [125] can serve as a fitness function characterizing the survival of the individuals (molecules that pass to the next generation).

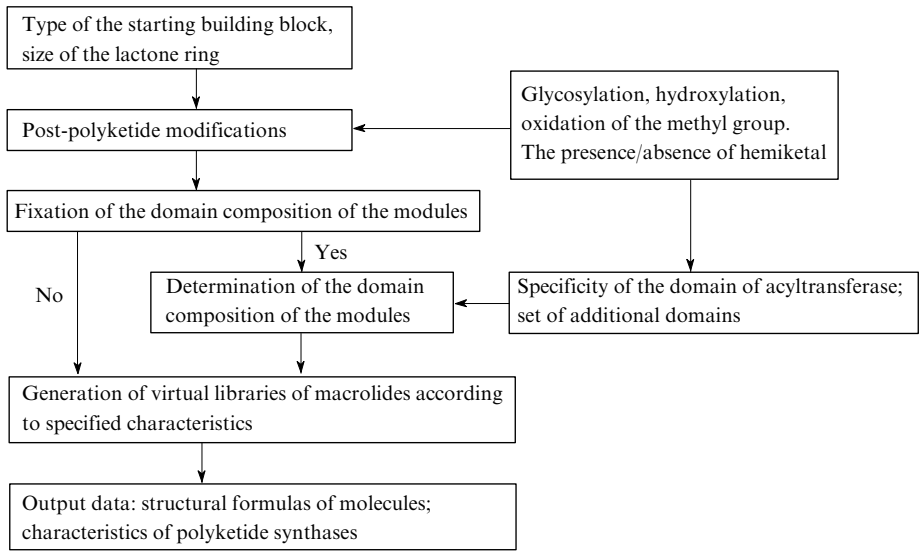
Template-based methods imply the generation of all possible molecular graphs with pre-specified constraints (on the number of nonhydrogen atoms, molecular weight and empirical formula). Then, in accordance with the generation task, certain vertices of the graph are replaced by particular atoms. First attempts to use templates for the structure generation have been made by designers of the Congen and Genoa programmes.[130]

The so-called superatom (a cyclic substructure containing several free valences) serves as a template. The catalogues in the Genoa programme includes $\sim$300 different superatoms. The simplified generation scheme consists in the following. In the first step, all possible superatoms (molecular graphs) are created according to the molecular formula. For each superatom, vertex graphs are generated based on the cyclic scaffold, from which all atoms are removed; the number of edges in the graph < 3. This process results in the generation of acyclic structures. In subsequent steps, cyclic and chelate scaffolds of the molecule are created. Initially, all bonds are removed from the chelate scaffold in such a way that all atoms do not contain free valences. Then cyclic scaffolds are formed from these chelate scaffolds. The chelate scaffold is a cyclic scaffold, whose vertices have free valences. Then all possible cyclic scaffolds, whose vertices have bonds with different types of atoms, are generated followed by the creation of all possible combinations of cyclic and acyclic scaffolds. Isomorphic structures are recognized and removed in all steps of the generation process.

This approach was used [2] for the generation of molecular structures containing up to 11 nonhydrogen atoms (C, N, O, F). In this study, all possible molecular graphs containing one to eleven vertices were initially generated. Then filters were applied to the set of generated graphs to remove the graphs corresponding to scaffolds of unrealistic chemical structures. In the next steps, filters were applied for the recognition of isomorphic graphs to prevent generation of repeating chemical structures. Each vertex of the graph was alternatively replaced by C, N, O and F atoms. More than a billion of possible structures were generated. After the application of filters, which took into account chemical stability and stereoisomerism, $\sim 26.4 \times 10^6$ unique structures of compounds with a molecular weight of $\sim$153 Da were obtained. This method of the generation of molecular graphs was employed for the design of molecules containing up to 15 nonhydrogen atoms (C, N, O, S, Cl).[3] A total of $\sim 28 \times 10^9$ structures with a molecular weight of lower than 210 Da were obtained. It should be noted that an increase in the molecular weight of compounds and, as a consequence, the necessity of the generation of molecular graphs with a larger number of vertices than those created by the authors of these studies inevitably leads to a combinatorial explosion. Hence, to design novel drugs it is necessary to create new libraries of compounds satisfying one or several chosen criteria.[3]

Transformations of molecules based on the data on biological processes are implemented in the algorithms of the BIOGENERATOR software.[131] This software generates virtual libraries of macrolides by modelling changes in the structure of a complex of Type I polyketide synthase,

Scheme 2



which catalyzes the assembly of the lactone ring in bacteria (mainly in the *Actinomyces* and *Streptomyces* families). Scheme 2 illustrates the operation of the BIOGENERATOR software.[132] Novel macrolides are designed by setting generation parameters. When setting a task, a user specifies the type of the starting building block (acetate/propionate) and the size of the lactone ring of macrolide determining the number of modules in polyketide synthase. In addition, the user specifies the site and type of post-polyketide modifications. Once these parameters are determined, virtual libraries of macrolides are generated. Fragments of the lactone ring and fragments corresponding to post-polyketide modifications of macrolides serve as building blocks.[132]

There are a number of methods, in which the reaction product is generated by the summation of the so-called reaction and reactant vectors.[133] Reaction vectors describing changes during the chemical process were designed[134] for the classification and discovery of novel chemical reactions. These vectors are created based on descriptors, which appear for the products of chemical reactions and are lost for the reactants. A set of descriptors of each reactant (for example, topological characteristics of the molecule) is transformed into a vector of the reactant involved in the reaction. The reactant vector (***R***) is created by the summation of the vectors for all compounds involved in the reaction:

$$\boldsymbol{R} = \boldsymbol{R}^1 + \boldsymbol{R}^2.$$

Analogously, the vector of each reaction product is created based on a set of descriptors of this product. The vector of the reaction products (***P***) is obtained by the summation of the vectors of all products

$$\boldsymbol{P} = \boldsymbol{P}^1 + \boldsymbol{P}^2.$$

The reaction vector (descriptor) is the difference between the product and reactant vectors

$$\boldsymbol{D} = \boldsymbol{P} - \boldsymbol{R}.$$

A set of descriptors of the generated compound (***X***) is determined by the equation

$$\boldsymbol{X} = \boldsymbol{D} + \boldsymbol{R}',$$

where $\boldsymbol{R}'$ is the vector of the reactant under study.
It is clear that the choice of the descriptor D for each new compound employed for the generation presents an individual complex problem. Consequently, the feasibility of the synthesis of compounds generated by this method substantially depends on the quality of the determination of the reactant vector – reaction vector pair. In addition, the completeness of the described chemical processes depends on the number of available original chemical reactions for compounds belonging to a particular chemical class of the training set.

It should be noted that, despite some common features of the methods of structure generation, there are a great diversity of algorithms for the solution of particular problems. In some methods, different combinations of algorithms for the generation of molecules from fragments or individual atoms are implemented depending on the specificity of tasks.

The problem of a combinatorial explosion is most often solved by performing the generation with constraints. There are deterministic and non-deterministic approaches.[35] In deterministic approaches, the affinity constants are determined or energies of the binding of the target molecule to a novel ligand are evaluated by molecular modelling methods[48] or QSAR.[26] In non-deterministic approaches, a genetic algorithm or its modifications are applied for the selection of individuals for the next generation.

When using the QSAR analysis, a researcher specifies which structural properties are responsible for the desired biological activity. Taking this into account, the researcher formulates the generation task, which usually includes the formation of a scaffold, the selection of fragments linked to the scaffold (substituents) and constraints. The latter are determined by the number of rings or the maximum distance from the atoms of the scaffold to the most distant atom of the substituent. The generated compounds corresponding to the structural formulas are tested for the similarity to the model. The GOLD-generator[135] is an example of generators based on this approach. In the authors' opinion, since QSAR models are often employed for the description of biological activity of numerous organic derivatives, virtual libraries created with the GOLD-generator are suitable for QSAR studies, because novel structures are generated by the addition of substituents to the central scaffold.

Melnikov *et al.*[136] performed the automated screening of all possible substituents for the central scaffold in the context of the generation task and selected structures that match the model.

In the probabilistic approach, a genetic algorithm (or its modifications) for the selection of leads is employed in each generation. The FLUX programme utilizes the pharmacophore alignment for this purpose.[98]

Despite the great efforts made by several research groups, there are no universal (simultaneously satisfying all criteria) generation algorithms for the design of novel synthetically feasible chemotypes with desired biological and physicochemical properties.[19] There are several approaches allowing for the compliance of novel structures with several criteria, for example, in the case of interactions with several target proteins,[48] but without any influence on the targets that are associated with side effects.[43] However, in these approaches the selection is performed after the generation procedure and, in addition, they require data on the three-dimensional structure of the target protein.

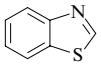
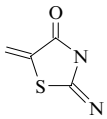
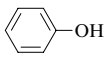
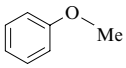
Consequently, the development of an algorithm for the fragment library generation based on the estimation of the contributions of fragments to interactions with a set of target proteins is still topical. In addition, the efficient use of fragment libraries requires the development of a generation algorithm for the design of novel chemotypes meeting the specified criteria.

We proposed to solve the problem of the generation of fragments for the design of synthetically feasible structures by applying rules that consider the energy of the bond to be cleaved. It is supposed that the probability of the bond cleavage is inversely proportional to the bond dissociation energy.[137, 138] When a molecular structure is split into fragments, the index numbers of the vertices that form the cleaved bond are preserved to be used in a generation

algorithm. In addition, information on atoms that form each cleaved bond are preserved in the code.

The problem of the multicriteria selection as applied to biological activity can be solved by employing the algorithm implemented in the PASS programme,[139] which enables the prediction of more than 3000 molecular mechanisms of action of pharmacological substances. The contribution of a fragment to each type of biological activity is estimated based on the average contribution calculated for each atom. In calculations of the estimates for individual atoms, the influence of the neighbourhood of each atom in the entire structure of the compound can be taken into account by using descriptors of multilevel neighbourhoods of atoms (MNA), which are generated on the basis of recursively defined rules and describe the structural neighbourhood of each atom.[140] The contributions of atoms are calculated from the occurrence frequencies of a set of MNA descriptors in the neighbourhood and those including a particular atoms in active and inactive structures of the training set. Then the estimate of the contribution of every fragment is normalized taking into account the occurrence frequency of the fragment in molecular structures, which do not exhibit or have particular biological activity (for example, interaction with the target protein). It should be noted that this approach takes into account the effect of each atom, which is present in the entire structure of the compound before its decomposition into fragments (by describing each atom by a combination of MNA descriptors), on particular biological activity. In addition, the estimation of the occurrence frequencies of the fragments generated after the application of the rules provides a possibility to take into

Table 3. Contributions of some fragments to the inhibition of COX-1, COX-2 and LOX.

Fragment	Object of inhibition	K [a]
	COX-1	89
	COX-2	100
	LOX	69
	COX-1	85
	COX-2	94
	LOX	79
	COX-1	100
	COX-2	90
	LOX	100
	COX-1	100
	COX-2	100
	LOX	84
	COX-1	85
	COX-2	87
	LOX	97
	COX-1	77
	COX-2	95
	LOX	93
	COX-1	73
	COX-2	100
	LOX	–13

[a] The coefficient characterizing the contribution of the fragment to the inhibition.

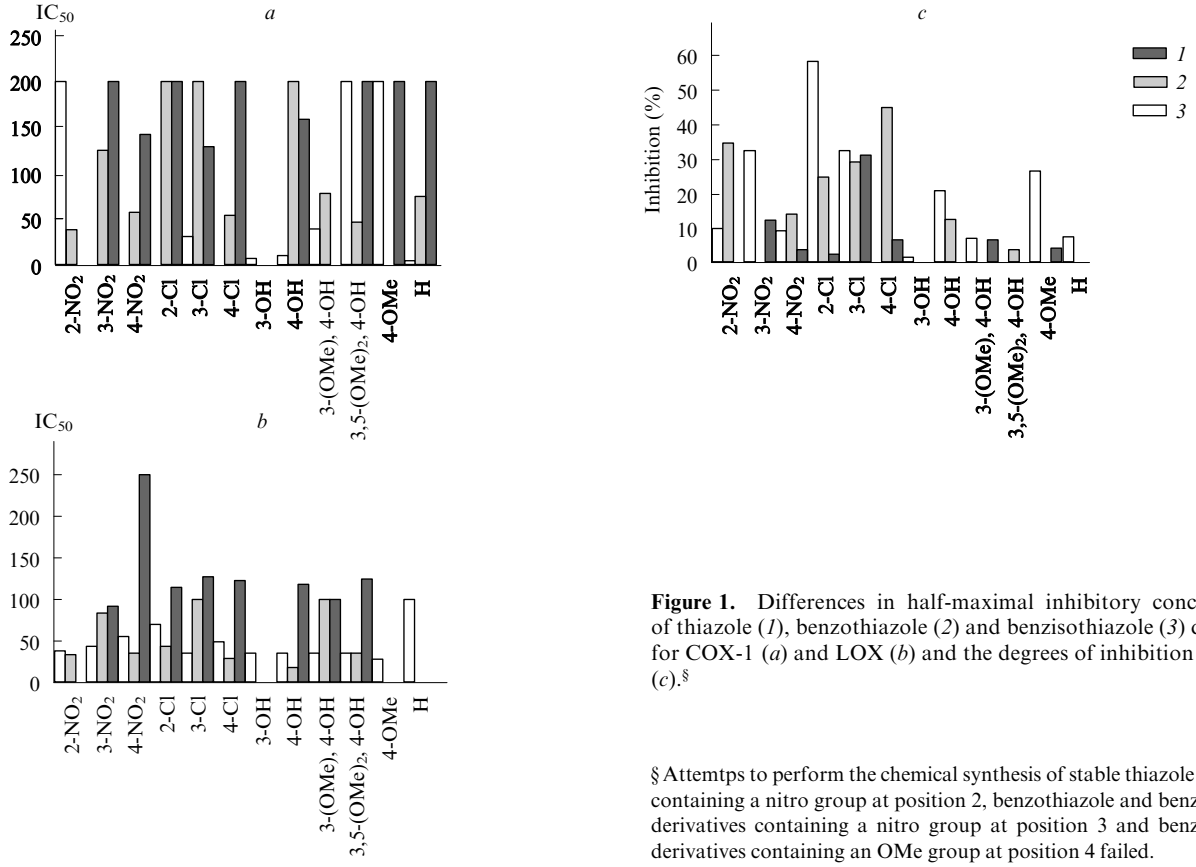


Figure 1. Differences in half-maximal inhibitory concentrations of thiazole (*1*), benzothiazole (*2*) and benzisothiazole (*3*) derivatives for COX-1 (*a*) and LOX (*b*) and the degrees of inhibition of COX-2 (*c*).§

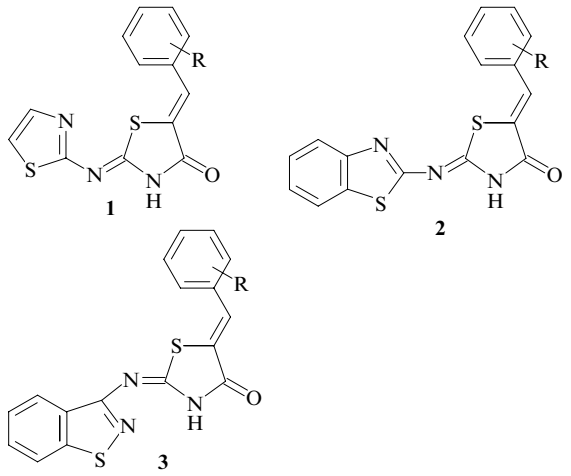
§ Attemtps to perform the chemical synthesis of stable thiazole derivatives containing a nitro group at position 2, benzothiazole and benzisothiazole derivatives containing a nitro group at position 3 and benzisothiazole derivatives containing an OMe group at position 4 failed.

account the *a priori* estimates of the contribution of the fragment as a whole to biological activity.

In the course of the generation, all possible combinations are created by connecting fragments at the vertices, which are involved in the bond cleavage. A new bond can appear if the code of the atom involved in its formation is identical to the code specified for the bond cleavage. It would be expected that these constraints will lead to the generation of synthetically feasible structures. In each step of the generation, structures corresponding to the desired types of biological activity can be selected.

It was shown that the developed generation algorithm ensures the formation of real structures.[141] However, to increase the feasibility of the synthesis of chemical compounds, it is necessary to develop additional criteria for rejection of structures, which seem to be synthetically inacessible.

Reliable differences in the occurrence frequency of fragments selected based on the estimation of their contributions to the inhibition of the angiotensin-converting enzyme (ACE) were observed for compounds, which have and do not have inhibitory properties against ACE.[141] Fragments selected based on the estimation of their contributions to the inhibition of cyclooxygenase-1 (COX-1), cyclooxygenase-2 (COX-2) and lipoxygenase (LOX) (Table 3), such as thiazole (**1**), benzothiazole (**2**) and benzisothiazole (**3**) derivatives, were shown to increase the affinity of inhibitors for these enzymes (Fig. 1).[141]



The proposed approach to the fragment library generation based on the estimation of the contributions to biological activity allows one to perform the multicriteria selection of the most promising candidates before the generation.

V. Prospects of the development of methods for the fragment-based lead design

The fragment-based *de novo* ligand design can be applied for the design of compounds that mimic functions of peptides involved in protein–protein interactions.[142–146] The latter can serve as the basis for the design of peptidomimetics. The problem of the construction of novel chemical compounds by modifying peptides has arisen from the difficulties encountered in developing peptidomimetics associated with the fact that natural peptides are rapidly cleaved by proteases. Besides, peptides have lower cellular membrane permeability and lower bioavailability as compared to drug molecules and they are rapidly metabolized.[147]

Consequently, there is a need to discover non-peptidic compounds capable of interacting with a particular target protein.

In some cases,[148] compounds that mimic functions of protein fragments in protein–protein interaction sites consist of several amino acids, in which peptide (amide) bonds are replaced by non-peptide bonds and which fulfill the role of a biological substrate or inhibitor. Stability of oligopeptides can be increased by introducing D-amino acid residues into the structure instead of L-amino acid residues.[149]

Besids, approaches were developed for the design of new scaffolds to replace the peptide backbone. Robinson[150] proposed that the replacement of the peptide backbone be used for the construction of ligands that mimic functions of peptides in protein interaction sites or fulfill the function of a protein inhibitor.

It is noteworthy that the GrowMol programme[119] for the fragment-based lead design was applied for the design of peptidomimetics, resulting in the discovery of piperidine derivatives of pepsin inhibitors. Although the possibility of designing novel non-peptide inhibitors of renin and pepsin has been shown earlier, the first step in the design with the use of the GrowMol programme involved the selection of known structures, for which the position of the ligand in the active site of the protein was determined by X-ray crystallography, *i.e.*, experimental data on the protein–ligand binding were required to tackle the problem. In addition, in most cases the structure generation resulted in the discovery of novel analogues of piperidine derivatives with inhibiotry properties against renin.[147]

It was shown[151] that biological activity of peptides and peptidomimetics can be predicted with the PASS computer programme. The prediction made by PASS can be used for the multicriteria optimization of the designed compounds (for example, high probability of the presence of desired types of activity and low probability of unwanted side effects and toxicity).

Nowadays, the design of fragment libraries containing fragments capable of interacting with a number of proteins is used less frequently for the discovery and design of peptidomimetics compared to its application for the construction of drug-like organic compounds. However, it would be expected that the targeted design of compounds that mimic functions of peptides in protein–protein interaction sites with the use of fragment libraries will be efficient.

The application of the fragment-based drug design using X-ray crystallography screening was exemplified by the discovery of a series of antitumour agents, in particular, of aurora kinase inhibitors (first- and second-stage clinical trials).[152]

NMR spectroscopy was applied for the design of inhibitors of the proteins Bcl-Xl and Abt-263 (second-stage clinical trials).[153]

The molecular structures of several compounds, which are in the second stage of clinical trials and are considered as drug candidates inhibiting the heat shock protein HSP90, were obtained by the integration of experimental and theoretical methods for the identification of fragments capable of interacting with a target protein.[154] Let us also mention inhibitors of the peroxisome-dependent receptor[154] and the blood coagulation factor Xa (second-stage

clinical trials).[155] Besides, dozens of molecules are in the pre-clinical trial stage.[61]

The rapid development of different methods applied in different steps of the fragment-based molecular design gives hope for the discovery of novel drugs, including multitarget agents.

It should be noted that the discovery of drugs, which are characterized by considerable chemical novelty and whose synthesis will be time- and labour-saving, requires the development of new theoretical approaches in the fragment-based molecular design.[156] One of the perspective directions is the development of original methods for the fragment library design and their application for the molecular drug design and the elaboration of efficient algorithms for the design of novel structures with the use of available libraries or by combining both approaches for the solutions of problems of the multicriteria selection.

References

1. A D Gorse *Curr. Top. Med. Chem.* **6** 3 (2006)
2. T Fink, J L Reymond *J. Chem. Inf. Model.***47** 342 (2007)
3. J L Reymond, R van Deursen, L C Blum, L Ruddigkeit *Med. Chem. Commun.* **1** 30 (2010)
4. J I Crowely, H Rapoport *Acc. Chem. Res.* **9**135 (1976)
5. G Barany, M Kempe, in *A Practical Guide to Combinatorial Chemistry* (Eds A W Czarnik, S H DeWitt) (New York: American Chemical Society, 1997) p. 51
6. K C Nikolaou, J Pastor, N Winssinger, F Murphy *J. Am. Chem. Soc.* **120** 5132 (1998)
7. F Guillier, D Orain, M Bradley *Chem. Rev.* **100** 2091 (2000)
8. S Kobayashi, R Akiyama *Pure Appl. Chem.* **73** 1103 (2001)
9. D G Hall, S Manku, F J Wang *J. Comb. Chem.* **3** 125 (2001)
10. Ya A Ivanenkov, E V Bovina, K V Balakin *Russ. Chem. Rev.* **78** 465 (2009) [*Usp. Khim.* **78** 503 (2009)]
11. R A Lewis *J. Comput. Aid. Mol. Des.* **4** 205 (1990)
12. Y Nishibata, A Itai *Tetrahedron* **47** 8985 (1991)
13. Y Nishibata, A Itai *J. Med. Chem.* **36** 2921 (1993)
14. S H Rotstein, M A Murcko *J. Comput. Aid. Mol. Des.* **7** 23 (1993)
15. D A Pearlman, M A Murcko *J. Comput. Chem.* **14** 1184 (1993)
16. Z W Luo, R X Wang, L H Lai *J. Chem. Inf. Comput. Sci.* **36** 1187 (1996)
17. A Miranker, M Karplus *Proteins* **23** 472 (1995)
18. C M Stultz, M Karplus *Proteins* **40** 258 (2000)
19. P S Kutchukian, E I Shakhnovich *Expert Opin. Drug Discov.* **5** 789 (2010)
20. I P Bangov, in *Handbook of Chemoinformatics* (Ed. J Gasteiger) (Weinheim: Wiley, 2003) p.178.
21. R A Carr, M Congreve, C W Murray, D C Rees *Drug Discov. Today* **10** 987 (2005)
22. H van de Waterbeemd, E Gifford *Nat. Rev. Drug Discov.* **2** 192 (2003)
23. C R Hopkins, A A Groom *Drug Discov. Today* **15** 430 (2004)
24. I D Kuntz, K Chen, K A Sharp, P A Kollman *Proc. Natl. Acad. Sci. USA* **96** 9997 (1999)
25. M L Verdonk, D C Rees *Chem. Med. Chem.* **3** 1179 (2008)
26. K Boda, T Seidel, J Gasteiger *J. Comput. Aid. Mol. Des.* **21** 311 (2007)
27. M H Todd *Chem. Soc. Rev.* **34** 247 (2005)
28. J Gasteiger, W D Ihlenfeldt, P Röse, R Wanke *Anal. Chim. Acta* **235** 65 (1990)
29. N S Zefirov, E V Gordeeva *Russ. Chem. Rev.* **56** 1002 (1987) [*Usp. Khim.* **56** 1753 (1987)]
30. E J Corey, T P Loh, T D Poper, M D Azimioara, M C Noe *J. Am. Chem. Soc.* **114** 8290 (1992)
31. E J Corey, A K Long, S D Rubenstein *Science* **228** 408 (1985)
32. H M Vinkers, M R de Jonge, F F D Daeyaert, J Heeres, L M H Koymans, J H van Lenthe, P J Lewi, H Timmerman, K Van Aken, P A J Janssen *J. Med. Chem.* **46** 2765 (2003)
33. M M Marsili, J Gasteiger, R E Carter *Chim. Oggi* **9** 1 (1984)
34. R Doenges, B Groebel, H Nickelsen, J Sander *J. Chem. Inf. Somrut. Sci.* **25** 42 (1985)
35. E Proschak, Y Tanrikulu, G Schneider, in *Chemoinformatics Approaches to Virtual Screening* (Eds A Varnek, A Tropsha) (Cambridge: Royal Society of Chemistry, 2008) p. 217
36. X Q Lewell, D B Judd, S P Watson, M M Hann *J. Chem. Inf. Comput. Sci.* **38** 511 (1998)
37. A C Pierce, G Rao, G W Bemis *J. Med. Chem.* **47** 2768 (2004)
38. P Schneider, G Schneider *QSAR Comb. Sci.* **22** 713 (2003)
39. U Fechner, G Schneider *J. Chem. Inf. Model.* **47** 656 (2007)
40. V J Gillet, W Newell, P Mata, G Myatt, S Sike, Z Zsoldos, A P Johnson *J. Chem. Inf. Comput. Sci.* **34** 207 (1994)
41. V J Gillet, G Myatt, Z Zsoldos, A P Johnson *Persp. Drug Discov. Des.* **3** 34 (1995)
42. H J Böhm *J. Comput. Aid. Mol. Des.* **8** 243 (1994)
43. D Douguet, H Munier-Lehmann, G Labesse, S Pochet *J. Med. Chem.* **48** 2457 (2005)
44. R Wang, Y Gao, L Lai *J. Mol. Model.* **6** 498 (2000)
45. H Zhao *Drug. Discov. Today* **12** 149 (2007)
46. W L Jorgensen *Acc. Chem. Res.* **42** 724 (2009)
47. G Schneider, M L Lee, M Stahl, P Schneider *J. Comput. Aid. Mol. Des.* **14** 487 (2000)
48. J Degen, M Rarey *Chem. Med. Chem.* **1** 854 (2006)
49. C A Nicolaou, J Apostolakis, C S Pattichis *J. Chem. Inf. Model.* **49** 295 (2009)
50. N P Todorov, P M Dean *J. Comput. Aid. Mol. Des.* **11** 175 (1997)
51. C M W Ho, in *Chemoinformatics in Drug Discovery* (Ed. I Oprea) (Weinheim: Wiley, 2005) p. 199
52. W R Moore *Curr. Opin. Drug Discov. Devel.* **8** 355 (2005)
53. J D Durrant, R E Amaro, J A McCammon *Chem. Biol. Drug Des.* **73** 168 (2009)
54. E M Duffy, W L Jorgensen *J. Am. Chem. Soc.* **122** 2878 (2000)
55. M Hornig, A Klamt *J. Chem. Inf. Model.* **45** 1169 (2005)
56. E Proschak, M Rupp, S Derksen, G Schneider *J. Comput. Chem.* **29** 108 (2008)
57. P J Hajduk, J R Huth, S W Fesik *J. Med. Chem.* **48** 2518 (2005)
58. P J Hajduk, J R Huth, S W Fesik *Mol. Int.* **6** 266 (2006)
59. M Fischer, R E Hubbard *Mol. Int.* **9** 22 (2009)
60. R Gozalbes, R J Carbajo, A Pineda-Lucena *Curr. Med. Chem.* **17** 1769 (2010)
61. G Chessari, A J Woodhead *Drug Discov. Today* **14** 668 (2009)
62. Yu Chen, B K Shoichet *Nature Chem. Biol.* **5** 358 (2009)
63. E R Zatler, H Mo *Curr. Top. Med. Chem.* **7** 1592 (2007)
64. P Cimmperman, L Baranauskiene, S Jachimovičiute, J Jachno, J Torresan, V Michailoviene, J Matuliene, J Sereikaite, V Bumelis, D Matulis *Biophys. J.* **95** 3222 (2008)
65. M C Lo, A Aulabaugh, G Jin, R Cowling, J Bard, M Malamas, G Ellestad *Anal. Biochem.* **332** 153 (2004)
66. G Rigaut, A Shevchenko, B Rutz, M Wilm, M Mann, B Séraphin *Nature Biotechnol.* **17** 1030 (1999)
67. A A Terent'ev, N T Moldogazieva, K V Shaitan *Usp. Biol. Khim.* **49** 429 (2009) [a]
68. A-C Gavin, M Bösche, R Krause, P Grandi, M Marzioch, A Bauer, J Shultz, J M Rick, A-M Michon, C-M Cruciat, M Remor, C Höfert, M Schelder, M Brajenovic, H Ruffner, A Merino, K Klein, M Hudak, D Dickson, T Rudi, V Gnau, A Bauch, S Bastuck,B Huhse, C Leutwein, M-A Heurtier, R R Copley, A Edelmann, E Querfurth, V Rybin, G Drewes, M Raida, T Boumeester, P Bork,B Seraphin, B Kuster, G Neubauer, G Superti-Furga *Nature (London)* **415** 141 (2002)

69. T Bürckstümmer, K L Bennett, A Preradovic, G Schütze, O Hantschel, G Superti-Furga, A Bauch *Nature Methods* **3** 1013 (2006)
70. B Ganem, Y T Li, J D Henion *J. Am. Chem. Soc.* **113** 6294 (1991)
71. S A Hofstadler, R H Griffey *Chem. Rev.* **101** 377 (2001)
72. A S Woods, J S Buchsbaum, T A Worrall, J M Berg, R J Cotter *Anal. Chem.* **67** 4462 (1995)
73. D A Erlanson, A C Braisted, D R Raphael, M Randal, R M Stroud, E M Gordon, J A Wells *Proc. Natl. Acad. Sci. USA* **97** 9367 (2000)
74. D A Erlanson, J A Wells, A C Braisted *Annu. Rev. Biophys. Biomol. Struct.* **33** 199 (2004)
75. P Schuk *Annu. Rev. Biophys. Biomol. Struct.* **26** 541 (1997)
76. A Kuglstatter, M Stahl, J-U Peters, W Huber, M Stihle, D Schlatter, J Benz, A Ruf, D Roth, T Enderle, M Hennig *Bioorg. Med. Chem. Lett.* **18** 1304 (2008)
77. T Neumann, H-D Junker, K Schmidt, R Sekul *Curr. Top. Med. Chem.* **7** 1630 (2007)
78. S S Antonysamy, B Aubol, J Blaney, M F Browner, A M Gianetti, S F Harris, N Hébert, J Hendle, S Hopkins, E Jefferson, C Kissinger, V Leveque, D Marchiano, E McGee, I Najera, B Nolan, M Tomimoto, E Torres, T Wright *Bioorg. Med. Chem. Lett.* **18** 2990 (2008)
79. A M Giannetti, B D Koch, M F Browner *J. Med. Chem.* **51** 574 (2008)
80. H Jhoti, A Cleasby, M Verdonk, G Williams *Curr. Opin. Chem. Biol.* **11** 485 (2007)
81. J P Overington, B Al-Lazikani, A L Hopkins *Nature Rev. Drug Discov.* **5** 993 (2006)
82. N Howard, C Abell, W Blakemore, G Chessari, M Congreve, S Hovard, H Jhoti, C W Murray, L C A Seavers, B L M van Montfort *J. Med. Chem.* **49** 1346 (2006)
83. P G Wyatt, A J Woodhead, V Berdini, J A Boulsridge, M G Carr, D M Cross, D J Davis, L A Devine, T R Early, R E Feltell, E J Lewis, R L McMenamin, E F Navarro, M A O'Brien, M O'Reilly, M Reule, G Saxty, L C A Seavers, D-M Smith, M S Squires, G Trewartha, M T Walker, A J A Woolford *J. Med. Chem.* **51** 4986 (2008)
84. J Tsai, J T Lee, W Wang, Y Zhang, H Cho, S Mamo, R Bremer, S Gillette, J Kong, N K Haass, K Sproesser, L Li, K S M Smalley, D Fong, Y-L Zhu, A Marimuthu, H Nguyen, B Lam, J Liu, I Cheung, J Rice, Y Suzuki, C Luu, C Settachatgul, R Shellooe, J Cantwell, S H Kim, J Schlessinger, K Y J Zhang, B L West, B Powell, G Habets, C Zhang, P N Ibrahim, P Hirth, D R Artis, M Herlyn, G Bollag *Proc. Natl. Acad. Sci. USA* **105** 3041 (2008)
85. G L Card, L Blasdel, B P England, C Zhang, Y Suzuki, S Gillette, D Fong, P N Ibrahim, D R Artis, G Bollag, M V Milburn, S-H Kim, J Schlessinger, K Y J Zhang *Nature Biotechnol.* **23** 201 (2005)
86. T Oltersdorf, S W Elmore, A R Shoemaker, R C Armstrong, D J Augeri, B A Belli, M Bruncko, T L Deckwerth, J Dinges, P J Hajduk, M K Joseph, S Kitada, S J Korsmeyer, A R Kunzer, A Letai, C Li, M J Mitten, D G Nettesheim, S Ng, P M Nimmer, J M O'Connor, A Oleksijew, A M Petros, J C Reed, W Shen, S K Tahir, C B Thompson, K J Tomaselli, B Wang, M D Wendt, H Zhang, S W Fesik, S H Rosenberg *Nature (London)* **435** 677 (2005)
87. M Mayer, B Meyer *Angew. Chem., Int. Ed.* **38** 1784 (1999)
88. S B Shuker, P J Hajduk, R P Meadows, S W Fesik *Science* **274** 1531 (1996)
89. K L Damm, H A Carlson *J. Am. Chem. Soc.* **129** 8225 (2007)
90. G M Morris, D S Goodsell, R S Halliday, R Huey, W E Hart, R K Belew, A J Olson *J. Comput. Chem.* **19** 1639 (1998)
91. R Huey, G M Morris, A J Olson, D S Goodsell *J. Comput. Chem.* **28** 1145 (2007)
92. C McMartin, R S Bohacek *J. Comput. Aid. Mol. Des.* **11** 333 (1997)
93. R Abagyan, M Totrov, D Kuznetsov *J. Comput. Chem.* **15** 488 (1994)
94. C N Cavasotto, R A Abagyan *J. Mol. Biol.* **337** 209 (2004)
95. G Jones, P Willett, R C Glen, A R Leach, R Taylor *J. Mol. Biol.* **267** 727 (1997)
96. M Rarey, B Kramer, T Lengauer, G Klebe *J. Mol. Biol.* **261** 470 (1996)
97. U Fechner, G Schneider *J. Chem. Inf. Model.* **46** 699 (2006)
98. B W Dymock, X Barril, P A Brough, J E Cansfield, A Massey, E McDonald, R E Hubbard, A Surgenor, S D Roughley, P Webb, P Workman, L Wright, M J Drysdale *J. Med. Chem.* **48** 4212 (2005)
99. P A Brough, W Aherne, X Barril, J Borgognoni, K Boxall, J E Cansfield, K M Cheung, I Collins, N G Davies, M J Drysdale, B Dymock, S A Eccles, H Finch, A Fink, A Hayes, R Howes, R E Hubbard, K James, A M Jordan, A Lockie, V Martins, A Massey, T P Matthews, E McDonald, C J Northfield, L H Pearl, C Prodromou, S Ray, F I Raynaud, S D Roughley, S Y Sharp, A Surgenor, D L Walmsley, P Webb, M Wood, P Workman, L Wright *J. Med. Chem.* **51** 196 (2008)
100. G L Warren, C W Andrews, A M Capelli, B Clarke, J LaLonde, M H Lambert, M Lindvall, N Nevins, S F Semus, S Senger, G Tedesco, I D Wall, J M Woolven, C E Peishoff, M S Head *J. Med. Chem.* **49** 5912 (2006)
101. R E Hubbard, I Chen, B Davis *Curr. Opin. Drug Discov. Devel.* **10** 289 (2007)
102. H-D Holtje, W Sippl, D Rognan, G Folkers *Molecular Modeling: Basics Principles and Applications* (New York: Wiley, 2008)
103. M Hohwy, L Spadola, B Lundquist, P Hawtin, J Dahmen, I Groth-Clausen, E Nilsson, S Persdotter, K von Wachenfeldt, R H Folmer, K Edman *J. Med. Chem.* **51** 2178 (2008)
104. T Lubbers, P Angehrn, H Gmunder, S Herzig *Bioorg. Med. Chem. Lett.* **17** 4708 (2007)
105. T Langer *Mol. Inf.* **29** 470 (2010)
106. S M Free, J W Wilson *J. Med. Chem.* **7** 395 (1964)
107. H Kubinyi, in *Methods and Principles in Medicinal Chemistry* Vol. 1 (Eds R Mannhold, P Kroogsgard-Larsen, H Timmerman) (Weinheim: VCH, 1993)
108. T Aoyama, H Ichikawa *Chem. Pharm. Bull.* **39** 1222 (1991)
109. O A Raevsky *Ros. Khim. Zh.* **50** (2) 97 (2006 [b]
110. Q S Du, R B Huang, Y T Wei, Z W Pang, L Q Du, K C Chou *J. Comput. Chem.* **30** 295 (2009)
111. B M Spowage, C L Bruce, J D Hirst *J. Chem. Inf.* **1** 22 (2009)
112. J R Fischer, U Lessel, M Rarey *J. Chem. Inf. Model.* **50** 1 (2010)
113. B C Pearce, D R Langley, J Kang, H Huang, A Kulkarni *J. Chem. Inf. Model.* **49** 1797 (2009)
114. B Nisius, U Rester *J. Chem. Inf. Model.* **49** 1211 (2009)
115. R S DeWitte, E I Shakhnovich *J. Am. Chem. Soc.* **118** 11733 (1996)
116. G Schneider, U Fechner *Nature Rev. Drug Discov.* **4** 649 (2005)
117. H Mauser, W Guba *Curr. Opin. Drug Discov. Devel.* **11** 365 (2008)
118. D K Gehlhaar, K E Moerder, D Zichi *J. Med. Chem.* **38** 466 (1995)
119. R S Bohacek, C McMartin *J. Am. Chem. Soc.* **116** 5560 (1994)
120. S H Rotstein, M A Murcko *J. Med. Chem.* **36** 1700 (1993)
121. D C Roe, I D Kuntz *J. Comput. Aid. Mol. Des.* **9** 269 (1995)
122. H Mauser, M Stahl *J. Chem. Inf. Model.* **47** 318 (2007)
123. P S Kutchukian, D Lou, E I Shakhnovich *J. Chem. Inf. Model.* **49** 1630 (2009)

124. F Dey, A Caflisch *J. Chem. Inf. Model.* **48** 679 (2008)
125. E W Lameijer, J N Kok, T Back, A P Ijzerman *J. Chem. Inf. Model.* **46** 545 (2006)
126. E W Lameijer, T Back, J N Kok, A P Ijzerman *Nat. Comput.* **4** 177 (2005)
127. V J Gillet, in *Evolutionary Algorithms in Molecular Design* (Ed. D E Clark) (Weinheim: Wiley-VCH, 2000) p. 49
128. S Kamphausen, N Holtge, F Wirsching, C Morys-Wortmann, D Riester, R Goetz, M Thürk, A Schwienhorst *J. Comput. Aid. Mol. Des.* **16** 551 (2002)
129. N Brown, B McKay, F Gilardoni, J Gasteiger *J. Chem. Inf. Comput. Sci.* **44** 1079 (2004)
130. M E Munk, C S Sodano, R L McLean, T H Haskell *J. Am. Chem. Soc.* **89** 4158 (1967)
131. S B Zotchev, A V Stepanchikova, A P Sergeyko, B N Sobolev, D A Filimonov, V V Poroikov *J. Med. Chem.* **49** 2077 (2006)
132. A P Sergeiko, Candidate Thesis in Biological Sciences, V N Orekhovich Research Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, Moscow, 2008
133. H Patel, M J Bodkin, B Chen, V J Gillet *J. Chem. Inf. Model.* **49** 1163 (2009)
134. US P. 0182094 A1 (2003)
135. S S Tratch, O A Lomova, D V Sukhachev, V A Palyulin, N S Zefirov *J. Chem. Inf. Comput. Sci.* **32** 130 (1992)
136. A A Melnikov, V A Palyulin, N S Zefirov *J. Chem. Inf. Model.* **47** 2077 (2007)
137. L Pauling *The Nature of the Chemical Bond and the Structure of Molecules and Crystals* (Intaca: Cornell University Press, 1960)
138. J A Beran *Laboratory Manual for Principles of General Chemistry* (Hoboken, NJ: Wiley, 2009)
139. D A Filimonov, V V Poroikov, in *Chemoinformatics Approaches to Virtual Screening* (Eds A Varnek, A Tropsha) (Cambridge: Royal Society of Chemistry, 2008) p. 182
140. A Geronikaki, P Eleftheriou, D Hadjipavlou-Litina, P Vicini, O Filz, D Filimonov, V Poroikov *Eur. J. Med. Chem.* **47** 111 (2011)
141. O A Filz, A A Lagunin, D A Filimonov, V V Poroikov, in *The 5th International Conference 'Genomics, Proteomics, Bioinformatics and Nanobiotechnologies for Medicine', Novosibirsk, 2010* p. 82
142. E Ko, J Liu, L Perez, G Lu, A Schaefer, K Burgess *J. Am. Chem. Soc.* **133** 462 (2011)
143. T T Tran, C Kulis, S M Long , D Bryant, P Adams, M L Smythe *J. Comput. Aid. Mol. Des.* **24** 917 (2010)
144. A Isidro-Llobet, T Murillo, P Bello, A Cilibrizzi, J T Hodgkinson, W R J D Galloway, A Bender, M Welch, D R Spring *Proc. Natl. Acad. Sci. USA* **26** 6793 (2011)
145. P Ung, D A Winkler *J. Med. Chem.* **54** 1111 (2011)
146. A V Veselovsky, A I Archakov *Curr. Comput. Aided. Drug. Des.* **3** 51 (2007)
147. M G Bursavich, D H Rich *J. Med. Chem.* **45** 31 (2002)
148. P A Hart, D H Rich, in *The Practice of Medicinal Chemistry* (Ed. C.Wermuth) (London: Academic Press, 1996) p. 393
149. I A Zamulaeva, A S Saenko, O V Semina, T N Semenets, V I Deigin, E I Celivanova, Ya V Malyutina *Byul. Eksp. Biol. Med.* **2** 214 (2006) [c]
150. J A Robinson *Acc. Chem. Res.* **41** 1278 (2008)
151. H B Martynova, D A Filimonov, V V Poroikov *Bioorg. Khim.* **26** 330 (2000) [d]
152. S Howard, V Berdini, J A Boulstridge, M G Carr, D M Cross, J Curry, L A Devine, T R Early, L Fazal, A L Gill, M Heathcote, S Maman, J E Matthews, R L McMenamin, E F Navarro, M A O' Brien, M O' Reilly, D C Rees, M Reule, D Tisi, G Williams, M Vinkovi, P G Wyatt *J. Med. Chem.* **52** 379 (2009)
153. C-M Park, M Bruncko, J Adickes, J Bauch, H Ding, A Kunzer, K C Marsh, P Nimmer, A R Shoemaker, X Song, S K Tahir, C Tse, X Wang, M D Wendt, X Yang, H Zhang, S W Fesik, S H Rosenberg, S W Elmore *J. Med. Chem.* **51** 6902 (2008)
154. D R Artis, J J Lin, C Zhang, W Wang, U Mehra, M Perreault, D Erbe, H I Krupka, B P England, J Arnold, A N Plotnikov, A Marimuthu, H Nguyen, S Will, M Signaevsky, J Kral, J Cantwell, C Settachatgull, D S Yan, D Fong, A Oh, S Shi, P Womack, B Powell, G Habets, B L West, K Y J Zhang, M V Milburn, G P Vlasuk, K P Hirth, K Nolop, G Bollaga, P N Ibrahim, J F Tobin *Proc. Natl. Acad. Sci. USA* **106** 262 (2009)
155. G Agnelli, S Haas, J S Ginsberg, K A Krueger, A Dmitrienko, J T Brandt *J. Thromb. Haemost.* **5** 746 (2007)
156. P J Hajduk, W R J D Galloway, D R Spring *Nature (London)* **470** 42 (2011)

[a] — *Biol. Chem. Rev. (Engl. Transl.)*
[b] — *Mendeleev Chem. J. (Engl. Transl.)*
[c] — *Bull. Exp. Biol. Med. (Engl. Transl.)*
[d] — *Russ. J. Bioorg. Chem. (Engl. Transl.)*