https://doi.org/10.1070/RCR5012

Preliminary modelling as the first stage of targeted organic synthesis

Ilya Yu. Titov,[a, b] Viktor S. Stroylov,[a, b] PolinaV. Rusina,[a, c] Igor V. Svitanko[a, b]*

[a] *N.D.Zelinsky Institute of Organic Synthesis, Russian Academy of Sciences, Leninsky prosp. 47, 119991 Moscow, Russian Federation*
[b] *National Research University 'Higher School of Economics' ul. Myasnitskaya 20, 101000 Moscow, Russian Federation*
[c] *Faculty of Chemistry, Lomonosov Moscow State University Leninskie Gory 1, stroenie 3, 119991 Moscow, Russian Federation*

The review aims to present a classification and applicability analysis of methods for preliminary molecular modelling for targeted organic, catalytic and biocatalytic synthesis. The following three main approaches are considered as a primary classification of the methods: modelling of the target – ligand coordination without structural information on both the target and the resulting complex; calculations based on experimentally obtained structural information about the target; and dynamic simulation of the target – ligand complex and the reaction mechanism with calculation of the free energy of the reaction. The review is meant for synthetic chemists to be used as a guide for building an algorithm for preliminary modelling and synthesis of structures with specified properties.
The bibliography includes 353 references.

Contents

1. Introduction 832
2. General features of structure – property correlations 832
3. Stages of development and complication of molecular modelling methods 833
4. Estimation of quantitative structure – property relationships (QSAR) with unknown target (Stage 1) 835
   4.1. Selection of structures for QSAR calculations 835
   4.2. Molecular descriptors 836
      4.2.1. 3D descriptors reflecting the three-dimensional structure 836
      4.2.2. Electrostatic (coulomb) interactions as the base for 3D-QSAR modelling 836
      4.2.3. Mapping of the molecular electrostatic potential 837
   4.3. Modelling of the odour of fragrances by constructing MEP-based model 838
   4.4. Determination of the structure – property relationships in the series of bicyclic ureas (glycolurils) with neurotropic activity 839
5. Molecular docking and its extensions in the modelling of targeted synthesis (Stage 2) 839
   5.1.Docking and virtual screening for identification of hits 839
   5.2. Combined use of QSAR methods and molecular docking 842
   5.3. Overcoming the conceptual limitations of classical docking 843
      5.3.1. Flexible protein docking 843
      5.3.2. On-top docking 844
      5.3.3. Molecular dynamics 846
      5.3.4. Free energy perturbation method 846
6. Quantum chemical methods in the predictive modelling (Stage 3a) 851
   6.1. Modelling of the mechanism and products of catalytic (including biocatalytic) reactions 851
   6.2. Asymmetric synthesis 852
   6.3. Examples of modelling of natural catalysis mechanisms 854
7. Combination of computer simulation and physicochemical methods (Stage 3b) 856
8. Approaches to modelling of organic heterogeneous catalysis 857
   8.1. Calculation of the ligand – solid catalyst conformational model 857
   8.2. 3D printing as a method for structuring of catalysts for targeted synthesis 857
   8.3. Modelling of the active component by 3D printing 858
9. Conclusion 859
10. List of acronyms and designations 860

Received 1 December 2020

© 2021 Uspekhi Khimii, ZIOC RAS, Russian Academy of Sciences and IOP Publishing Limited

1. Introduction

Molecular modelling (MM) represents computational studies of structures and properties of molecular systems. This review presents the classification and retrospective analysis of the methods used for preliminary modelling of chemical structures for targeted organic (including catalytic) synthesis. The review covers the published sources where modelling is a stage preceding the synthesis of compounds with specified properties rather than a method to interpret the reaction that has already been carried out. In other words, we consider publications in which the choice of a desired structure is not limited to the preferences of a synthetic chemist or the availability of required chemicals. A part of the review is devoted to modelling of the reaction mechanisms, which makes it possible to determine the composition of reaction products before the synthesis has been carried out.

There are numerous published reviews on the methods of molecular design and on the objects of targeted synthesis.[1–6] Since we were aiming to avoid repetitions, we inevitably paid much attention to our own research in computer simulations. We also gave references to the papers published in recent months and not yet mentioned in other review publications.

Detailed description of theoretical methods and mathematical apparatus is beyond the scope of the review, because the information is meant for synthetic chemists, who understand the need for preliminary modelling in the targeted organic synthesis, but are unfamiliar with the complex algorithm of actions, and not for specialists in mathematical methods. The goal of the review is to survey methods of computer modelling in organic and pharmaceutical chemistry, so that researchers could choose the required protocol and correctly set the task for a computational chemist. In this way, it is possible to avoid hundreds of 'blind' syntheses.

The material is presented according to the following principles:

— the review is meant, first of all, for synthetic chemists and can be used as a guide to build the algorithm of preliminary modelling and synthesis of structures with specified properties;

— the review presents a number of examples of studies not yet included in the existing reviews addressing drug development, simulation of chemical reaction mechanisms and use of computation results to control chemical reactions;

— the review is not focused on only the biological activity of chemical compounds, but on the synthesis *per se* in which the predictive modelling can be used to determine the structure and activity of a not yet synthesized final product. The main body of the review is devoted to catalytic synthesis, which corresponds most closely to the topic 'formation and degradation of the ligand – target complex'. This part also includes biocatalysis, the elementary event of which is coordination of a drug candidate to a therapeutic target.

2. General features of structure – property correlations

For any science, a correlation of structural elements repeated in space or in time with a property of the system comprising these elements leads, after analytical treatment and conclusions, to correct planning of research or practical actions. A set of robust correlations can be used to formulate the general theory of quantitative structure – activity relationships (QSAR).

Synthetic chemistry is not an exception. Everything that is done in the field of molecular modelling can have only one application in the synthetic chemistry: transition from extensive to intensive data accumulation, that is, switching from a multitude of syntheses with the hope to obtain a target compound with desired properties to the pinpoint synthesis of the compound that was obtained as a molecular model *via* predictive modelling.

Despite the obvious progress in the molecular modelling, rational planning of studies in the field of organic synthesis is still a fundamental problem to be solved. This is due, first, to the belief of synthetic chemists in their own intuition based on practical experience and knowledge of the literature and, second, to their long-standing area of research (studying of preferable classes of compounds), and even to the inventory of available chemicals. As a result, only one out of $\sim$5000 biologically active compounds that have been synthesized is brought to *in vivo* assays, and only a small fraction of the tested compounds is promoted to clinical trials as drug candidates, which makes the whole process very expensive.[7]

Switching from targeted organic synthesis to the prediction of a desired property on a model structure followed by the synthesis of the model is based on the following necessary conditions:

— high accuracy and reliability of molecular modelling, *i.e.*, relatively high predictive power of the method;

— selectivity of the method to a particular type of activity, *i.e.*, a degree of ligand affinity for the selected target or efficiency of ligand activation by the reaction catalyst;

— complexity of the obtained model, since a model unstable in air or requiring a multistep synthesis may prove to be inapplicable because a targeted synthesis must be scalable;

— feedback, that is, the experimentally confirmed presence and degree of the specified type of activity of the target

I.Yu.Titov. Researcher at the Laboratory of Molecular Modelling and Targeted Synthesis, ZIOC RAS.
E-mail: il_tit@bk.ru
Current research interests: chemistry of natural compounds, chemistry of biologically active compounds, targeted synthesis.
V.S.Stroylov. Candidate of Chemical Sciences, Researcher at the Laboratory of Molecular Modelling and Targeted Synthesis, ZIOC RAS.
E-mail: victor.stroylov@gmail.com
Current research interests: molecular docking, chemistry of biologically active compounds.
P.V.Rusina. Student of the Faculty of Chemistry, Moscow State University.
E-mail: polina.rusina@chemistry.msu.ru
Current research interests: *in silico* drug design, selectivity prediction, development of anticancer therapy.
I.V.Svitanko. Doctor of Chemical Sciences, Head of the Laboratory of Molecular Modelling and Targeted Synthesis, ZIOC RAS, Head of the Base Department of Organic Chemistry of the Faculty of Chemistry – HSE University.
E-mail: svitanko@mail.ru
Current research interests: molecular docking, chemistry of biologically active compounds, organic synthesis.

Translation: Z.P.Svitanko

structures synthesized according to the computed model. Practical validation of the method is a necessary condition for any modelling, especially for targeted synthesis (unvalidated mathematical constructions and modelling performed to explain a synthesis that has been already carried out are beyond the scope of this review).

Predictive analysis of existing data and the buiding and synthesis of a model with desired characteristics would eliminate the following problems of the modern targeted synthesis:[8]

(1) necessity of broad 'blind experiment' (expensive synthesis of a multitude of structures that may be expected to possess the desired property). Even taking account of literature analysis and experience of a synthetic chemist, the accuracy of this experiment is very low;

(2) impossibility of *a priori* prediction of the selectivity and toxicity of the obtained products. For example, the experimental verification of the antibacterial activities of 142 thousand compounds synthesized at the Institute of Organic Chemistry, Russian Academy of Sciences, carried out at the University of Queensland,[9] showed the percentage of non-toxic compounds active against gram-positive bacteria to be 0.11% and that against gram-negative bacteria to be 0.07%. For a commercial library of compounds, the percentage of useful structures was even lower: 0.03 and 0.008%, respectively. In-depth analysis of publications and 'search by analogy' insignificantly increase these values;

(3) scaling of the experiment. It is known that only a few products of targeted synthesis would reach the pre-industrial or preclinical stage, and the percentage of such products out of the total number of synthesized compounds would be much lower for empirical synthesis than in the case of pinpoint synthesis;

(4) inefficient involvement of young researchers in the work only as operating staff until they have accumulated the required amount of synthetic chemist's intuition and have analyzed the critical mass of literature.

It is evident that the use of preliminary predictive modelling reduces the resource intensity of the experiment by a factor of several tens, with a considerable cost saving coming from testing of the activity of products. Furthermore, having gained a useful tool such as methods for building a model whose structure is unpredictable in advance, young researchers may escape from the routine topics pursued by their research team and from doctrines restricting the search for new non-infringing structures with a specified property. Of course, this does not mean that they no longer need synthetic experience and in-depth analysis of publications. Mathematical methods in chemistry represent nothing but optimization tools and do not claim to replace humans.

The validation of methods is illustrated by examples of their practical use for predicting the products of targeted synthesis both for medicinal chemistry (calculation of an elementary step of biocatalysis) and for catalytic synthesis as a whole.[10]

Thus, this review deals with methods of predictive modelling of biological activity and determination of reaction mechanisms and free energy of reactions, illustrated by selected examples of homogeneous and heterogeneous catalysis, *i.e.*, those fields of the synthesis in which ligand coordination to the target may give the idea about the ligand activity towards the target and on the course of the reaction and composition of (bio)catalytic reaction products.

3. Stages of development and complication of molecular modelling methods

Methods of molecular modelling can be classified according to increasing accuracy of mathematical (sufficient number of objects, data processing) and physical (spatial resolution, number of calculated coupled physical parameters) representation of molecular structures. It is clear that this sequence would also be chronological, reflecting the degree of development of computational methods. We will consider the following stages of complication of molecular modelling methods.

Stage 1: *no experimental data are available on the mechanism of the desired activity of components of the primary complex, that is, the active compound and the target*. The physical picture of the ligand–target binding can be conceived only theoretically, *i.e.*, it can be based on available empirical data on the activity of a certain group of ligands whose physicochemical parameters can be easily calculated.[11–14]

The information that can be derived from a set of structures that *a priori* possess the desired property is gained at this stage by statistical analysis. General characteristics that can be responsible for the given property (descriptors) are defined for a selection of structures. Descriptors can be represented by sets of structural details, the nature and arrangement of chemical bonds, molecular dimensions and other physical parameters, *e.g.*, the spatial configuration of the molecular electrostatic field.[9,15]

The basic issue that is debated in Stage 1 is how to choose the set of descriptors to describe a property on the basis of analysis of a training set comprising both active and inactive compounds. Correspondingly, methods used for the reliable selection of the most significant descriptors are also different.

After a coherent set of descriptors has been selected, they can be applied to new compounds with unknown activity that are not included in the training set. With some reservations, this result can be declared to be a statistical quasi-model of a property of the ligand.[16] The quasi-model can be either a physical object or a molecular graph.[17]

Thus, the construction of the model in Stage 1 includes statistical treatment of characteristics of ligands that are either active or not. In the general case, physical characteristics of the target are not regarded as investigation objects in this stage.

The methods for statistical treatment of the data on the training set of ligands, construction of the model of a property on this basis and application of the resulting model to a compound with an unknown property is commonly called QSAR/QSPR (quantitative structure–activity/quantitative structure–property relationships). Certainly, applying this term only to Stage 1 narrows down its essence, as all of the subsequent stages such as docking, quantum chemical calculations, correlation between spectral data and structural features are still the same QSAR.

The principles of QSAR molecular modeling are well illustrated in the presentation of A.Tropsha.[18]

Stage 2: *experimentally determined data on the structures of interacting substrates are avalable*, both for the active substance (ligand) and for the target (receptor). Using computational methods, it is possible to model and vividly demonstrate the physical picture of ligand–target interac-

tion (Stage 2, docking, substrate positioning into the target) [19, 20] and, after calculation of the binding energy, it can be used to quantitatively express the degree of ligand activity towards a particular target.

A basic disadvantage of the classical docking in Stage 2 is that the calculation usually concerns a system of two coordinating objects. However, since a real biological ligand – target system is located in a solution rather than in a vacuum, it is necessary to take into account all system components, as the solvent can significantly affect the ligand activity. Classical docking implicitly considers solvation, but this may be insufficient.

One more common drawback of both statistical methods of Stage 1 and calculation of the binding energy in Stage 2 is the static nature of both the ligand in the training set and the ligand – target complex. In the former case, it is necessary to either include compounds with a static 3D structure into the training set or take into account all possible conformations of each compound. In the latter case, docking actually sums up the energies of point interactions of the ligand and target functional groups in the static complex.[21]

To attain the optimal coordination (the maximum binding energy), it is necessary to check all possible configurations of a conformationally labile ligand, *i.e.* to generate a multitude of alternative poses (pose is the spatial arrangement of the ligand relative to the target) for calculating all possible ligand – target configurations. Considering the requirements to computational resources, this is a drawback equivalent to the inevitable 'swelling' of conventional QSAR datasets due to a large number of possible conformations of active and inactive structures.

In addition, any variation of docking is based on the known structure of the target active site. Meanwhile, if the target has been detected experimentally, but its active site has not been detected or is vaguely defined, or there are several potential active sites, docking gives ambiguous results.[22] Considering these factors, the predictive power of the classical docking can be increased using the following extensions:

(*a*) protocol to take account of the solvent effect and other induced external effects on the strength of the complex (Stage 2a),[23] which would inevitably increase the required computational power by increasing the number of objects of the system;

(*b*) method for determining the optimal active conformations of the ligand and the target in the coordinated state;[24] this function will be called conformation filtering (Stage 2b);

(*c*) method for solving the problem of the static nature of the ligand – target complex. Ligand coordination is a dynamic, equilibrium process; therefore, to claim that the model has been constructed, it is insufficient to merely calculate the energy for the complex formed by the most stable ligand conformation with the target. It is also necessary to take into account the dynamic dissociation of the complex under conditions as close as possible to real conditions.[25] Without considering the process dynamics, docking would remain a mere summation of the interaction energies of the pair of static structures;

(*d*) method for the dynamic search for the complementary active site of the target if it is vaguely defined or has low activity.

For solving problem *a* of increasing the number of objects in the calculation of a solvated complex, it is reasonable to calculate the total energy of the complex, without considering the number of objects in the complex, *i.e.*, to pass from arithmetic summation of particular interactions to integration of the system energy as a whole. The extension to the classical docking taking into account free energy perturbation (FEP) is a successful way of such integration,[26, 27] which is applicable not only to calculations of biological complexes.[28] The method is effective with any number of system components.

For solving problem *b*: if the extensive calculation method (checking of the ligand and target conformations) is rejected, it would also be possible to use an integral characteristics: the energy curve of the complex obtained by mutual dynamic repositioning of both components using the FEP method. By docking the labile structure of the ligand with either rigid or flexible structure of the target in the coordinates 'free energy – complex geometry – time', it is possible to find the local minima in this curve and thus derive a stable structure of the initial complex.[29] This approach is rather dynamic than static; hence, by adding calculation of the probability of complex decay into components to the energy characteristics, one can pass to variant *c*, that is, a real dynamic process. Recall that this does not concern the dynamics of two components, but concerns the dynamics of a solvated complex and takes account of other external impacts that could be considered by using FEP in the previous stage.

Problem *d* arises in those cases where docking does not give results or gives incorrect prediction with a known active site. This may happen if the active site is vaguely defined and the ligand – target coordination has a low binding energy, or it is inactive towards the given type of activity, or there are several such sites.[8] If the active ligands found empirically (in a blind experiment) still show binding to the given target, it is necessary to find other binding sites on the target surface. This cannot be done in a way other than testing the whole surface with the ligand with recording the binding energy curve. The testing algorithm is the key factor: if testing is done consistently or randomly, with the results of previous iterations being neglected in every next iteration, the search for the active sites is restricted by the required computational power.[22]

Stages 1 and 2 considered so far address the ligand – target coordination. The described methods are generally operable for identification of drug candidates or, for example, the structure of the ligand(s) – catalyst complex. Examples of using these methods in targeted synthesis are discussed in the subsequent Sections of the review.

If we deal with synthesis, the ligand coordination to the target is the first and important elementary step, but it is far from being always the rate-limiting step. Therefore, in the modelling for targeted synthesis (and not only for drug design), it is important to additionally determine the dynamics and the mechanism of the whole reaction giving the active product by calculating the possible transition states and composition of the products of synthesis. This requires methods of quantum chemistry. This level of calculations can be called Stage 3a.

Methodologically, it is important that, like in the case of FEP, this approach operates with integral energy parameters, *i.e.*, the number and nature of objects do not matter: this may be a drug – protein complex or a catalyst – reactant complex. The homogeneity of the reaction mixture also makes no difference. For example, if the

structure of a heterogeneous catalyst is simplified to a single-atom case, it is quite possible to determine and visualize the ligand position on this catalyst by quantum chemical methods.[30]

The computational methods are supplemented and validated by structural studies using physicochemical methods. This is why the consideration of spectroscopic data (Stage 3b) is a part of integrated analysis of the mechanism of synthesis. By summarizing some or all approaches to molecular modelling outlined above and using both computational methods and characteristics of real physical objects,[31, 32] one can obtain much more accurate results on the monitoring of chemical reactions,[32] control of the reaction selectivity [33] and affecting the reaction pathway,[34] *e.g.*, for adjustment (control) of the carbon catalyst structure in acetylene trimerization.[35]

The structural adjustment of the target attracts considerable interest. This does not concern adjustment of the structure of the naturally established target protein, but concerns the fabrication of the catalyst (both homogeneous and heterogeneous one) by physical methods, *e.g.*, by 3D printing of the catalyst with a desired activity.[36]

Consideration of the synthetic pathways to biologically active compounds and calculation of catalytic reaction pathways and catalyst activity within a single publication are based on a substantial similarity of algorithms and methods of modelling of the products of targeted synthesis. The goal of this consideration is to demonstrate that virtually any organic synthesis can be predicted in order to obtain practically useful compounds with a target property.

In the modelling of the ligand – target complex, a biologically active compound or a reactant of a catalytic reaction can serve as the ligand. The target can be represented as either a protein or a homogeneous or a heterogeneous catalyst. Physical laws and description of their effect on the chemical reaction are important for physical modelling. Certain corrections to this calculation are made by considering the nature of interacting substrates and by visualizing the interactions using physicochemical methods.

The following Sections consider molecular modelling methods in the order of increasing complexity and, hence, increasing accuracy of the prediction of properties of compounds.

4. Estimation of quantitative structure – property relationships (QSAR) with unknown target (Stage 1)

4.1. Selection of structures for QSAR calculations

After analysis of the published data on the activities of compounds exhibiting a certain property, one can select a group of structures to compose the QSAR training set. The training set should include the whole range of active structures varying from completely inactive to the most active ones. The result of QSAR prediction is checked by the validation set including compounds with known activity that are calculated in the same way as compounds of the training set. This set would characterize only the effect of activity rather than the activity mechanism. In the best case, the training set should include compounds that exhibit the desired activity ranging from zero (no activity) to unity (maximum activity), desirably with a dense coverage of this range. In this case, the statistically built model would show a probability scattering of the values. The degree of coincidence of a new unknown compound with the model can also be characterized rather accurately.[37]

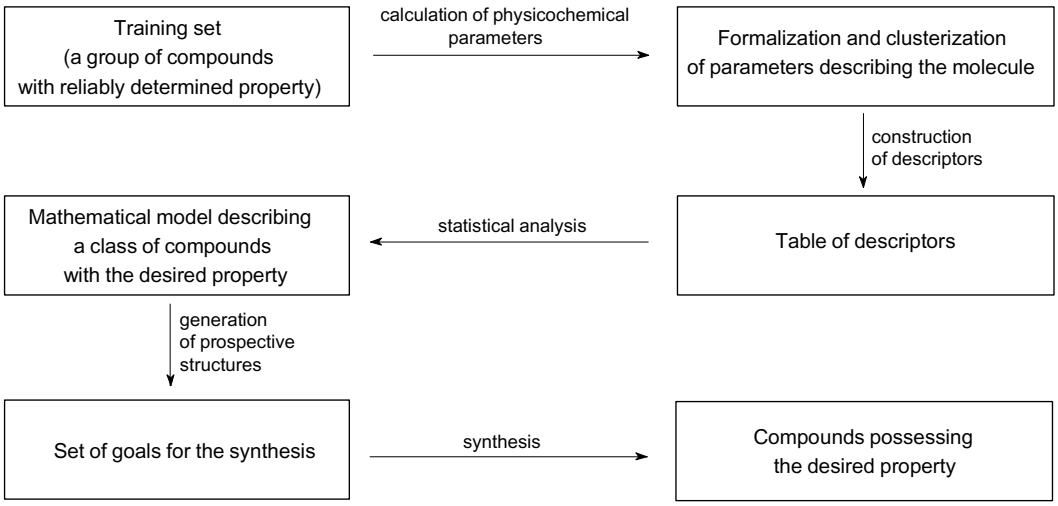
The next stage is to describe the training set compounds in such a way as to include physicochemical and other characteristics defining the model (descriptors). Thus, descriptors and characteristics that should be included for training set structures are related by a feedback. The descriptors are not necessarily a set of physical objects with spatial coordinates.

The set of descriptors (if it actually describes the required property that is determined in the validation experiment) is a discrete model of this property. Generation of structures corresponding to these descriptors gives model compounds for targeted synthesis.

Data on the biological or catalytic activity of compounds (in a broad meaning of these notions) are based on the measurements of the efficiency of interaction between the compound and its target. From this, it directly follows that in the absence of data on the activity mechanism, any activity models can be based only on the observed results of interaction and on various modelling approaches. Scheme 1 shows the algorithm of statistical predictive modelling for the targeted synthesis.

It is noteworthy that the adequacy of selection of the descriptors and structures for the training set can have a

Scheme 1



crucial effect on the result of modelling. In the case of biological activity modelling, the training set should meet the following criteria:[38]

— no chemical reactions take place upon interaction with the target;

— the number of structures in the training set is definitely greater than the number of degrees of freedom in the final model;

— accurate quantitative data on the compound activity are presented;

— known data on activity are separated for the training and validation sets;

— the sets must include both active and inactive compounds;

— structure flexibility leads to an increasing number of structures for calculations, since various conformations are also included in the sets.

4.2. Molecular descriptors

Mathematical, physical, physicochemical and chemical descriptors are derived from structural characteristics of the set, ranging from the molecular formula (so-called 1D), connectivity matrix of atoms in a two-dimensional (2D) structure to a three-dimensional, conformation-dependent (3D) one[39,40] and to higher-dimensional structures that consider the relative orientation and the temporal dynamics of molecules (4D and higher).[41] The descriptors thus formed are used in the modelling to predict the properties. The key factors are the analysis and final choice of descriptors significant for the property. Comprehensive analysis of molecular descriptors from the standpoint of applicability for modelling can be found in a number of reviews.[14,15,42]

4.2.1. 3D descriptors reflecting the three-dimensional structure

In terms of QSAR, it is stated that effective ligand binding is mainly determined by non-covalent interactions, *i.e.*, by induced effects of the fields generated by the ligands themselves or their environment. This statement has defined the initial form of 3D-QSAR as the comparative molecular field analysis (CoMFA).[39,43] This analysis uses a statistical approach called the partial least squares (PLS) of latent variables.[44]

The ideology of switching from 1D and 2D descriptions of molecules to three-dimensional (3D) characteristics in calculating the activity descriptors of any type was first proposed by N.S.Zefirov at the seminar on QSAR held at the N.D.Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, in September, 1989. His key statement (cited from memory) was as follows: 'It's time to move from the reference to atoms and topological indices to a molecule as a whole, that is, to pass from atomic and molecular connectivity matrices to real physical objects with real physical characteristics: geometry, molecular surface and the total electrostatic potential.' The date of this seminar can be taken as the birthday of 3D-QSAR in Russia.

The CoMSIA extension to 3D-QSAR[45] offers the possibility of including electrostatic interactions, adding hydrogen bonds to non-covalent effects and taking account of molecule hydrophobicity. There are also methods that restrict the amount of calculations of the possible conformations of labile ligands: GRID[46] belonging to the Molecular Discovery software package;[47,48] Cresset software,[49] which uses the ligand field to correct the ligand position relative to the target; and PHASE module of the Schrödinger software package.[50]

Preliminary data on the biological binding showed high accuracy of 3D-QSAR in COMBINE[51] and AFMoC systems;[52] more accurate techniques, HASSLE[53] and MSTD,[54] have been developed. Finally, there exist QMOD (Quantitative MODeling) systems[55] in which the binding pocket is first designed and then ligands are placed inside in the most complementary way. As this is done, the pocket is modified by means of iteration feedback: COMPASS 3D-QMOD.

Currently, the largest amounts of experimental data on biological activities are present in ChEMBL,[56] Pubchem[57] and PASS[58] databases, which include in total more than 300 million records on the biological activities of chemical compounds, enough to form a multitude of training and validation sets. The average accuracy of PASS prediction for more than 5 000 types of biological activity exceeds 0.96. The GUSAR program was developed for quantitative analysis of the structure – activity relationships,[59] while the DIGEP-Pred program was meant for predicting gene expression profiles.[60]

Local versions of the PASS and GUSAR programs, PASS online,[61] and 19 free Web services based on them are available. More than 24 000 users from approximately 100 countries utilize the Way2Drug/PharmaExpert Web services.[62] Promising models for targeted synthesis of biologically active compounds were formulated for more than 830 000 predictions. The PharmaExpert program is meant for analysis of the results of PASS- and GUSAR-based predictions and for the search for chemical compounds with desired ranges of biological activity. By combining the PASS, GUSAR and PharmaExpert programs, one can evaluate the pharmacotherapeutic, side and toxic effects of new compounds.

Quite an efficient type of statistical treatment with a feedback is ViralChEMBL, a semi-automatic filtering procedure for antiviral activity,[63] which represents a set of biological activity data obtained by comparing data on the antiviral activity of ChEMBL structures, data on the viral taxonomy provided by the International Committee on Taxonomy of Viruses (ICTV) and structural information.

The efficiency of this procedure is much higher than that of the classical procedure; it provides an antiviral activity profile for almost any structure from the ChEMBL database. Self-organization maps of the activity profiles of promising models (machine learning) represent a step forward in the development of this approach. Machine learning in the selection of descriptors, in combination with the molecular graph theory, is also used in the Charming QSAR & QSPR system.[17]

Fairly good results are provided by the multi-target QSAR (mt-QSAR), in which chemical binding in the training set is correlated with the molecular geometry and supplemented by contributions of physicochemical characteristics and biological activities.[64]

4.2.2. Electrostatic (coulomb) interactions as the base for 3D-QSAR modelling

Coordination of objects with spatially distributed molecular potentials is based on interaction between the electric fields of the objects. For the initial coordination of molecules, dipole – dipole interactions, van der Waals forces and solvolysis are equivalent, their difference being determined

only by electrostatic properties, as was found by Ariens *et al.*[65] back in 1964.

Electric fields act over longer distances than geometric factors, *i.e.*, if the electric fields are not complementary, that is, the charges of the interacting surfaces of the objects are incompatible (in the absence of external impacts on the system), the chemical binding may fail to occur even if the interacting structures perfectly match in geometry. Therefore, the obvious fact that the biological activity is determined by the efficiency of complex formation between the ligand and the target,[66,67] in combination with the key–lock theory, which implies complementarity of the ligand and target geometries in the resulting complex,[68–71] should be supplemented in this case by complementarity of their electrostatic fields.[72–74] Like the molecular geometry, the configuration of the molecular field can be considered, with equal reason, to be a principal factor of the key–lock concept.

The electric field of a molecule is a vector quantity, representing the sum of vectors of the unit electric fields present in the system. The distribution of the 3D potential generated by the molecule is a function of the molecular geometry and is called the molecular electrostatic potential (MEP).[75]

Comparison of MEPs for evaluating the electrical complementarity of binding species is one of the 3D-QSAR methods.[76,77] As regards comparison of the MEP characteristics of molecules, note the main benefit of this method, namely, the fact that neither the sorts of atoms that form the molecule nor the arrangement and binding of the atoms are significant for the calculation, but only a secondary factor, the generated field, matters. Similar MEP distributions can be generated by absolutely different structures; therefore, the 3D-QSAR training set may include molecules of any sort that possess the desired property. The verification of the model compliance for a new compound is not restricted to geometry either. In the next Section, we demonstrate how this feature is applied to analysis of fragrances.

The MEP description can be either integrated or discrete. The discrete description of MEP in 3D-QSAR uses points characterized by two parameters, the coordinates and the charge. For the sake of simplicity, the 3D configuration of MEP can be restricted by the molecular surface of MEP. A comparison of such surfaces results in the number of coinciding points as a quantitative parameter.

4.2.3. Mapping of the molecular electrostatic potential

Electrostatic potential is a feature of a point charge, a discrete value, while the electrostatic field generated by the molecule is a vector. The researcher is faced with the question of what should be compared and which descriptors are decisive when it is necessary to determine the complementarity of electrostatic fields of two or more binding species.

It is evident that the mapping of the integrated vector configurations of electric or magnetic fields requires huge computational resources. Therefore, it is more appropriate to compare the point (discrete) descriptions of the molecular electrostatic potential (MEP). The MEP is described using the interaction energy between the molecule and a proton removed to a certain distance (test proton).[78] The coordinates of MEP points that are significant for structure comparison in the QSAR method serve as one of the property descriptors.

According to the Faraday law, the charges of every point of MEP are additive values derived from the charged parts of the molecule. The electrical potential q_{tot} at a distant point in space is given by

$$c_{tot}\, q_{tot} = c_1 q_1 + c_2 q_2 + \ldots + c_k q_k\,, \qquad (1)$$

where $q_1 \ldots q_k$ are the added charges (*e.g.*, atom-centred charges) and $c_1 \ldots c_k$ are contributions of these point charges inversely proportional to the squared distance.[79]

Nevertheless, this simplified calculation of MEP by summing up the unit fields from point sources shows results comparable to the results of more rigorous calculations using semiempirical methods.[80]

Masek *et al.*[81] and Svitanko and co-workers[82,83] took full advantage of the use of electrostatic factors, independent of the nature of point sources: in calculations and comparison of MEPs, they took into account similar types of atoms and induced steric and electrostatic fields, that is, the hydrophilicity or hydrophobicity of a molecule or its parts. A hydrogen bond calculation in terms of MEPs within the QSAR approach was reported by Ghafourian and Dearden;[84] the accuracy was compared with that of analogous MNDO, AM1 and PM3 calculations.[85]

The MEP maps are composed by the following methods:

— scalar cuts of the vector field at certain distances, *i.e.*, the molecular potential surface is described by discrete elements of a surface molecular graph[86] (Fig. 1). Each surface point is characterized by two parameters: the relative coordinates and the charge. The key points reflecting both the molecular electrostatic potential and the molecular geometry are selected at the bulges and concavities of the molecular surface, but not inside or outside this surface. As a rule, 200 to 400 points with a step of 0.8–2.0 Å are selected for a ligand. The purpose of QSAR modelling in this case is to construct three-dimensional structural descriptors representing the key points;[82]

— several concentric charge surfaces that specify the charge distribution of the MEP, or the Connolly's molecular surface algorithm;[87]

— CHELP procedure,[88] in which a larger number of points for description of the MEP surface is allowed, *i.e.*, the discretization is more accurate.

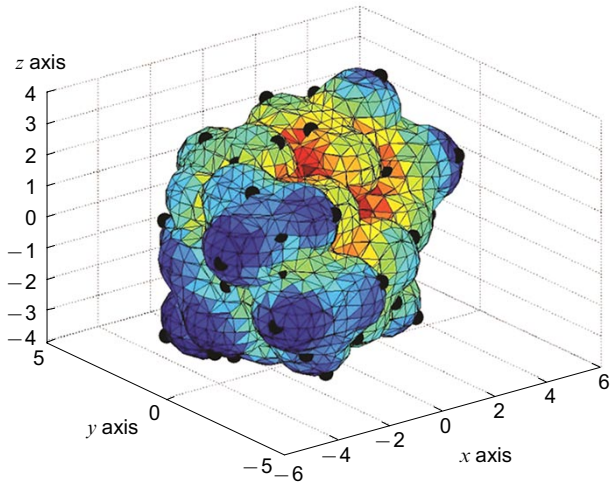


Figure 1. MEP key points on the molecular surface.[82] Reproduced with the permission of Elsevier.

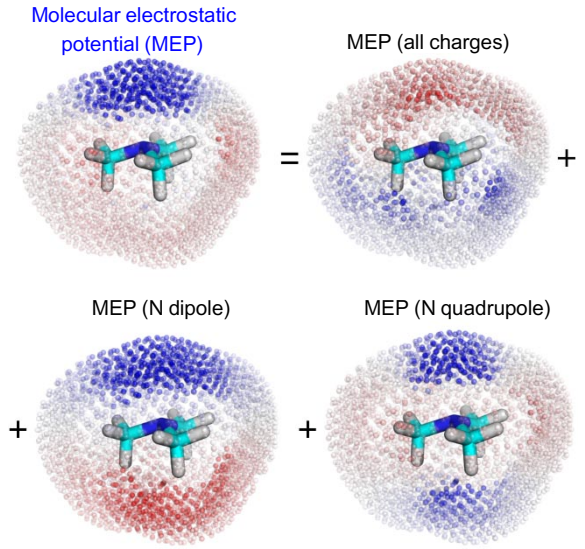


Figure 2. Construction of MEP using single charges and refinement with dipole and multipole structures for amines.[92] Reproduced with the permission of Elsevier.

More complex programs like PDM88 define the molecular electrostatic potential as a function of bond-centred interactions,[82] *i.e.*, MEP charges generated by one or several chemical bonds, which generate the overall electrostatic field.[89]

V.Palyulin's research team reported a multifield description (dipoles, quadrupoles) of MEP for a number of sulfur,[90] nitrogen[91] and phosphorus[92] compounds (Fig. 2). As a result, without additional quantum chemical calculations, the authors obtained dependences for the effect of substituents on the 3D structure of MEP,[92] *i.e.*, they actually provided a multifield modelling of a property in terms of 3D-QSAR.

4.3. Modelling of the odour of fragrances by constructing MEP-based model

Despite the complexity of generation of impulses in olfactory receptors upon the interaction with a fragrance molecule, the classification of odours was among the first and most accurate QSAR applications. This is due to high accuracy of differentiation of flavours by expert perfumers, who are able to determine the contribution of a particular odourant to a composition with an accuracy of a few percent; therefore, the results of calculations of fragrance activities can be validated without instrumental or medical studies.

Most of the theories of odour formation and transmission have not been experimentally confirmed to a sufficient extent.[93] Mention should be made of the key–lock theory, which is based on the formation of an odourant–specific target complex. The targets for particular odours were found in many animals and in humans. Luan *et al.*[94] and Rojas *et al.*[95] classified the structures and functional groups responsible for various odours.

It was found that the active site of olfactory receptors, like the active sites of many other receptors including therapeutic targets,[96–101] is sensitive to the substrate stereochemistry.[102,103] For example, the odours of stereoisomers of some fragrances may range from the presence to complete absence. The stereochemical factor included in QSAR made it possible to build a model using the molecular simulation software CATALYST and to synthesize six camphor stereoisomers — camphor odourants.[102]

The odotope theory, or shape theory, states that, instead of the rigid key–lock mechanism, receptors are bound to a specific universal receptor on the epithelium, and the molecule generates a nerve impulse corresponding to its structure, which is transmitted to brain, and brain perceives it as a definite odour.[104] It appears that different molecules can activate the universal receptor in different ways, generating different odours. However, the shape theory cannot explain, for example, why structurally different molecules can have the same odour and why structurally similar molecules can smell quite differently.[105] The weak-shape theory also cannot account for the difference between the odours of deuterated and non-deuterated compounds.[106]

The mechanism of odour perception, olfactory receptors, odour maps, olfactory system structure and classifiers of the odour theory were formulated by Richard Axel from the Columbia University (USA) and by Linda Buck from the Fred Hutchinson Cancer Research Center in Seattle (USA), who were awarded the Nobel Prize of 2004 in physiology and medicine for these studies.[107,108]

The above theories were summarized in relation to the amber odour by Kovatcheva *et al.*[109] The basic QSAR studies dealing particularly with the structure–amber odour relationship were carried out in the early 1990s by P.Vlad and I.Bersurker using a classical selection (Fig. 3) of amber odourants.[110]

An evident way to use MEPs in the structure–odour QSAR for building the model is to compare MEP descriptions, or to put it simpler, to compare the MEP points on the molecular surface (key points, see Fig. 1).

For comparison of surface descriptions in terms of QSAR, it is possible to use the description of surfaces by adjoining two-dimensional structures of an invariable length.[82]

For the description to be accurate, the number of points must be as great as possible. The coincident points of surfaces of the training set constitute a discrete description

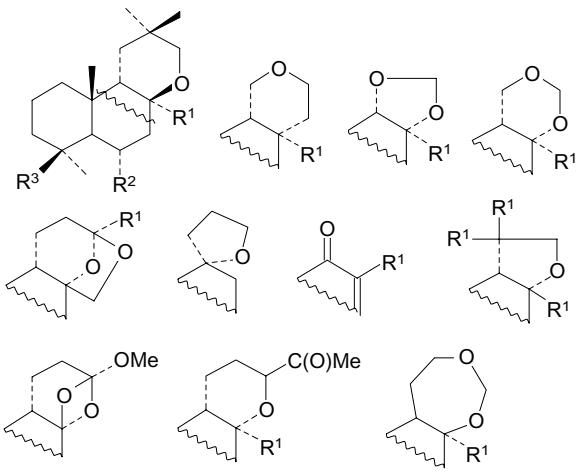
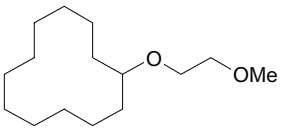


Figure 3. Examples of a training set of compounds with amber odour.

for the model of a property applicable for testing molecules for which this property is unknown.[83, 111]

The use of MEP in QSAR analysis can serve for determining the activity of compounds that considerably differ from the training set compounds in geometry, but produce molecular surface charge distributions corresponding to the model set. An analysis of the amber odour by this method using the symbol descriptions of the structures of steric isomers demonstrated [112] that an additional molecule structurally differing from the sesquiterpenes of the classical training set corresponds to the MEP model of the amber odour. This is ambrolignan **1**, which actually has a woody amber odour.[113]

Structure 1



1

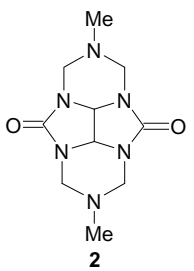
Analysis of the MEP training set consisting of aromatic musk odourants resulted in an equally high accuracy of prediction of the musk odour.[80, 114]

4.4. Determination of the structure – property relationships in the series of bicyclic ureas (glycolurils) with neurotropic activity

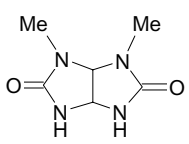
In the late 1970s, researchers from the N.D.Zelinsly Institute of Organic Chemistry, Russian Academy of Sciences, and the Kazan Medical Institute showed that saturated bicyclic bis-ureas (glycolurils, BBUs) have a neurotropic activity.[115] However, although a broad range of glycolurils and their structural analogues have been synthesized since that time [116 – 118] and psychoactive compounds have been empirically identified among them,[119, 120] attempts to determine the complementary receptor by docking have not met with success.[114] As a result, QSAR is still the major method for predicting the activities of these compounds.

The strategy of 3D-QSAR application to fragrances using the MEP surface was reproduced by Svitanko *et al.*[121] to process a training set of BBUs. It was shown that the calculation of electrostatic potential is quite applicable to modelling of the pharmacological activity of potential medicines. The mathematical model thus obtained allowed for the synthesis of new structures **2** and **3** with neurotropic activity, which showed the best fit to the model. Having supplemented the statistical methods with machine learning, the authors successfully classified the conformational dependences and identified a more labile molecule **4**,[122] belonging to the class of thioglycolurils, which showed neurotropic activity.

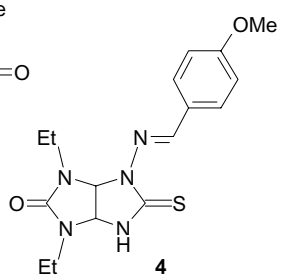
Structures 2 – 4



2



3



4

The diversity of developed calculation methods using 'only collections of ligands' for various molecular systems was fully reflected in the classical book of Kubinyi *et al.*[123] and in a recent comprehensive review.[14]

The existing extensions to 3D-QSAR (4D, 5D, *etc.*), which use more advanced mathematical methods for data processing, are summarized in a review of Damale *et al.*[124]

For the subsequent presentation, it should be noted that Vahedi *et al.*[125] utilized QSAR and modern mathematical statistics methods [multiple linear regression (MLR), support vector machine (SVM) and artificial neural networks (ANN)] to evaluate the activity of poly-adenine-ribose-polymerase (PARP) inhibitors.

Thus, the diversification and some subjectivity of the sorts of molecular descriptors, in some cases, cast doubt on QSAR as a base for designing synthetic models.[126] However, there are quite a few published examples of using QSAR in the targeted synthesis, and this method will be used as long as there are objective difficulties in determining the active site structure of the target for the ligand – target coordination. QSAR can be useful for preliminary selection of the potentially active ligands even in those cases where the structure of the target is known.[14]

5. Molecular docking and its extensions in the modelling of targeted synthesis (Stage 2)

5.1. Docking and virtual screening for identification of hits

The progress in X-ray diffraction and spectroscopic methods, which resulted in determination of more than 100 000 three-dimensional protein structures and provided structural information on the key drug targets, gave rise to the structure-based drug design (SBDD) methods. Molecular docking is one of the fastest and frequently used SBDD methods. This method can predict both binding affinity of the ligand for the protein and the structure of the protein–ligand complex. Thus, it enables the determination of the ligand specificity for a target for further optimization.

More than 60 docking tools and programs have been developed over the last two decades for both fundamental research and commercial purposes. Examples are DOCK, AutoDock, FlexX, ICM, Glide and many other. All programs were evaluated for their ability to reproduce the correct mode of ligand binding to the biological target and to identify the known active compounds in virtual screening tests. Recently, it was reported that the best docking programs are able to predict the experimentally determined coordinations in ligand – target complexes with root-mean-square deviations (RMSD) from 1.5 to 2 Å (Ref. 21) (note that the average resolution for X-ray diffraction data in the protein database is 2.1 Å).[127]

In 2017 – 2020, modelling of kinase inhibitors using the Lead Finder software was brought to the level surpassing the world analogues developed by pharmaceutical giants, which was confirmed by the D3R Grand Challenge (D3R GC) analysis. This is an annual competition of *in silico* drug designers organized by the University of San Diego (USA) in cooperation with the largest pharmaceutical companies. The Lead Finder system, developed by researchers from the N.D.Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, proved to be efficient in the prediction of ligand IC_{50} values based on the 3D structure of the protein – ligand complex ('structure-based scoring') where it ranked fifth among 26 participants. In the prediction of ligand IC_{50} on the basis of the ligand structure ('ligand-

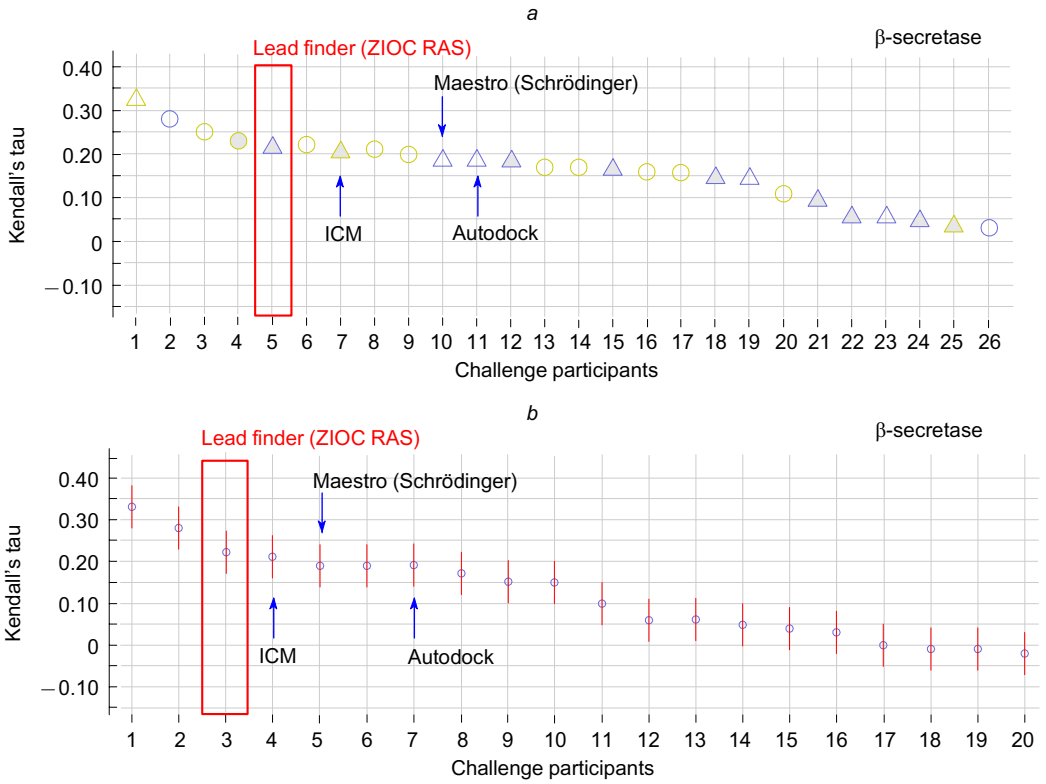


Figure 4. Choice of the most active compounds (*a*) and accuracy of structure prediction (*b*) for β-secretase in terms of D3R Grand Challenge. The Figure was prepared by the authors using the data of Ref. 128.

based scoring') it ranked third among 20 groups (Fig. 4 *a*,*b*), being superior to top commercial systems for predicting the biological activity.[128]

Currently, there is a sufficient body of information to confirm the benefits of docking in drug development. Preliminary docking may reduce the cost of high-throughput screening (HTS) by decreasing the number of molecules to be checked to a relatively small number of probable drug candidates. A blind prospective study of inhibitors of enzymes such as heat shock protein (HSP90), protein phosphatase (PTP-1B), β-lactamase and papain-like cysteine protease demonstrated that preliminary docking increases the hit rate (the ratio of the number of active ligands to the number of tested ligands) by 2 – 3 orders of magnitude. It is worth noting that in the HTS of commercially available compounds on an industrial scale, even a 10% increase in accuracy is often sufficient to justify the financial costs of the whole enterprise.[129]

A vivid example and proof of the advantage of the *in silico* approach over the classical HTS is the discovery of a new inhibitor of TGF-β receptor I. One and the same molecule (HTS-466284) was detected independently by virtual screening and by traditional HTS. The compound found experimentally required *in vitro* screening of a large library of compounds in TGF-β-dependent cell analysis, which took more time and was more expensive than the cost of development of the computational analogue.[130]

One more benefit of docking is a relatively short time required for the search for active compounds. For example, the whole process of development of serotonin receptor (5HT1A) inhibitors from *in silico* screening to phase I clinical trials took less than two years after the study was initiated.[130] In relation to several real drug discovery projects of Merck & Co. Inc., it was demonstrated that virtual screening (VS) plays a key role in accelerating the search for ligands for particular targets with different amounts of initial information available.[131] In the first example, the availability of crystal structures of the protein and known allosteric inhibitors made it possible to use the VS approach to gain the highest probability of detecting new allosteric HIV integrase inhibitors. In the second example, the VS was carried out to accelerate the generation of high-coordinate complexes for coagulation factor, the FXIa protein.

It is believed that the quality of hits in the docking-based virtual screening increases with increasing size of the compound set; therefore, the active search for platforms that allow large-scale VS is in progress.[132] For example, the new VirtualFlow platform, capable of efficient generation and docking of ultra-large libraries of molecules, was used for VS of 1 billion compounds using a Kelch-like ECH-associated protein 1 (KEAP1) model.[133] The VirtualFlow platform was utilized to identify a set of molecules, the best of which binds to KEAP1 with a nanomolar affinity (the dissociation constant K_d is 114 nM).

Docking of 170 million compounds resulted in the detection of the phenolate inhibitor of AmpC, which revealed a group of new efficient inhibitors of AmpC β-lactamase. This group of compounds had not been evaluated previously as AmpC inhibitors, because of the absence of phenolates in the small libraries that were tested. New D4 receptor agonists with efficiency ten times exceeding the efficiency of compounds known previously were obtained in the same way.[134]

In the next study, not only were new BRD4 ligands actually discovered, but it was also possible to practically

evaluate the utility of the generally available big data for the construction of robust predictive models. The obtained models were applied for the virtual screening of a collection comprising 2 million compounds. This was followed by virtual screening of the subset comprising 2992 most active molecules selected by the VS procedure. Experimental testing confirmed 29 hits, which amounts to 1% of selected candidates. The VS hit rate was 2.6 times as high as that for the equivalent random screening.[135]

In recent years, the attention of drug designers has been attracted by the docking-based inverse virtual screening (IVS) technique, consisting in the identification of the potential protein targets for ligands. Despite the shortcomings of the method caused by incompleteness of available protein databases, neglect of the protein flexibility and low accuracy of evaluation functions, there are examples of successful use of this approach for identification of targets for natural compounds. The SELNERGY system containing 400 protein structures and FlexX were used to identify the target for ε-viniferin, a natural phenol (PDE4), and meranzin, an essential oil component (COX1, COX2, PPARγ), respectively.[136]

A database of targets comprising 163 proteins was used to identify protein kinases PDK1 and PKC as the targets for xanthohumol and isoxanthohumol by means of the AutoDock Vina program. This result was confirmed by *in vitro* studies. Additionally, IVS can be used to predict the side effects in the initial stage of development of new drugs. The INVDOCK program was applied to eight clinical drugs (aspirin, gentamicin, ibuprofen, indinavir, neomycin, penicillin G, 4H-tamoxifen and vitamin C), resulting in the identification of 83% of all known targets responsible for the side effects of these drugs.[136]

Apart from the large-scale drug development projects, docking-based virtual screening is widely used in the common laboratory practice to find biologically active compounds for targeted synthesis and for the development of more active compounds based on known ligands.[137–143] A brief description of some docking applications is given below.

Acetylcholinesterase (AchE) inhibition increases the level of the acetylcholine neurotransmitter and symptomatically improves the impaired cognitive function. Identification of AchE inhibitors was performed by the docking-based virtual screening of a standard lead set of the ZINC database containing more than 6 million small molecules using the GOLD software. Nine of the ten selected hits bind to the enzyme with affinities from 0.5 to 0.7 μM; eight of these compounds are non-toxic.[144]

In one more study, molecular docking was utilized to predict the activity of tamoxifen, isoprenaline, chlorpromazine and carbamazepine against glutathione S-transferase (GST), acetylcholinesterase (AchE) and butyrylcholinesterase (BchE), which are potential targets in the treatment of Alzheimer's disease. Furthermore, the results of docking gave a better insight into the mechanism of inhibition by the above drugs and the key structural features responsible for the ligand activity.[145]

Virtual screening can be useful for the identification of potential drugs in libraries of so-called dark chemical matter, that is, sets of compounds that have never exhibited biological activity despite extensive HTS studies. Docking-based screening was carried out against GPCRs, particularly, adenosine A2A and dopamine D4 receptors. The top-ranked GPCR ligands for adenosine A2A were experimentally evaluated; the most active compounds had a submicromolar affinity for the receptors. This approach may promote development of drugs with improved selectivity profiles.[146]

An attempt was made to use virtual screening to find molecules active against both dopamine D2 and serotonin 5-HT2A receptors and selective for the histamine H1 receptor, and also active against κ-opioid receptors with a selectivity higher than that for μ-opioid receptors.[147] While the hit rate varied from 40 to 63%, the selected ligands proved to be equally active also against the targets to which the absence of binding was assumed. Although the authors[147] identified a ligand with a 21-fold selectivity for D2 and 5-HT2A receptors *versus* the selectivity for the H1 receptor, this should rather be considered as an exception. Thus, addressing the problem of search for selective ligands may require new strategies in this field.

Consider an example of development of small molecules that occupy the haem-binding site in indoleamine 2,3-dioxygenase 1 (IDO1). This enzyme is an intracellular haem-containing dioxygenase associated with the immunosuppression in cancer. A combined strategy of virtual screening including pharmacophore screening based on the highly specific haem-binding group (HmBG) and molecular docking gave four compounds that exhibited correct binding at the haem iron-binding site. Further structural optimization afforded the compound S18e3, which showed high *in vivo* antitumour efficacy.[148]

It is assumed that the galactokinase 1 (GALK1) inhibitor, which decreases the production of galactose 1-phosphate (GAL1P), may be considered as an agent for treatment of galactosemia — a hereditary disease caused by metabolic disorders in the conversion of galactose to glucose (mutation of the structural gene responsible for the synthesis of galactose-1-phosphate uridyl transferase). Hu *et al.*[149] reported the discovery of a new series of GALK1 inhibitors by virtual screening of 230 000 compounds. As a result of a multi-factor prioritization procedure (contraction of the range of considered objects with each request), 350 compounds were selected. Seventy five of these compounds exhibited inhibitory activity in the biochemical assay of GALK1. In addition, several phenylsulfonamides with excellent ADME characteristics *in vitro* were selected for the subsequent consideration. These compounds were found to decrease the level of GAL1P in the fibroblasts of galactosemia patients. This result may serve as the starting point for the subsequent development of drug candidates for the modulation of the GALK1 inhibitor for the treatment of galactosemia.[149]

Zhou *et al.*[150] carried out virtual screening to find new ligands for an androgen receptor (AR) on the basis of the crystal structure in the complex with its agonist. Altogether 58 structurally diverse compounds were selected; they showed IC_{50} values of approximately 4 μM, which is even lower than this value of anzalutamide (IC_{50} = 13.87 μM), a second-generation AR antagonist, which is currently present on the drug market.[150]

Virtual screening of G-protein coupled receptors (GPCR) based on high-resolution structures would probably become a preferable approach to the development of next-generation drugs. The applicability of this approach was demonstrated by discovery of new ligands for melatonin MT1 receptor. Currently, drugs acting on melatonin receptors are prescribed for circadian disorders (jet lag, shift work, *etc.*), insomnia and depression. Stein *et al.*[151]

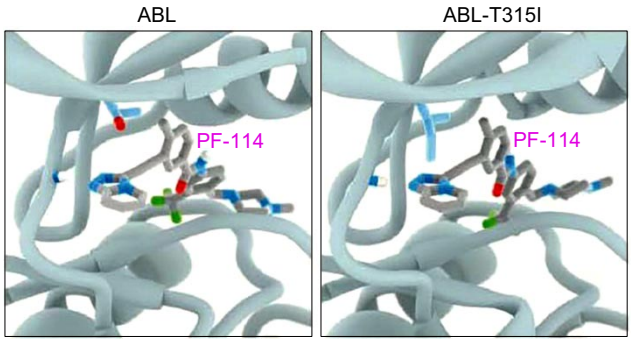


Figure 5. Binding of the PF-114 inhibitor at the active sites of ABL and ABL-T315I kinases.[152] Reproduced with the permission of Springer Nature.

intended to identify MT1-selective ligands using a VS library containing more than 150 million molecules. As a result, two new MT1-selective agonists were found. The VS hit rate was 39%, with some primary candidates exhibiting nanomolar affinity for MT1.

A selective tyrosine kinase inhibitor, PF-114 (Fig. 5), was developed using molecular docking (LeadFinder) and the subsequent targeted synthesis.[152] The PF-114 agent inhibits T315I and other mutants of the BCR/ABL protein, which cause resistance to existing drugs, and is meant for the treatment of drug-resistant chronic myeloid leukemia.[153]

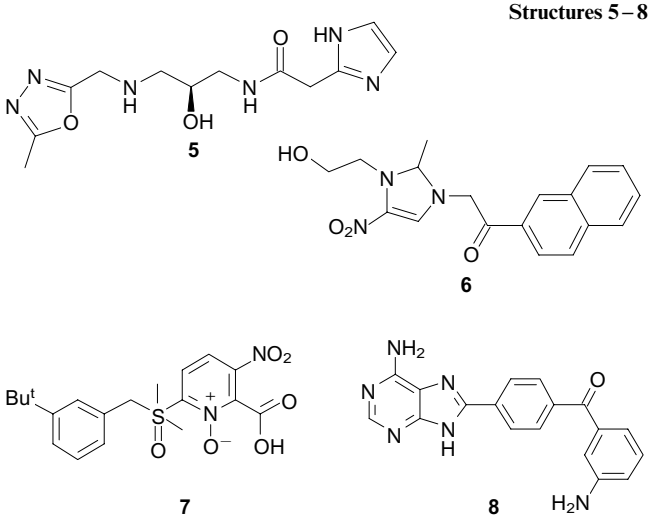
5.2. Combined use of QSAR methods and molecular docking

Allosteric modulators of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors play an important role in the therapy of neurological disorders such as schizophrenia, Alzheimer's disease and attention deficit hyperactivity disorder in adults.[154] The combined use of molecular docking, CoMFA modelling and pharmacophore analysis gives a consistent picture of structural requirements for the molecular design of modulators of AMPA receptors — drug candidates for the correction of functional disorders of the central nervous system.[155] The targeted synthesis based on the modelling resulted in the preparation of several such structures,[156, 157] which showed biological activity in line with the predicted values.[158, 159] The highest activity at nanomolar concentrations was found for tricyclic 3,7-diazabicyclo[3,3,1]nonane (bispidine) derivatives.[160]

A similar combined use of pharmacophore modelling, molecular docking and 3D-QSAR for analysis of a training set comprising 66 cyclooxygenase-2 (COX-2) inhibitors resulted in the generation of a model that interpreted the descriptors of COX-2 inhibitors and made it possible to select the most efficient structure for this purpose.[161] The sequence of filtering (decrease in the number of objects by an order of magnitude in each stage) was as follows: pharmacophore analysis of the database of molecules, virtual screening (QSAR), molecular screening (molecular dynamics), visual expert evaluation and final model **5**. Compound **5** inhibits COX-2 when present in nanomolar concentrations.

The same sequence of filtering was applied in the search for models of indoleamine-2,3-dioxygenase (IDO1) inhibitors.[162] The experimental verification confirmed the efficiency of structure **6**.

Structures 5 – 8



Preliminary 3D-QSAR (CoMFA and CoMSIA) filtering with the same sequence also proved to be necessary for the treatment of databases to generate models of SARS-Cov-2 inhibitors based on pyridine oxides.[163] The best *in vitro* efficacy was found for compound **7**.

In a study of inhibitors of Janus kinase (JAK1), which is involved in the development of AIDS and autoimmune

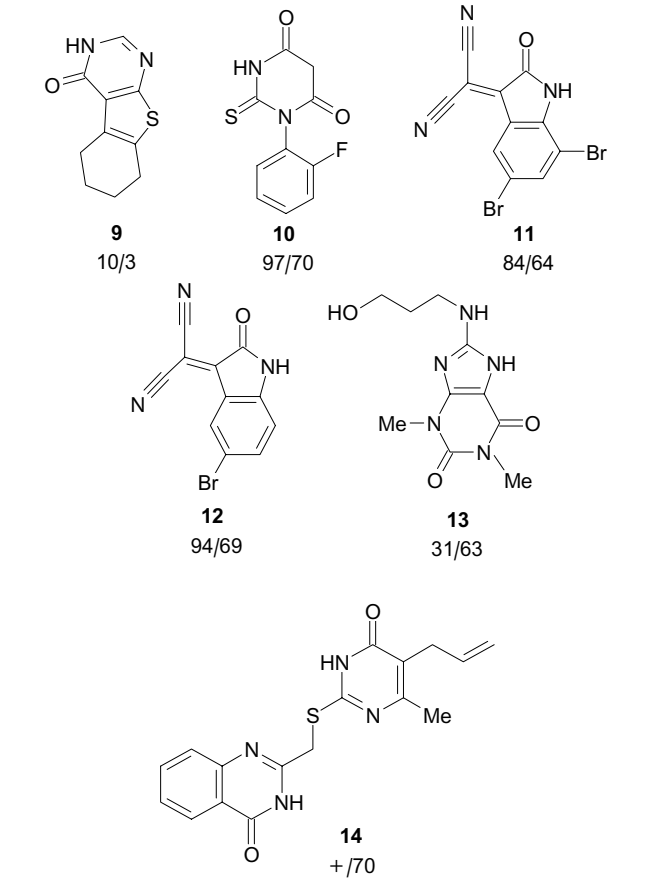


Figure 6. Examples of structures (**9** – **13**) with identical activities according to molecular docking and 3D-QSAR calculations (IC_{50}, calculated/experimental) and structure **14**, which was recognized to be inactive according to molecular docking data, but was active according to QSAR and experimental results.

diseases such as rheumatoid arthritis and other, a training set consisting of 151 compounds was used for 3D-QSAR and a docking model was generated in parallel.[164] After comparison of the results, several models of active inhibitors were identified, *e.g.*, compound **8**.

The combined use of QSAR and molecular docking also proved to be effective in the modelling of anti-tuberculosis drug candidates.[165]

Apart from the above applications of QSAR for data pre-filtering, mention should be made of a study describing parallelization of active structure modelling by QSAR and molecular docking.[166] The authors carried out parallelized (3D-QSAR and molecular docking) search for new potential inhibitors of poly(ADP-ribose)polymerase-1, which catalyzes poly-ADP-ribosylation of various protein and DNA substrates. The training set comprised 86 active, 34 inactive and 76 control compounds, taken for molecular docking calculations described above. Upon the use of QSAR, the five models (structures **9**–**13**) (Fig. 6) identified by docking were supplemented by one more structure (**14**), which was defined by docking as inactive. The experimental verification showed the activity of compound **14**.[166]

In recent years, combining QSAR and molecular docking methods has become a popular approach. For example, combined calculations and subsequent synthesis and biological assays were used to model inhibitors of tyrosine kinases,[167, 168] topoisomerase IV subunit[169] and synthetic analogues of anti-inflammatory aminoazaadamantanes.[170]

Thus, 3D-QSAR can be efficiently used to supplement physical modelling methods.[171] The efficiency of the method increases when machine learning is employed in various stages of the process.[172]

5.3. Overcoming the conceptual limitations of classical docking

5.3.1. Flexible protein docking

The methods used most frequently for predicting the ligand binding imply that the structure of the receptor remains invariable. However, most of proteins are flexible when binding the ligand, which causes local shifts of side chains or pronounced conformational changes in the tertiary protein structure. Consequently, docking of a rigid receptor, taking into account only one conformation, leads to false negative results. This is due to the fact that a protein can exist in constant motion between various conformations with similar energies, which is usually neglected in the docking.[173]

Modelling of protein alteration upon the interaction with the ligand requires a lot of time because of a large number of degrees of freedom present in the system. The DESRES group used molecular dynamics simulation to dock ligands into a receptor conformation requiring an induced movement of amino acids to achieve the correct pose (the spatial position of the ligand relative to the target).[174] However, reliable binding required 30–300 microseconds of complete dynamics simulation per ligand. Furthermore, the robustness of this protocol across numerous protein–ligand complexes was unknown, since the computational restrictions made it possible to study only a small number of examples.

For addressing this problem, several methods have been developed that take into account the protein flexibility in docking calculations. One of such approaches is the induced fit docking (IFD), which allows flexible parts of the receptor to move during docking *via* iterative selection of the ligand pose with the subsequent adjustment of positions of side chains. This method is based on the hypothesis that a conformational change of a bound protein can be induced by a ligand. Detailed descriptions of various IFD algorithms can be found in the literature.[173, 175–178]

It was shown that the accuracy of the IFD method in predicting the poses is higher compared with traditional docking. Rigid receptor docking for 21 cases of pharmaceutically relevant targets gives an average RMSD of 5.5 Å, while the average RMSD for the ligand in the case of flexible receptor docking is 1.4 Å.[179]

Induced fit docking is especially relevant for docking of ligands into poorly accessible targets.[180, 181] As an example, consider UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA), which is accessible only in the closed form with a binding site pocket which is not optimal for docking. The IFD method was used to simulate the induced fit effect for ligand binding to MurA; this resulted in a larger binding site, which was used in virtual screening to find compounds with binding constants of $\sim$1 μM.[182]

The known crystal structures of Mnk kinase contain amino acids with side chains, which prevent ligands from access to the ATP binding site. This precludes the use of docking protocols with flexible ligands and rigid proteins, which are applied most often to screen large libraries of compounds. Virtual screening based on IFD modelling of the protein active site revealed a number of Mnk inhibitors, which can be used for further optimization as potential therapeutic agents for the treatment of malignant blood diseases. The authors used a special IFD protocol based on mutations (M-IFD), which allows docking of the known Mnk1 inhibitor ligands into the active site and consists in temporary replacement of the amino acid residues Phe192, Val63 and Leu55 by alanine with subsequent IFD. This gave a Mnk1 structure with an accessible ATP-binding pocket, which was subsequently used for screening.[183]

A conceptually different approach for taking account of the protein flexibility in the docking experiments is ensemble docking (ED). This method is based on the 'conformational selection', which describes the molecular recognition as a process in which the ligand chooses the most complementary conformation of the protein from the ensemble. The ligand is docked into the pre-specified multiple protein conformations, which can be found either experimentally (X-ray crystallography or NMR spectroscopy data) or by computer selection, *e.g.*, by molecular dynamics.[184]

The ensemble docking requires much more computational resources than the use of a single structure or IFD. However, there are reported examples that illustrate the considerable increase in the number and diversity of biologically active candidates by means of ED-based virtual screening,[185] the promising use of this method for predicting toxicity[186] and increase in the quality of prediction of binding as a whole in comparison with classical docking,[187] which justifies the computational cost.

Additionally, the ED approach is often employed to predict the ligand binding to complex targets, including the targets, for which only a partial crystal structure is available or there is no *a priori* knowledge on the small molecule-binding sites, and the targets the binding to which is accompanied by conformational changes.[184, 188–190]

For example, this method proved to be useful for the discovery of new molecules able to bind to the surface of blood coagulation factor Xa (FXa) and modulate its interaction with the factor Va (FVa), without affecting the

enzyme properties of Xa. The strategy was to identify the binding pockets on the FXa surface in the areas that presumably interact with the partner protein FVa by docking into 'random' binding pockets of the FXa – FVa complex in various conformations determined by molecular dynamics. Experimental verification of selected compounds revealed 31 molecules (out of 124 tested molecules) as nonactive site inhibitors and 10 compounds with desired properties out of 31 selected inhibitors. It is noteworthy that docking using the crystal structure alone did not predict binding for 8 out of these 10 compounds. In other words, these compounds were found to be efficient upon docking into some binding sites that were formed only in the molecular dynamics modelling.[191]

A new trend of ED studies is the application of this method for the multiple prediction of toxicity using protein clusters that are involved, most often, in the acute toxicity and clinical failures of drugs.[192] For example, off-target activation of the adenosine receptor A2A present in the CNS tissues is related to side effects such as pain, asthmatic attacks, convulsions and other neurological disorders, which prevent the clinical use of drugs that have affinity for this receptor. When only one receptor crystal structure was used, only 58 known ligands were ranked among the 10% top-scoring molecules (*i.e.*, among the top 1174 compounds) according to docking results, which is close to the random distribution. However, when ED was used, 343 known ligands were ranked among 10% top scoring molecules.[186]

Whereas the induced fit processes are kinetically determined, the conformational selection is thermodynamic. It is quite probable that both of these processes play an important role in molecular recognition and each of them has its own scope of applicability for particular systems: stronger and longer-range ligand – protein interactions tend to favour the induced fit mechanisms, whereas weaker interactions are favourable for the conformational selection. Meanwhile, both these mechanisms can coexist in one and the same process. In several cases, it was actually shown that the protein backbone close to the bound conformation is chosen by the ligand *via* the conformational selection. Subsequently, further changes take place to optimize the intermolecular interactions by induced fit simulation.[184]

5.3.2. On-top docking

Despite the concerns about the cost and accuracy of the currently available high-throughput screening (HTS) methods, the opinion that docking-based virtual screening will completely replace HTS is premature. Currently, there are no studies in which both HTS and VS are carried out using the same library for a broad range of targets with strictly defined criteria and subsequent direct comparison of the results. However, there are examples in which both methods can be successfully used and enhance each other. In particular, HTS was employed to test 400 000 compounds for the inhibitory activity against the protein tyrosine phosphatase 1B, where 85 hits had IC_{50} less than 100 μM, while for six compounds IC_{50} was less than 10 μM (the hit rate was 0.021%). Out of the 365 compounds identified by VS, 127 compounds had IC_{50} below 100 μM (the hit rate was 34.8%). Remarkably, compounds identified by both methods showed high structural diversity.[193]

In actual drug development practice, the information on comparison of HTS and VS is variable. Moreover, in some cases, it is difficult to establish the influence of HTS or VS on the discovery of new drugs, in particular because it takes 12 – 14 years from the choice of the target to the release of the drug on the market. There are data[193] that typical hit rates of HTS in the search for the most active compounds are in the range from 0.01 to 0.14%, whereas this characteristics for VS are assumed to range from 1 to 40%.

It should be noted that, despite the presence of reported promising predictions of ligand poses by molecular docking or VS confirmed by crystallographic methods,[194] it is not always possible to identify the correct pose or even the binding site by these methods. This is illustrated by virtual screening of a library of 1013 compounds from the collection of the USA National Cancer Institute (NCI) using the crystal structure of human farnesyl diphosphate synthase (hFPPS) in complex with an allosteric inhibitor using the AutoDock Vina and Glide programs. The goal was to identify the ligand that binds to the allosteric site. Fifty compounds were selected for experimental verification. Two compounds (taxodione and arenarone) with IC_{50} in the range of 1 – 3 μM were chosen *via* virtual screening. In the crystal structures of hFPPS with taxodione and arenarone, both ligands were bound to the homoallylic rather than the allosteric site of the enzyme. The structures were modelled for binding in the allosteric site; however, a study by X-ray crystallography indicated that ligand binding to the homoallylic site is preferable. This is considered as an example of successful docking for the detection of active compounds, despite incorrect prediction of the structure of the ligand – target complex.[193]

The main cause for the fact that virtual screening of labile ligands still cannot completely replace experimental methods for evaluating the activity is the lack of approaches that are able to reliably predict the free energy of the protein – ligand binding; therefore, the results of modelling often do not satisfy researchers.[195] Apart from the problem of considering a multitude of protein conformations and inadequate estimation of the entropy contributions to the binding free energy found by docking, the key issue in the drug development is to identify the real binding site of the ligand. Otherwise, the efforts to calculate the binding energy will be futile.[196]

The Protein Data Bank (PDB) often does not contain experimental data for the chosen protein – ligand complex, especially for new proteins and ligands that have recently attracted attention of researchers. Although most of available methods for prediction of binding sites were verified using training sets of proteins with known ligand binding, a comparative study of ten representative prediction methods developed over the past 15 years implies that approximately 30% of known biologically active sites were not detected.[197]

These data indicate that the existing methods for predicting protein – ligand binding sites suffer from limitations and, hence, they are unable to verify the true site in the ranked list of putative sites. Moreover, they neglect the dynamic nature and chemical environment of the real protein – ligand binding in a living body.

The existing demand for increasing the accuracy of docking without the explicit procedure of detection of the protein active site resulted in the development of computational methods based on putative binding sites. For example, in the multiple grid approach developed in 2018, the predicted ligand binding site is covered by multiple inner grids instead of a single grid located at the ligand binding site in the classical docking[198] (Fig. 7). During the computations, a selected site point is appended to the list of grid

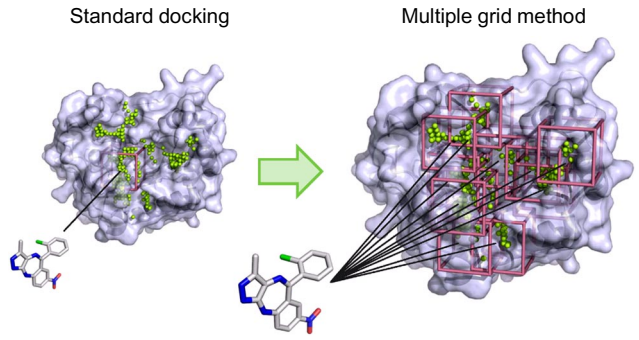


Figure 7. Multiple grid concept.[198] Reproduced with the permission of Elsevier.

centres. Then all site points included in the grid the centre of which is located at the extracted site point are removed. The extraction and removal procedures are repeated until all points of the site, in which the inner and outer box sizes are fixed at 10 and 30 Å, respectively, are removed. Verification by re-docking with the Astex dataset showed an increase in the prediction accuracy from 27.1 to 34.1% (top-scoring poses). These results emphasize the limitations and problems of scoring functions and the necessity of developing more accurate methods.

A new promising method for correction of the prediction of the binding energy, on-top docking, comprises explicit estimation of the ligand interaction energy with the protein surface, which is used as a base for estimating the actual energy of binding in the active site (Fig. 8). Since shifts of the scoring function would affect both the binding energy in the active site and the binding energy to the protein surface, it can be expected that the difference between the energies would provide much better characteristics of the ligand activity than the uncorrected binding energy in the traditional docking.[22]

The on-top docking algorithm developed by Medvedev *et al.*[22] markedly improves the results of virtual screening for many protein targets (the improvement is up to 48%), with the computation time increasing 26-fold. The improvement depends on the type of active site and the total charge of the protein. All proteins that benefit most of all from the use of this method have open active sites; this may be attributable to the fact that such proteins are less able to select ligands according to their shape. In addition, these proteins usually have non-standard total charges at pH 7. Most likely, this is due to the fact that the conventional force fields used in docking, being parameterized using experimental data, 'learned' to implicitly take into account the interaction of standard (having standard charges) proteins with the surface. It is clear that conventional methods do not operate on proteins with non-standard charges, but they are efficiently corrected by the new method. Moreover, the on-top docking actually brings the distribution of inactive ligands closer to an ideal case in which they are located at the bottom of the list in a random order.

On-top docking has already found application in the search for biologically active compounds for various targets. For example, it was used to screen the KEGG database of natural compounds for putative inhibitors that are able, similarly to curcumin, to prevent the conversion of the normal prion protein (PrP) to the amyloid protein, which causes some neurodegenerative diseases. The top-scoring candidate — 3,4-dimethoxycinnamic acid (DMCA) — was tested *in vitro*. It was shown that DMCA binds to PrP with K_d = 405 nM. Also, it is noteworthy that the use of on-top docking made it possible to propose a DMCA binding site that stabilizes the tertiary structure of PrP. An *in vitro* study of the effect of DMCA on PrP oligomerization and fibrillation showed that DMCA affects the formation of PrP oligomers: particularly, the content of oligomers decreases by 30–40% and the viability of neuroblast-like SH-SY5Y cells treated with prion oligomers increases. This makes DMCA a promising compound for the prevention of neurodegenerative diseases.[199]

Stroylov and Svitanko[200] reported identification of new putative inhibitors of the SARS-CoV-2 main protease ($3CL^{pro}$) among registered drugs by means of on-top docking. Screening demonstrated that sulfur-containing drugs show high affinity for the active site of the SARS-CoV-2 main protease ($3CL^{pro}$). Although some publications have already reported the potential applicability of this structural class of drugs, an additional molecular dynamics study showed that these compounds dissociate from the active site of the enzyme, while only the drug disulfiram preserves stable interaction with the active site. Recently, disulfiram was reported to be a putative $3CL^{pro}$ inhibitor.[201] Previously, it was shown that disulfiram inhibits proteases of the

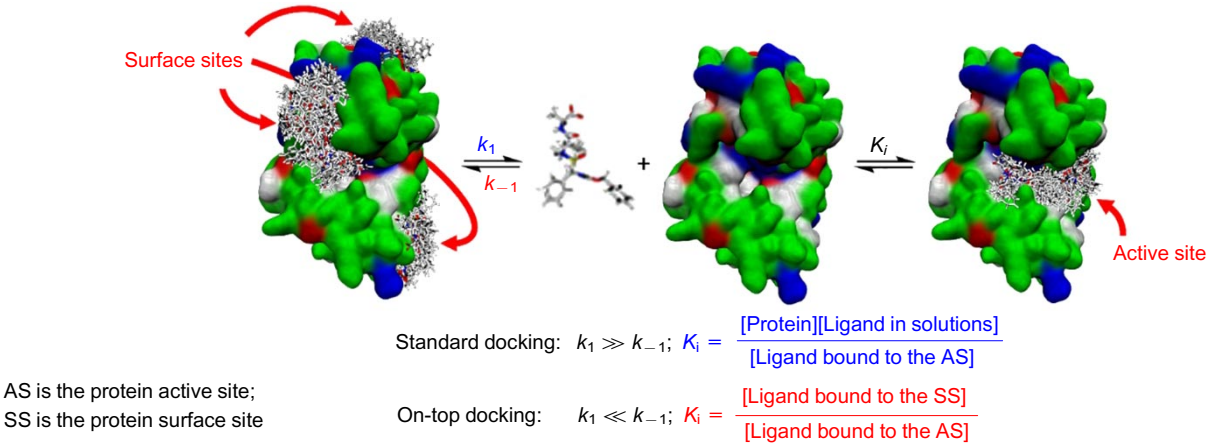


Figure 8. On-top docking concept.

MERS and SARS coronaviruses *in vitro* (IC_{50} of 14 and 24 μM, respectively).[202] In the study,[200] it was also reported an independent identification of one more drug as a potential $3CL^{pro}$ inhibitor, that is, neratinib, known to inhibit tyrosine kinase and used as an adjuvant therapy in the treatment of HER2-positive breast cancer. This drug candidate was also identified during the on-top docking and successfully passed the molecular dynamics simulation without dissociation from the $3CL^{pro}$ active site.

5.3.3. Molecular dynamics

Glide WaterMap. An interesting extension to the classical docking is considering the binding site solvation while evaluating the docking poses implemented in the Glide WaterMap software package.[203] The most probable positions and thermodynamic characteristics of water molecules are calculated from the molecular dynamics trajectory of a free protein with a fixed backbone. The mobility of amino acid side chains is thus implicitly taken into account. Despite the attractiveness of this approach, its application is rarely mentioned in the literature. Perhaps, development of new convenient protocols integrating WaterMap into the docking procedure would give an impetus to the versatile use of this method.[203]

Molecular dynamics (MD) as a tool for refining the results of docking. Despite the fact that docking can predict optimal ligand position in the receptor active site, not all key interactions between the ligand and the protein can be accurately determined from the docking results. Usually, it is recommended to perform the MD simulation of the complexes derived from docking, because this may help to refine the key protein – ligand interactions.[204] In a recent study,[205] it was shown that the high-throughput molecular dynamics (MD) simulation of the protein – ligand complex using the output data of AutoDock Vina docking improves the results of docking regarding the discrimination between active ligands and decoy compounds. The authors used 56 protein targets (of seven different classes) and 560 ligands (280 active and 280 decoy compounds) and demonstrated improvement of the area confined by the ROC curve and the axis of the false positive rate (ROC AUC) by 22% from the initial value of 0.68 (AutoDock Vina) to final 0.83.[205]

Molecular dynamics is a computational technique that can predict the time-dependent behaviour of a molecular system. Hence, it is an invaluable tool in the structure-based drug design (SBDD). Although quite a few reviews on the use of MD for the development of biologically active compounds are available from the literature,[206 – 211] below we briefly describe the most recent studies in which MD is applied to refine the results of docking.

High interest in the use of MD simulation was observed in the discovery of drugs based on protein kinases involved in a number of diseases including cancer.[212] Cyclin dependent kinase 2 (CDK2) is a potential target for the treatment of cancer, while heterocyclic purine base moieties attract particular attention as scaffolds for the development of CDK2 inhibitors. For studying the interaction mechanism and the structure – activity relationships (SAR) and for developing new compound candidates as potential CDK2 inhibitors, systematic molecular modelling of purine derivatives as CDK2 inhibitors was carried out by combining 3D-QSAR, virtual screening, molecular docking and MD. Molecular docking identified the key interactions between the ligand and the protein affecting its activity. The MD simulation of the selected complexes additionally confirmed their stability. This study provided valuable information for the development of new potential CDK2 inhibitors.[213]

Serum- and glucocorticoid-regulated kinase 1 (SGK1) plays an important role in cancer progression and metastasis by regulating inflammation, apoptosis, hormone release, nervous excitability and cell proliferation. The SGK1 glucocorticoid is considered as a potential drug target for the treatment of cancer, diabetes and neurodegenerative diseases. Virtual screening of natural compounds from the ZINC database revealed four molecules possessing a significant binding capacity and specificity for the SGK1 binding pocket. The results of docking were supplemented by 100 ns MD simulation, which demonstrated that binding of one of the compounds stabilizes the SGK1 structure.[214]

There are also examples demonstrating the applicability of MD for refining the results of flexible docking. In a retrospective study, a combination of IFD, molecular dynamics and 3D-QSAR was used to investigate the interaction between the naphthoquinone inhibitors and Hsp90 receptor. The search for the binding site using a combination of SiteMap and IFD resulted in the conclusion that naphthoquinone has the highest affinity for the N-terminal site of the Hsp90 receptor. Furthermore, MD simulation was applied to additionally study the details of interactions between the protein and the ligands. The results obtained by the combination of methods were in line with the results of 3D-QSAR and molecular docking models. Thus, it was shown that the combined strategy of molecular docking, molecular dynamics and 3D-QSAR provides a useful way towards the rational design of more potent Hsp90 inhibitors and the synthesis of new highly potent antitumour agents based on the naphthoquinone scaffold.[215]

Using new type-$I_{1/2}$ anaplastic lymphoma kinase (ALK) inhibitors with a 1-pyrine-3-piperidinecarboxamide scaffold, which can bind to both the ATP-binding site of ALK and the neighbouring hydrophobic allosteric pocket, it was demonstrated that ensemble docking and molecular mechanics/generalized Born surface area (MM/GBSA) calculations involving the application of molecular dynamics give better predictions than the traditional 'rigid' docking. Thus, inclusion of the receptor flexibility in the prediction of binding poses and binding affinity of type-$I_{1/2}$ ALK inhibitors is an important aspect for modelling of a property.[216]

5.3.4. Free energy perturbation method

The free energy perturbation (FEP) method was proposed more than 60 years ago[217] to calculate the free energy difference between states of the molecular system. Since then, this method formed the basis of a large group of protocols that use this principle. Being a potent evaluation method based on enumeration of the microstates of the system, the FEP method has a serious limitation associated with the necessity of considerable overlap between the distributions of these microstates in the molecular dynamics trajectory of the system.[218]

In recent years, the FEP method has been converted from a tool that can be used only by molecular modelling experts into a more easily operated and user friendly system. For example, the developers of the BRIDGE system[219] implemented the FEP method in a hardware-independent Docker container for cloud computation.

A challenge involved in the FEP calculations is high computational complexity, which often requires access to supercomputer resources. An actively developing trend is the application of high-performance video cards for scientific calculations. In the NAMD molecular dynamics software package, this computational approach is supported, in particular, for the FEP method, thus accelerating the calculations by approximately 30 times compared to the use of the central processor alone.[220]

One more bottleneck of the FEP calculations is high complexity of preparation of the input files. Often, this task can be accomplished only by highly skilled computational chemists. The authors of the QligFEP software package[221] substantially simplified the preparation for FEP calculations by designing an automated protocol for preparing input files for the FEP-tailored Q software package.[222] Apart from the essentially automated workflow, the authors noted that the accuracy of prediction of the binding energy attained for the test set of ligands was not inferior to that of the commercially available FEP+Schrödinger protocol.[26] Amber is yet another software package in which the FEP strategy is actively developed. In a recent study,[223] the authors used a set of 134 protein–ligand complexes, the ff14SB/GAFF1.8 force field and hardware acceleration by video adapters and attained an accuracy and speed not inferior to those inherent in the use of the FEP+Schrödinger package.

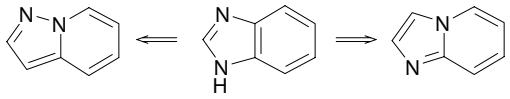
It should be emphasized that the preparation of the 3D model of the protein–ligand complex for FEP calculations has a primary significance for obtaining reliable results even when the most advanced protocols are used. Thus recently[224] it was shown that an accurate reconstruction of the backbones and side chains that are absent in the experimental structure and mutations and also preliminary relaxation of the structure of the complex by MD simulations considerably increase the accuracy of calculation of the binding free energy according to the FEP+protocol.

Also, an important role is played by the substantial conformational flexibility in the internal degrees of freedom of the ligand. The role of this factor is especially pronounced in the case of macrocyclic ligands, and its compensation requires a considerable increase in the time of MD and FEP simulations.[225] This factor can prove to be very important if the ligand has slow rotational degrees of freedom that can be sampled only during very long simulation times. In this case, it may be useful to estimate thermodynamics of these conformational transitions, which can also be done by FEP calculations.[29]

In 2020, researchers from Merck analyzed the prospective application of FEP in real large-scale projects on the search of drugs, which started with the FEP-based design of molecules followed by evaluation of their activity *in vitro*.[226] The results for 12 protein targets and 19 series of chemical compounds showed highly accurate predictions obtained for three sets of data (RMSE < 1 kcal mol^{-1}),† a medium accuracy attained for four sets (RMSE < 1.5 kcal mol^{-1}) and an acceptable prediction accuracy obtained for eight more sets (RMSE < 2 kcal mol^{-1}). The cases where the predicted affinity deviated from the experimental value by more than 3 kcal mol^{-1} were attributed to the following causes.

† RMSE is the root-mean-square error.

Scheme 2



(1) Drawbacks of the force fields. Roos *et al.*[227] distinguished several cases of overestimated changes in the relative binding free energy for the final compounds caused by improper force fields employed to describe the initial compounds, resulting in underestimation of the initial bind-

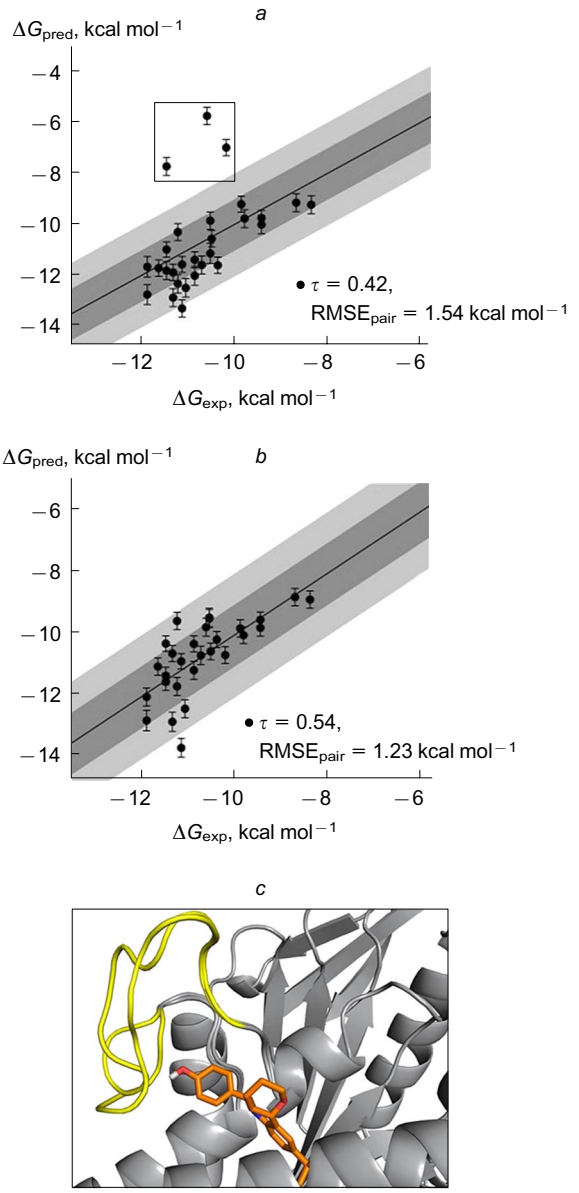


Figure 9. Influence of the protein conformation on the FEP accuracy for Eg5:[226] the initial FEP results show all compounds with a phenol group as outliers (enclosed in the rectangle) (*a*); result after loop remodelling. The predictions within 1 and 2 kcal mol^{-1} from the experimental value are marked by dark- and light-grey areas, respectively (*b*); the modified loop is shown in yellow (*c*). Here ΔG_{pred} and ΔG_{exp} are the predicted and experimental changes in the Gibbs energy, respectively. Reproduced with the permission of the American Chemical Society, *J. Chem. Inf. Model.*, **60**, 5457 (2020)[226] Copyright (2020).

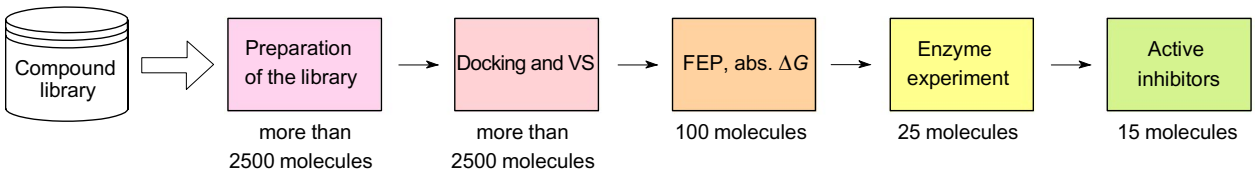


Figure 10. Algorithm of increasing the efficiency of the search for active COVID-19 protease (M^{pro}) inhibitors with FEP determination of the binding free energy for repurposed drugs.[230]

ing free energy. For example, the energy ($\Delta\Delta G$) was overestimated in the case of aromatic heterocyclic compounds (Scheme 2). Despite the fact that the force field used in this study is the most advanced version of the OPLS (optimized potentials for liquid simulations) force field, which is widely applied in the modelling of biological systems and includes a fairly large number of torsion potentials, the problem of describing some interactions remains unsolved.

(2) Structural features of the protein. It was found for the Eg5 protein that the protein structure, in particular the loop in the vicinity of the ligand, affects the prediction accuracy. After remodelling of the loop located close to phenol-containing ligands, the accuracy of FEP increased Fig. 9 *a*,*b*). In addition, the authors suggested that in some cases, the outliers may be caused by a change in the protein conformation induced by the cofactor or an additional protein domain that was not a part of the crystallization construct.

It is worth adding that the insufficient simulation time is responsible for the outliers. In relation to the development of β-secretase (BACE1) inhibitors, it was shown that the simulation that lasted for 5 ns/λ (a separate molecular dynamics simulation is carried out for each intermediate state with particular λ) gives a mean error of 0.91 kcal mol^{-1} in comparison with the experiment, and the simulation accuracy can be improved to 0.59 kcal mol^{-1} when the time is 20 ns/λ.[228] Thus, the prediction accuracy of the change in the relative binding energies by the FEP method can reach 1.0 kcal mol^{-1}. Nevertheless, there are data indicating that FEP still gives unsatisfactory results in many cases. This problem can be solved by improving the simulation characteristics such as force fields and simulation time and by development of tailored FEP protocols for particular protein targets.[229]

The FEP simulation is used, most often, to calculate minor changes in the binding free energy upon slight changes in the chemical structure of the low-molecular-weight ligand. However, it can also be used to calculate the absolute binding free energy, although the accuracy is much lower. This approach was successfully employed in a recent study[230] to correct the results of molecular docking in the search for COVID-19 protease (M^{pro}) inhibitors (Fig. 10).

The targeted FEP optimization of the ligand structure improved binding by more than 2000 times. As a result of optimization, a subnanomolar phosphodiesterase PDE-10 inhibitor was obtained (Fig. 11).[231]

A reliable prediction of the binding affinity for a ligand–receptor complex is the main goal of the drug design on the basis of the receptor structure. The 'alchemical' methods (first of all, FEP method) are popular for predicting the binding free energy in biomolecular complexes. However, highly flexible protein–ligand systems impair the accuracy of FEP calculations, mainly because of an insufficient sample size. Singh and Li[232] proposed a computational protocol that combined FEP-based calculation of the absolute binding free energy with the free energy landscape method for improved prediction of the binding free energy for flexible complexes. The proposed method was successfully applied to a set of compounds of various classes inhibiting the p53 protein–MDM2 binding.

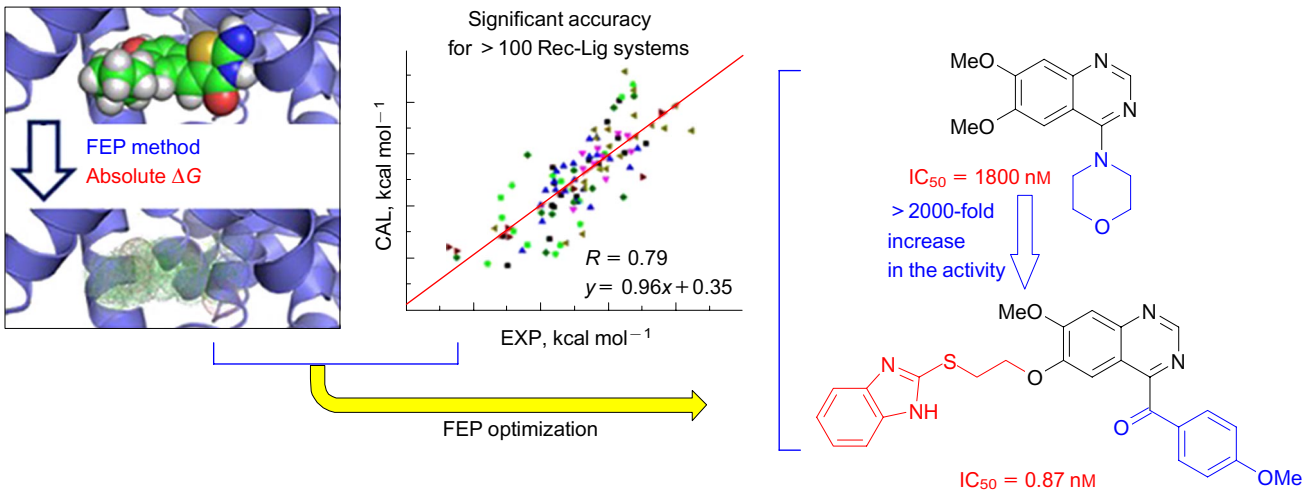


Figure 11. Scheme of the development of phosphodiesterase PDE-10 inhibitors by FEP. Reproduced with the permission of the American Chemical Society, *J. Med. Chem.*, **62**, 2099 (2019)[231] Copyright (2019).

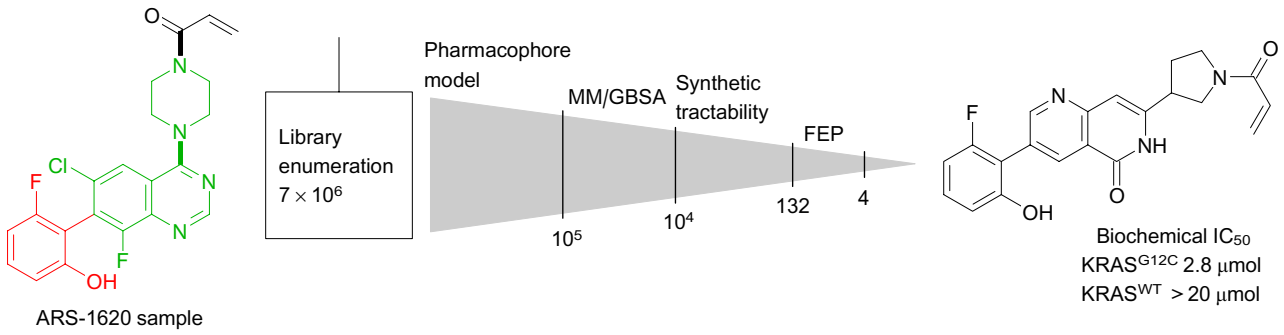


Figure 12. Scheme of the development of $KRAS^{G12C}$ inhibitors by the FEP method.[233]

The FEP method was successfully used to optimize inhibitors of $KRAS^{G12C}$, an important target for cancer therapy. Thus, a highly active inhibitor with a new type of binding was obtained (Fig. 12).[233]

Phospholipase A2 (PLA2) is the major component of the snake venom. The enzyme PLA2 catalyzes the Ca^{2+}-dependent hydrolysis of the 3-sn-phospholipid ester bonds, thus releasing fatty acids and lysophospholipids. Inside the victim's body, the PLA2 enzyme of the snake venom induces either direct or indirect pathophysiological action, in particular anticoagulant, inflammatory and neuro-, cardio- and myotoxic actions. Chinnasamy *et al.*,[234] who looked for new PLA2 inhibitors, used virtual screening and molecular docking of compounds present in the traditional Chinese medicine database and found the drug candidate TCM3290. Additional studies by molecular dynamics and FEP methods also showed the potential activity of this compound, which was subsequently confirmed experimentally.

Cyclin-dependent kinase 7 (CDK7) participates in the regulation of mRNA processing and controls the cell cycle by activating some CDKs.[235] Since the activity of CDK7 in various types of tumours is increased, CDK7 inhibitors were proposed as anticancer drugs. All inhibitors developed previously have considerable drawbacks, such as hepatotoxicity, which prevents their use; therefore, the search for new CDK7 inhibitors is a relevant task.[236]

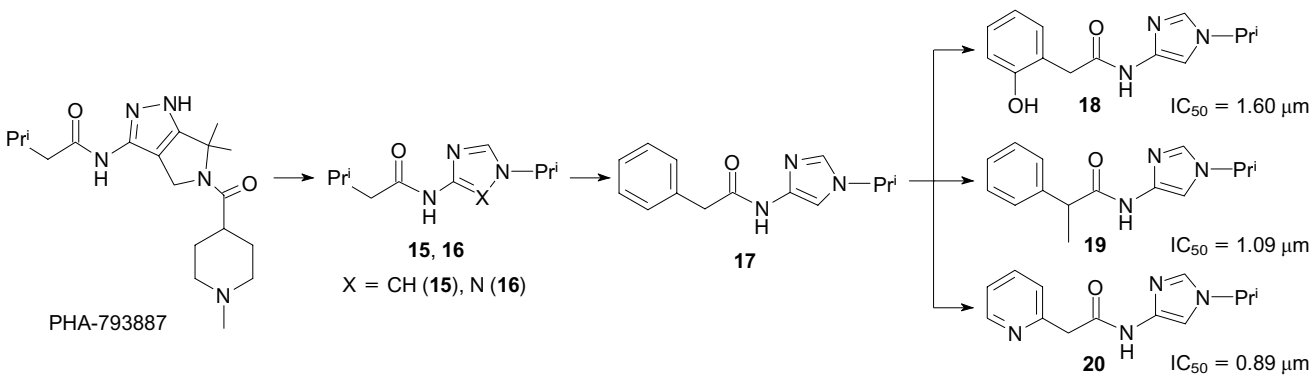
Using molecular dynamics (MD) and free energy perturbation methods (FEP/MD), a new CDK7 inhibitor scaffold free from the hepatotoxic action was developed on the basis of PHA-793887, a multi-CDK inhibitor, which was withdrawn from clinical trials because of hepatotoxic action (Scheme 3, compound **17**).[237]

In order to evaluate the effect of an additional nitrogen atom on the binding, MD simulation and the subsequent analysis of trajectories were carried out for imidazole and 1,2,4-triazole derivatives (**15**, **16**) bound in the kinase active site. It was found that the additional nitrogen atom in compound **16** does not participate in the binding. The results were confirmed by FEP/MD calculations and *in vitro* experiments. The evaluation of compounds **15** and **16** using the LiverTox website demonstrated that, whereas the probability of hepatotoxicity for 1,2,4-triazole derivative **16** was high, compound **15** was free from this drawback.[237] The replacement of the isopropyl group by a phenyl group ensured the possibility of synthetically easy insertion of substituents (structure **17**). As a result of MD simulation, three inhibitors were synthesized on the basis of the obtained scaffold (Fig. 13). According to theoretical and experimental data, the highest binding affinity for CDK7 among these compounds was inherent in compound **20**.[237]

It is known that inhibition of microtubule formation by affecting the tubulin protein causes cell death *via* apoptosis. The idea of using various regions of tubulin as targets for natural compounds proved to be a potential strategy for the treatment of haematological tumours. However, the clinical development of natural inhibitors, tubulizines, as drugs is difficult due to limited natural resources and instability of drug conjugates. In addition, the existing drug resistance and side effects, such as peripheral neuropathy, are the major problems involved in the use of existing drugs. Therefore, the development of new synthetically available tubulin inhibitors is a relevant task.[238]

In 2019–2021, several research teams reported the development of new inhibitors of tubulin polymerization using various advanced molecular modelling techni-

Scheme 3



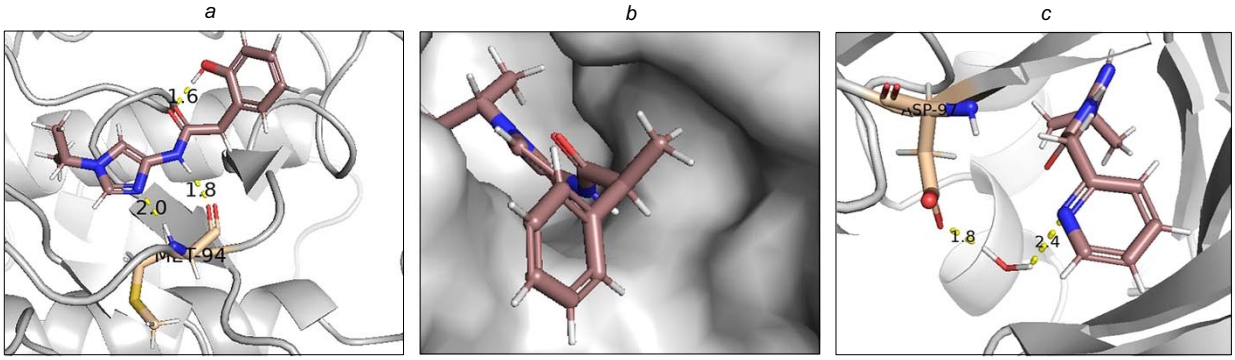


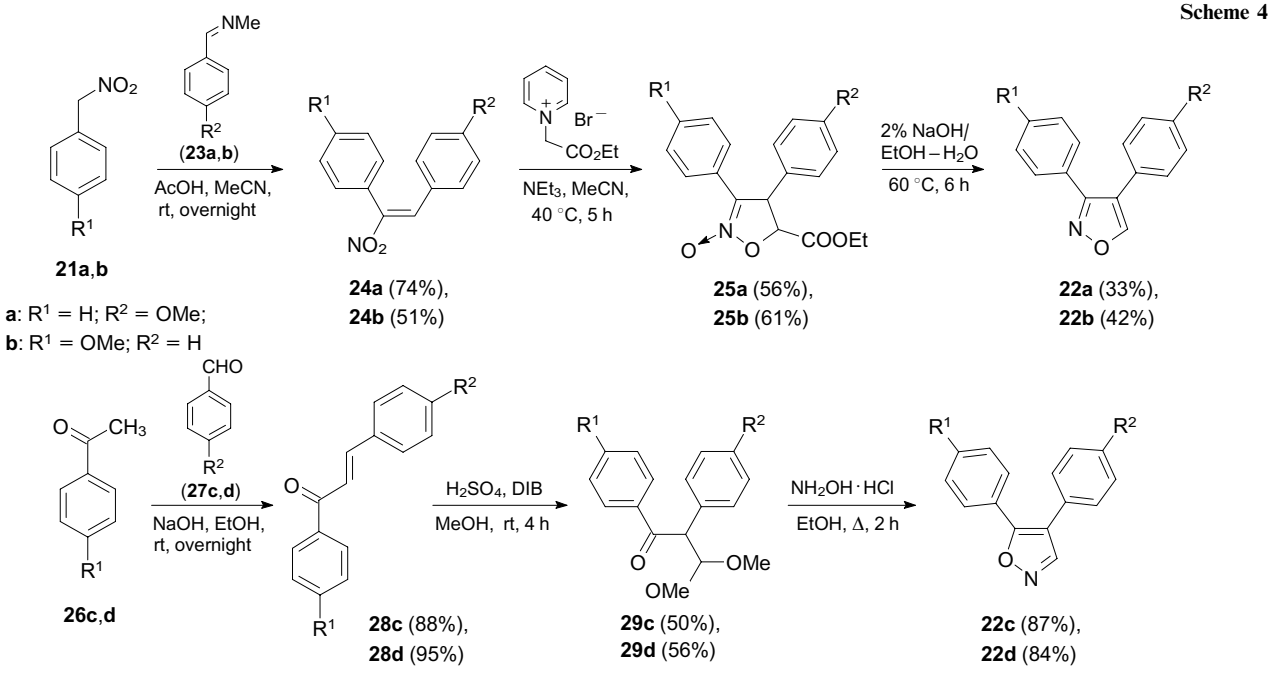
Figure 13. Binding of compounds **18** (*a*), **19** (*b*) and **20** (*c*) to the CDK7 active site.[237] Reproduced with the permission of Elsevier.

ques.[239 – 242] The MD and FEP/MD methods were applied to predict the affinity of new inhibitors of tubulin polymerization to the colchicine binding site.[243] The corresponding molecules were synthesized using readily available benzaldehydes, acetophenones and arylnitromethanes (Scheme 4). The calculated inhibitor effects were confirmed by *in vivo* studies using phenotypic analysis of sea urchin embryos. Compounds **22a** and **22c** with unsubstituted phenyl rings in the vicinity of N or O heteroatoms of the isoxazole linker exhibit antimitotic destabilizing activity against microtubules comparable with the activity of the initial combretastatin A-4, a known natural inhibitor of tubulin.[243]

Most recent advances in the application of FEP to overcome the limitations of docking. Ruiz-Blanco and Sanchez-Garcia[244] proposed an approach based on the free energy perturbation (FEP) theory and the central limit (CL) theorem for estimating the free energy of transition between the states of a system directly from the molecular dynamics trajectories of only two final states. In the proposed CL-FEP method, the energy found by calculation of the molecular dynamics trajectory with an explicit model of the solvent is used to estimate the absolute change in the free energy. In this implementation of the method, no additional parameters are introduced, and accurate estimates of the free energy can be obtained without stratification (ordering of the system objects according to strata to develop a hierarchical description of the system). The applicability of CL-FEP was first illustrated in four different type systems ranging from host – guest to protein – peptide complexes, for which the deviation from experimental values was less than 1 kcal mol^{-1}. The extended validation set included 25 host – guest and protein – ligand systems, and the mean absolute error of free energy changes calculated by CL-FEP was in the range from 1.1 to 1.4 kcal mol^{-1} between different sets; this falls within the range of accuracy of the most robust free energy approaches applied to these systems. The CL-FEP method provides an objective estimate of the free energy change and also uses the bootstrap algorithm, which allows assessing both the convergence and robustness of calculations.

Wade *et al.*[245] proposed and implemented a conceptually similar method of single-step free energy perturbation to identify those protons in the ligand the replacement of which by fluorine would enhance binding to the receptor. The accuracy of the methods was tested

Scheme 4



R^1 = H; R_2 = OMe (**c**); R^1 = OMe; R^2 = H (**d**); DIB is diacetoxyiodobenzene

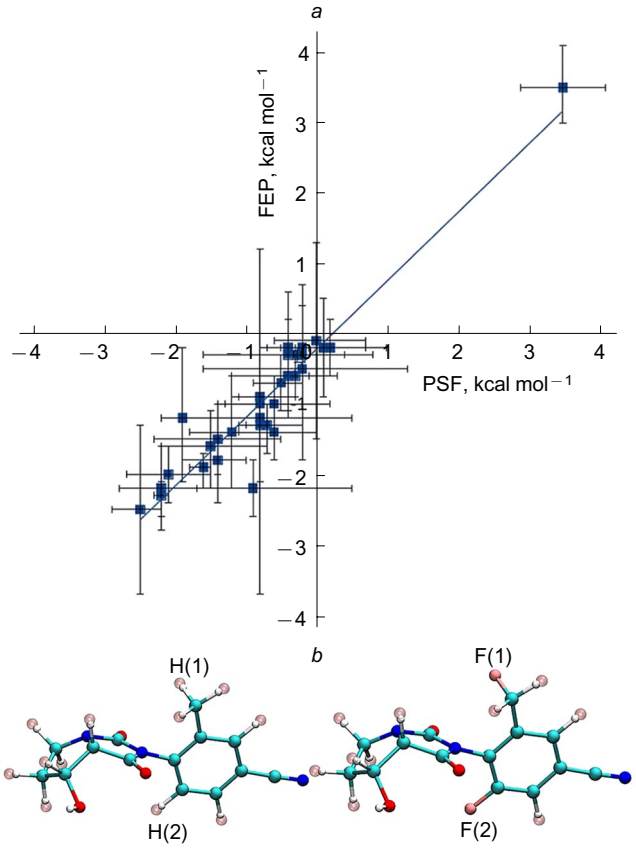


Figure 14. Accuracy of the single-stage free energy perturbation method (*a*) and an example of replacement of hydrogen atoms in the androgen receptor inhibitor by fluorine atoms (*b*).[245] Reproduced with the permission American Chemical Society, *J. Chem. Inf. Model.*, **59**, 2776 (2019).[245] Copyright (2019).

using nine systems: renin, DPP4, menin, P38, factor Xa, CDK2, AKT, JAK2 and androgen receptor (AR). Comparison of the results with the results of classical FEP showed a very good correlation, while the requirements for computational resources were reduced more than 10-fold (Fig. 14).

Free energy perturbation calculations can be used together with other methods for free energy calculation to improve their characteristics. For improving the correlation between the calculated and experimental results, Rifai *et al.*[246] combined the FEP method with the linear interaction energy for explicit calculation of the free energy and entropy of ligand solvation.

Wade *et al.*[247] described one more interesting method of using FEP calculations. Instead of the alchemical transformation of the ligand or its complete disappearance in the binding site, the authors used a genetic algorithm for local optimization of the partial charges on ligand atoms in order to maximize its binding affinity for the protein. The greatest change in the charge was restricted in such a way that the dot product of the original and final charge vectors was no less than 0.95. The pattern of charge changes thus obtained can clearly indicate the preferable positions for introduction of electron-donating or -withdrawing substituents into the ligand to increase the receptor affinity (Fig. 15).

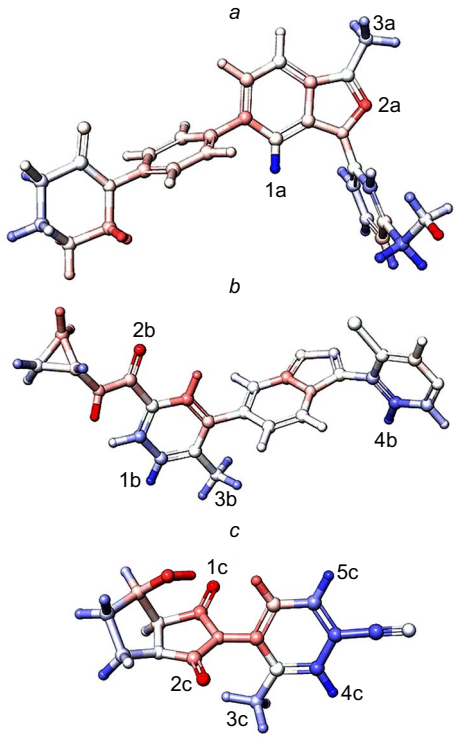


Figure 15. Ligands of FXa (*a*), P38 (*b*) and AR (*c*) receptors. The atoms are coloured depending on the change in the charge relative to the initial partial charges. More electronegative atoms are shown in blue, more electropositive atoms are in red. The areas chosen for chemical modification are numbered.[247]

6. Quantum chemical methods in the predictive modelling (Stage 3a)

The basic principles of molecular design formulated in the literature[248,249] also apply to calculations that include a quantum chemical component.[250,251] Computational studies by no means always end in the synthesis of the desired product; however, studies on quantum chemical (or hybrid, that is, involving molecular mechanics methods) modelling in the targeted synthesis have been increasing geometrically in recent years. We will consider particular examples that have not been included in relevant reviews so far.

6.1. Modelling of the mechanism and products of catalytic (including biocatalytic) reactions

Modelling of a catalytic reaction, that is, prediction of active structures, is impossible without knowledge of the reaction mechanism and dynamics. Apart from the static problem of elucidation of the energy of ligand–target coordination, in the case of biological activity (the first step of biocatalysis), it is necessary to consider dynamic factors: the formation and decay of complexes of reactants and products with the target (catalyst). In the molecular dynamics modelling, quantum chemical methods supplement or replace molecular dynamics simulation. The combined use of quantum mechanics and molecular mechanics (QM/MM) provides a more comprehensive picture of reactivity than separate calculations by either of these methods, and the modelling of the coordination step and the mechanism of the catalytic reaction potentially allows one to attain the desired selectivity of the targeted synthesis.

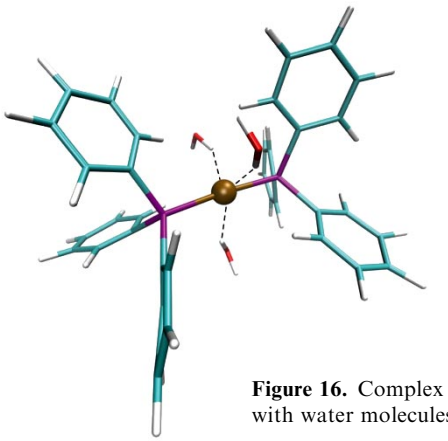


Figure 16. Complex of $Pd(PPh_3)_2$ with water molecules.[255]

Some typical examples were reported by Khrenova *et al.*[252, 253] and Ankudinov and Perekalin.[254] Khrenova and Tsirelson[253] used the QM/MM method to demonstrate that the formation of an intramolecular hydrogen bond is the rate-limiting step of the enzymatic hydrolysis of cephalosporin antibiotics by metallo-β-lactamase. One more characteristic example is the use of a similar approach to discriminate between active and inactive substrates in the enzymatic peptide bond hydrolysis by matrix metalloproteinase 2.[252]

A combination of molecular mechanics and quantum chemical methods was used to demonstrate the role of water in the destruction of the $Pd(PPh_3)_2$ complex in the last step of the Suzuki – Miyaura reaction. The quantum mechanical optimization of the $Pd(PPh_3)_2$ complex with water demonstrates that hydrogen rather than oxygen is the reacting atom of water, that is, the Pd···H – O bond is formed (Fig. 16).[255]

The Suzuki – Miyaura reaction, like the Diels – Alder reaction, is a popular object for studying the mechanisms of catalytic transformations,[256 – 259] particularly, the solvent effect on the reaction pathway. The role of water in the solid-phase Suzuki – Miyaura reaction was also clarified by Pentsak and Ananikov.[260] Thus, solvation with protic solvents, which is most often considered in a simplified form, can be an important detail of understanding of the mechanism and products of a catalytic reaction.

6.2. Asymmetric synthesis

Presently, various methods for preliminary modelling of asymmetric synthesis, including statistical methods, are widely used for the preparation of desired enantiomeric products.

— Machine learning for the construction of 3D-QSAR descriptors for the selection of the optimal solvent for asymmetric synthesis.[261]

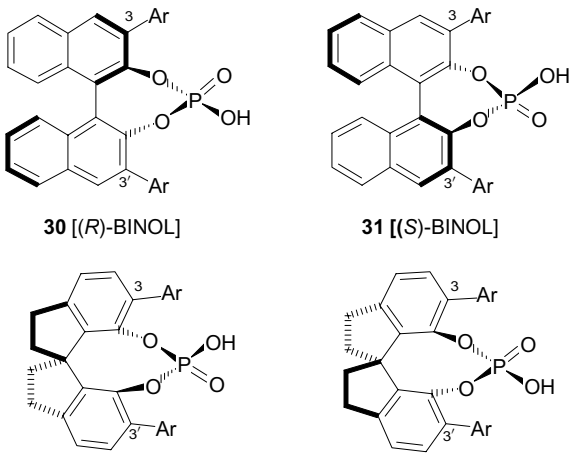
— Pre-filtering methods of 3D-QSAR in the catalyst optimization:[262] training set analysis, verification of descriptors and construction of the quasi-model by selecting a subset of active catalysts (Fig. 17). The idea of applicability of QSAR principles in any molecular modelling is successfully utilized from time to time at least for prefiltering of potentially active structures. For example, Zahrt *et al.*[263] described in detail an approach to application of computational methods in this case, which did not differ from the classical QSAR analysis of properties of a ligand (but not a target):

(a) generation of training sets of chiral catalysts relying on empirical data;

(b) separate parameterization of the whole structure and some flexible substituents;

(c) building of a mathematical model based on the quasi-dynamic energy – conformation curves;

Structures 30 – 33



30 [(*R*)-BINOL] **31** [(*S*)-BINOL]

32 [(*R*)-BINOL] **33** [(*S*)-BINOL]

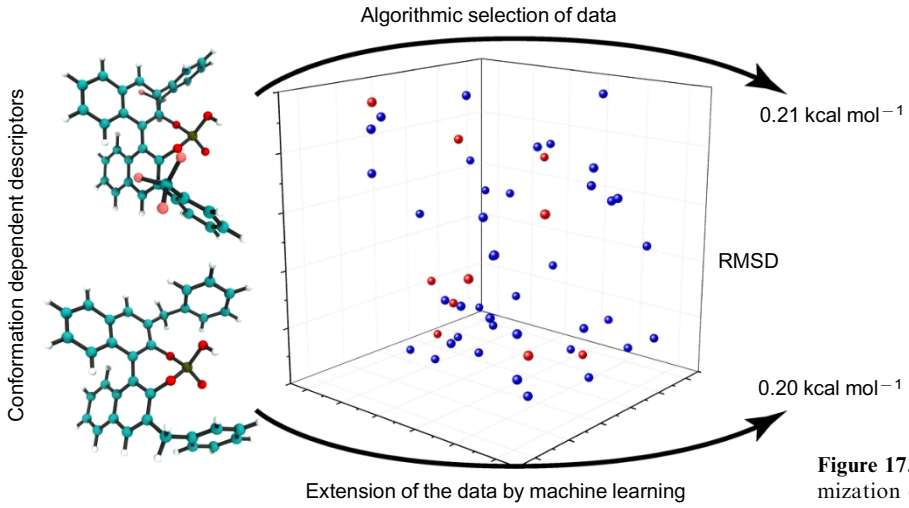


Figure 17. Application of 3D-QSAR for optimization of enantioselective catalysts.[262]

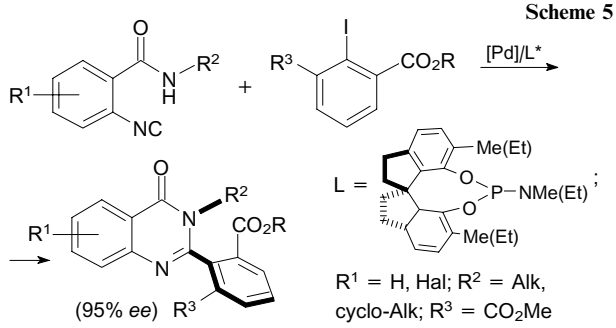
(d) prediction of the optimal catalyst structure by minimizing the potential energy of the ligand.

It can be seen that the use of authors' acronyms such as QSSR (quantitative structure–selectivity relationships), principles of description of molecular interaction surfaces [molecular interaction field (MIF)] (MEP [264] is the same by definition, see above) does not change anything in the essence of optimization of targeted synthesis by statistical methods.

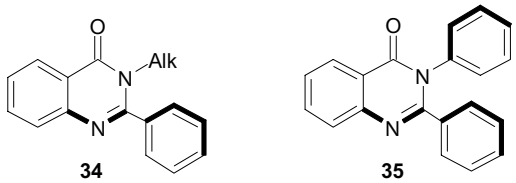
— Targeted asymmetric synthesis in the presence of chiral catalysts,[265,266] enantioselective catalysis of the Minisci reaction,[267] synthesis of indole derivatives,[268] increase in the efficiency of asymmetric catalysis [269] involving 1,1′-bi-2-naphthol (BINOL, **30**, **31**) and 1,1′-spirobiindane-7,7′-diol (SPINOL, **32**, **33**) — derivatives of chiral phosphonic acids.

A similar example is the use of a chiral ligand for the formation of axially (N_2C-C and $N-R^2$ bonds) asymmetric rings (Scheme 5),[270] which are subsequently converted to the target asymmetric aryl- and diaryl-substituted 4-quinazolones **34**, **35**.

Scheme 5



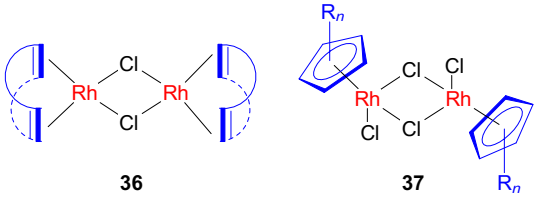
Structures 34, 35



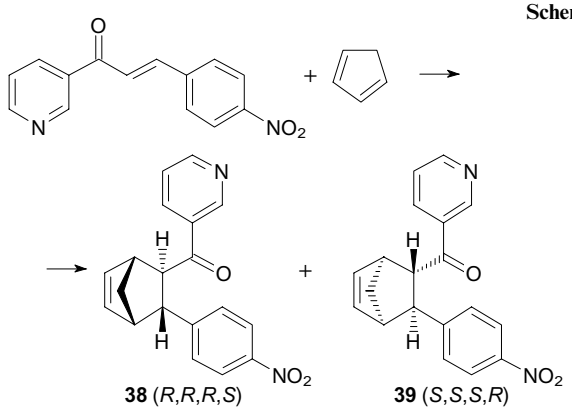
— Asymmetric catalysis based on rhodium complexes with asymmetric dienes **36**, **37** is reported in a review.[271] The same publication describes methods for the synthesis of these complexes with various structural types of diene ligands and dependence of the composition of products of asymmetric catalysis on the structure of the asymmetric ligand.

The dependence of the product composition in Cp*Rh(III)-catalyzed cyclization and cyclopropanation on the interplay between the steric features of the ligand and redox potential of the ligand–target complex was discussed [272] on a large number of examples.

Structures 36, 37



Scheme 6

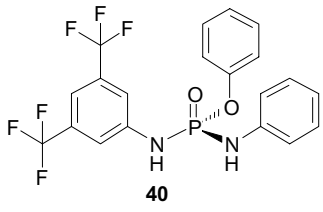


— Asymmetric synthesis using chiral auxiliary reagents.[273] As an example of docking-based optimization of the catalyst for a particular reaction, mention can be made of practical modelling of asymmetric synthesis involving asymmetric induction.[274]

The Diels–Alder addition of (*E*)-3-(4-nitrophenyl)-1-(pyridin-3-yl)prop-2-en-1-one to cyclopentadiene (Scheme 6) affords two enantiomeric products — (*R*,*R*,*R*,*S*) (**38**) and (*S*,*S*,*S*,*R*) (**39**). The calculation of the energy of the complex formed by the chiral oxazaborolidine catalyst with the reaction products (molecular dynamics + DFT) showed that the free energy of interaction with the catalyst is lower by 5.75 kJ mol^{-1} for compound **38** than for compound **39**. This attests to enantioselectivity of the reaction with the predominant formation of product **38**, which was actually demonstrated in experiment.[274]

Wåhlander *et al.*[275] predicted the composition of the Diels–Alder reaction products formed in the presence of chiral phosphamides **40**, which form a complex (hydrogen bonding of amide groups) with the starting acrolein.

Structure 40



— Hybrid method combining two preceding ones: an asymmetric catalyst is formed *in situ*, for example from $[Rh(C_2H_4)_2Cl]_2$ and bicyclo[2,2,2]octadiene, during asymmetric arylation of *N*-tosylarylamines [276] or [4 + 2]-cycloaddition of (*R*)-α-phellandrene to methyl propiolate.[277]

— Chiral solvents,[278,279] in particular chiral ionic liquids.[280]

— Spontaneous asymmetric synthesis.[281]

— Heterogeneous catalysis by metal-organic frameworks (MOFs).[282]

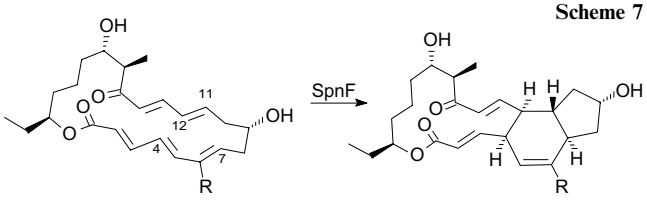
An interesting example of applying a combination of various methods, including FEP, is described by Velez *et al.*[283] The force field for simulation of ionic liquids [optimized potentials for liquid simulation-ionic-liquid virtual site (OPLS-VSIL)] was found to correctly describe the properties of the bulk phase and the local ionic interactions for a wide range of imidazolium-based ionic liquids. The authors studied the Diels–Alder reaction between cyclo-

pentadiene and methyl acrylate in the [bmim]PF_6 ionic liquid (1-butyl-3-methylimidazolium hexafluorophosphate) to validate the OPLS-VSIL force field. Combined *ab initio* quantum mechanical and molecular mechanical calculations (QM/MM), together with free energy perturbation and Monte Carlo simulations (FEP/MC) using the M06-2X/6-31G(d) and OPLS-VSIL functionals resulted in the activation free energy of 14.9 and 16.0 kcal mol^{-1} for *endo-cis* and *exo-cis* Diels – Alder reaction pathways, respectively (the experimental $\Delta H^{\ddagger}$ value is 14.6 kcal mol^{-1}). The *endo*-selectivity was correctly predicted, with the calculated *endo* : *exo* ratio being 73 : 27%. For comparison, the previous version of the OPLS-2009IL force field also gave $\Delta G^{\ddagger}$ of 14.9 kcal mol^{-1} for the *endo*-reaction, but the OPLS-2009IL field could not adequately model the highly organized solvent interactions occurring between the cation and the transition state in the Diels – Alder reaction.

6.3. Examples of modelling of natural catalysis mechanisms

The biosynthesis of spinosin, a natural insecticide,[284] particularly using the enzyme SpnF,[285] can be considered as the first example of enzymatic catalysis of the Diels – Alder reaction. It was found that in this reaction, SpnF catalyzes the intramolecular [4 + 2]-cycloaddition of 22-membered macrolactone to give a six-membered ring (Scheme 7).[286]

Scheme 7



The structure of SpnF resembles that of SAM-dependent methyltransferase; however, the participation of SAM in reactions of this type has not been studied before. Only few enzymes that are known to catalyze pericyclic reactions were reported; however, for none of them was catalysis of the Diels — Alder reaction reliably demonstrated.[287] Since the 3D structure of SpnF has been determined experimentally, the catalytic activity can be modelled in the explicit form.[288] If the SpnF-catalyzed reaction actually follows the [4 + 2]-cycloaddition mechanism (see Scheme 7), a detailed study of this reaction is of enormous interest for both fundamental and applied science.[289]

The reaction shown in the Scheme is formally a Diels – Alder reaction; it is not surprising that for long period of time, there were no doubts concerning the reaction mechanism. After the substrate conformations were studied by molecular mechanics, it became possible to identify the catalytic site of the protein in accordance with the putative mechanism (Fig. 18).

However, under the assumption that conjugated double bonds present in the substrate molecule can be involved in the cyclic transition state, thus changing the ring size, the reaction mechanism may significantly differ from the mechanism suggested initially.[285]

The theoretically possible alternative mechanisms of this reaction were investigated in several studies both theoretically using quantum mechanics methods[290 – 292] and experimentally using the kinetic isotope effect.[293] A recent publication[294] describes an exhaustive search for the possible transition state conformations for four proposed mechanisms of the catalytic reaction: Diels – Alder, bis-pericyclic, biradical and alternative Diels – Alder reaction mechanisms. The possible mechanisms of SpnF-catalyzed [4 + 2]-cycloaddition — Diels – Alder (DA) reaction, bis-pericyclic reaction (Cope rearrangement, BPC), biradical reaction (BR) and alternative Diels – Alder reaction (altDA) mechanisms — are shown in Scheme 8.

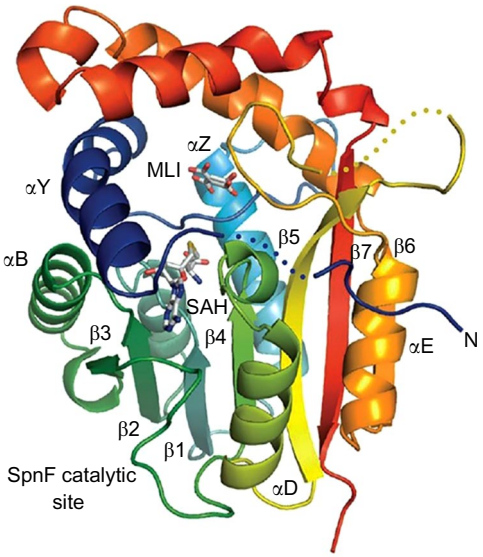


Figure 18. Catalytic site of SpnF for the postulated [4 + 2]-addition mechanism.[286] Reproduced with the permission of Springer Nature.

Quantum chemical modelling of more than 750 transition states (TS) for all mechanisms shown in Scheme 8 demonstrated that the Cope (BPC) transition state is 7 kcal mol^{-1} lower in energy than the transition state of the Diels – Alder (DA) reaction. Hence, the preferable mechanism is bis-pericyclic:[295]

$$\mathbf{S} \rightarrow [\mathrm{BPC}] \rightarrow \mathrm{I} - [6+4] \rightarrow [\mathrm{Cope}] \rightarrow \mathbf{P}$$

Similar calculations of the mechanisms and the corresponding product compositions for the Diels — Alder cyclization were reflected in a review[296] and in original papers.[297 – 299] The Cope rearrangement in the last step was also confirmed by Jeon *et al.*,[300] who used a similar DFT calculation of the potential energy surface of the reaction.

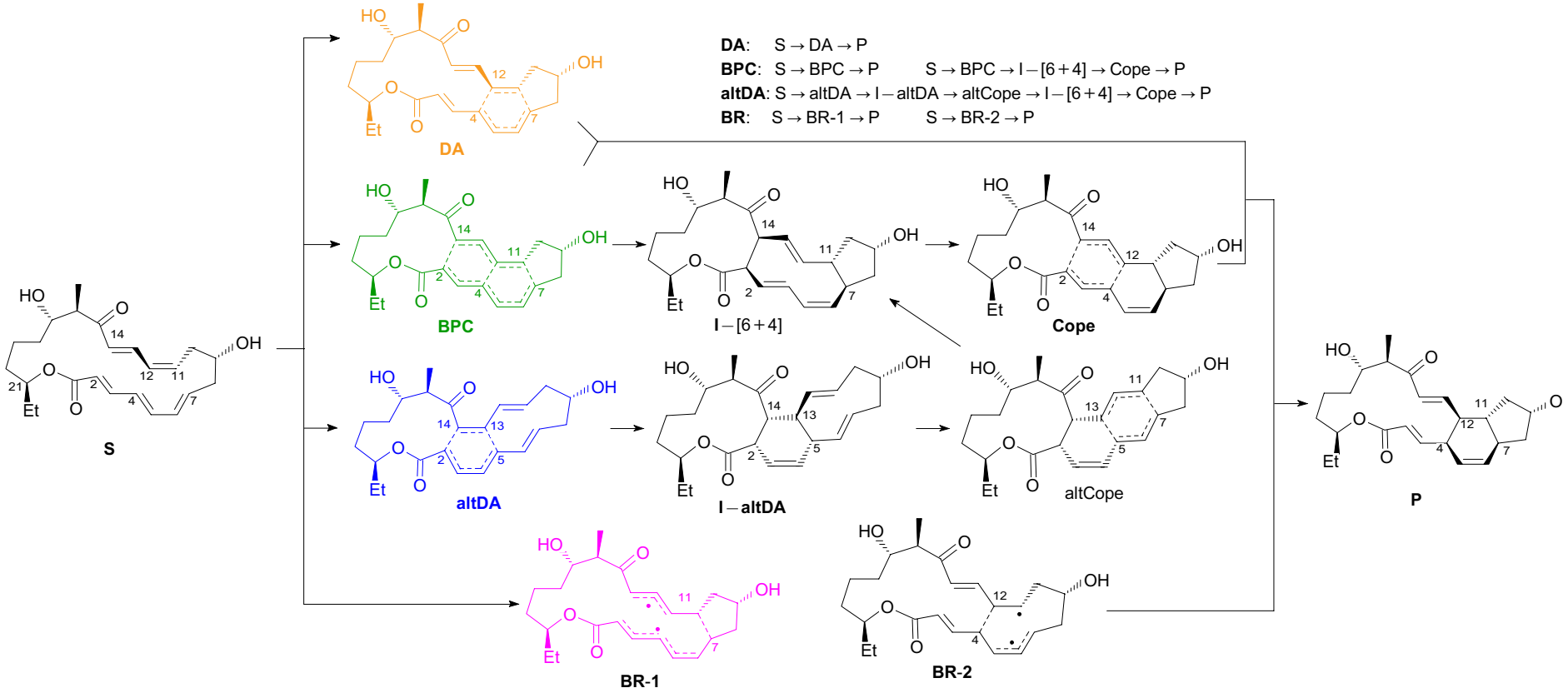
The fact that the mechanism of transannular [6 + 4]-cyclization is real both in the absence of a solvent and in water was confirmed[295, 301] in relation to the biosynthesis of heronamide A (Scheme 9).

It was shown[294] that the ratio between the [6 + 4] and [4 + 2] mechanisms is 16 : 1 in the gas phase and 56 : 1 in water.

Examples of computer simulation of enzymatic catalysis are rather numerous.[302 – 304] Tan *et al.*[305] reported an example of practical application of the calculated Diels – Alderase enzyme catalysis for planning the synthesis of varicidins (Scheme 10).

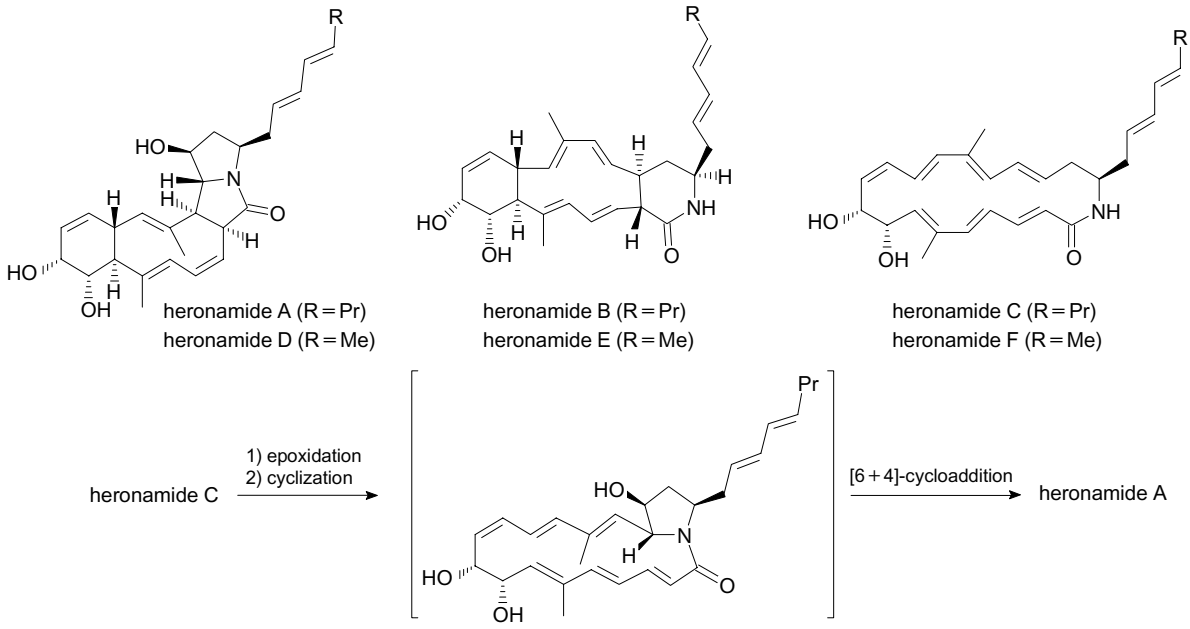
The real mechanisms of catalytic (including biocatalytic) reactions in the living nature provide excellent possibilities for gaining a more comprehensive natural science view of the world.[306]

Scheme 8

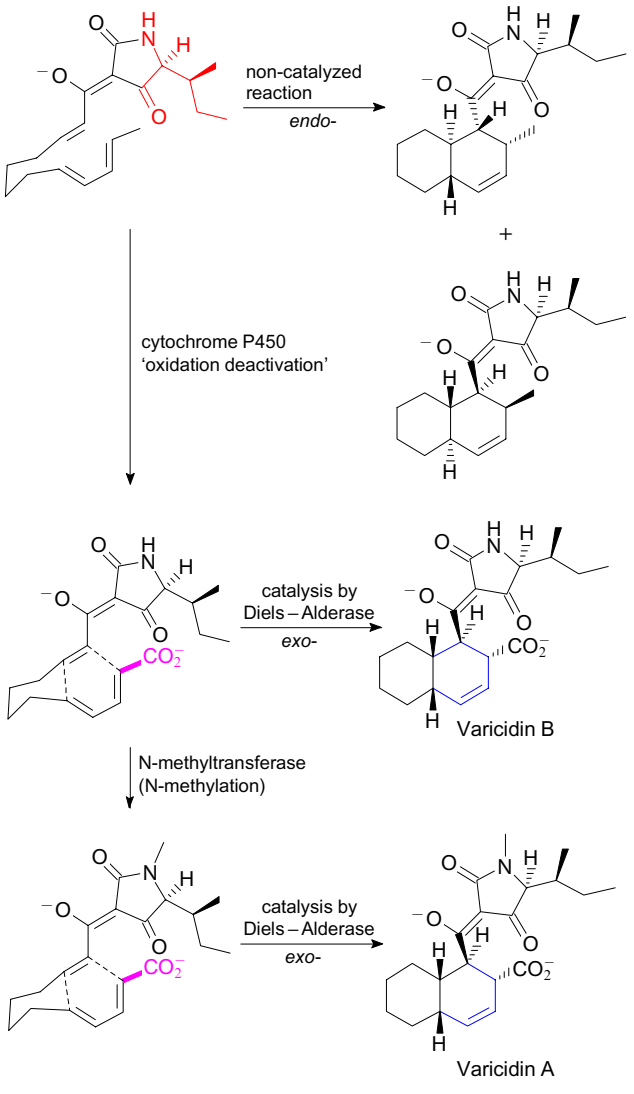


Possible mechanisms of SpnF-catalyzed [4 + 2]-cycloaddition: Diels–Alder (DA) reaction, bis-pericyclic reaction (BPC), biradical reaction (BR) and alternative Diels–Alder reaction (altDA) mechanisms. The Scheme is reproduced with the permission of the American Chemical Society, *JACS*, **139**, 3942 (2017).[294] Copyright (2017).

Scheme 9



Scheme 10



7. Combination of computer simulation and physicochemical methods (Stage 3b)

The structure – property relationships studied by molecular mechanics methods, even with the described extensions and refinements, cannot model transition states of catalytic reactions with a sufficient accuracy. Therefore, in this case, a toolbox of theoretical calculations in combination with physicochemical methods such as X-ray diffraction, NMR and electronic spectroscopy can be applied.

The following combinations can be mentioned as examples of this integrated approach:

— methods of quantum chemistry combined with high-precision X-ray diffraction experiments[307] markedly extended the range of investigation objects and problems that can be solved, which vary from more precise determination of protein structures to gaining any desired physicochemical characteristics of crystals. Without physicochemical data on the exact electronic structure of protein targets and catalyst targets,[308, 309] any other calculations except for QSAR are impossible;

— the structure – spectrum relationships, the use of advanced techniques of NMR spectroscopy for the design of biologically active compounds and their subsequent targeted synthesis are described in relevant reviews[310, 311] and in publications on NMR spectroscopy of transition metal complexes such as cisplatin, transplatin[312] and their derivatives,[313] four-component relativistic DFT calculation of relative ^{1}H, ^{13}C, ^{15}N and ^{17}O NMR chemical shifts of iron, cobalt and nickel bis-glycinates[314] (Fig. 19).

Quantum chemical calculation of a structure with a tautomer form complementary to adenine, that is, the antitumour agent[315] 5-fluorouracil, was confirmed by gas phase electron diffraction (GED) data: the H – N(3) – C(4)=O moiety was identified as the active group responsible for the biological activity.[316]

Ananikov and co-workers[317, 318] developed a set of procedures for analysis of transition metal-based catalytic systems. The integrated approach included innovative procedures of mass-spectrometric analysis of reaction mixtures on a real time basis for the direct monitoring of the

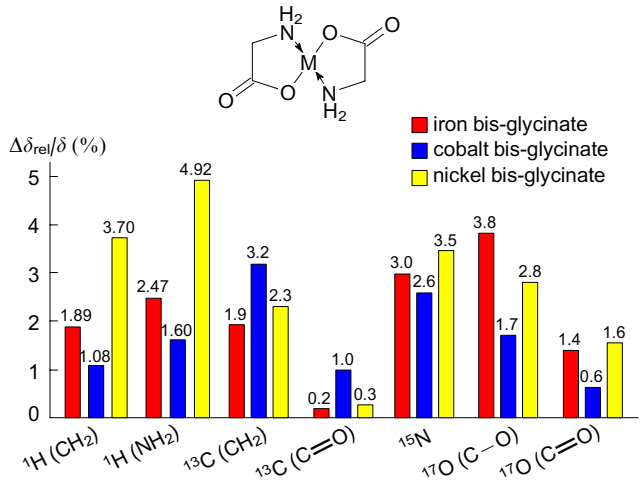


Figure 19. DFT calculation of the relative ^{1}H, ^{13}C, ^{15}N and ^{17}O NMR chemical shifts of iron, cobalt and nickel bis-glycinates.[314] Reproduced with the permission of Elsevier.

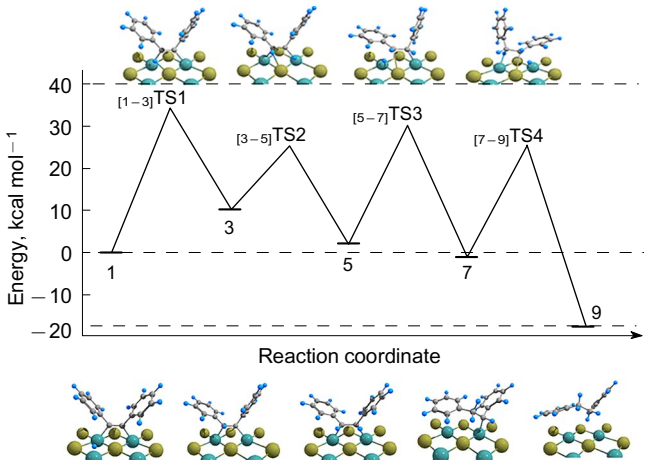


Figure 20. Energy profile of hydrogen addition in the first stage of hydrogenation of the diphenylacetylene molecule on the Pd–Ag catalyst surface.[321] Reproduced with the permission of Elsevier.

conversion of dynamic catalytic systems from the nanoscale state to the molecular state. The data of mass-spectrometric measurements are supplemented by studying the behaviour of substrates by NMR spectroscopy, which gives a complete view of the molecular processes taking place in the homogeneous phase. Transition from a homogeneous to heterogeneous system is monitored by transmission and scanning electron microscopic examination of the samples. The reverse transition is detected by mass spectrometry.

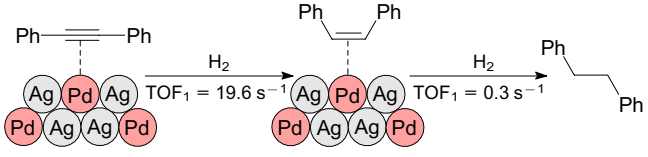
8. Approaches to modelling of organic heterogeneous catalysis

8.1. Calculation of the ligand–solid catalyst conformational model

A reasonable continuation of the mechanistic studies of catalytic reactions and transformations is theoretical study of heterogeneous catalysis using objects convenient for calculations, *e.g.*, single-atom catalysts. For this purpose, a study of diphenylacetylene hydrogenation on single-atom palladium/silver catalysts was undertaken. It is known from experimental data that the rate of alkyne to alkene hydrogenation on these catalysts is much higher than that for the alkene/alkane pair.[319,320] The mechanistic study of diphenylacetylene hydrogenation on the surface of single-atom Pd/Ag catalyst was carried out[321] by density functional theory (Scheme 11, reproduced with the permission of Elsevier). The structure of the catalytic particle was modelled as a cubic packing of alternating palladium and silver atoms.

The reaction was modelled as a successive stepwise addition of two hydrogen molecules to a diphenylacetylene molecule. The main goal of the study was to calculate the structural and energy parameters of the transition states for

Scheme 11



all four steps of the reaction (Fig. 20). The calculations showed that diphenylacetylene is adsorbed and activated on two neighbouring sites of the Pd–Ag–Pd palladium clusters.

The obtained results on the effect of ligand coordination in the heterogeneous ligand–target (catalyst) complex on the selectivity of catalytic synthesis performed on single-atom catalysts such as Pd–Ag[322,143] Pd–Cu,[323,324] Pd–Au,[325,326] Pd–Ni[327] and Pd–Zn (Ref. 328) indicate that preliminary modelling can be used to adequately predict the course of the reaction.

In more complex cases where multiatomic catalysts are used, researchers often turn back to statistical QSAR methods. This can be exemplified by the transition metal catalysis of ethylene polymerization by iron bis-iminopyridine complexes[329,330] in combination with quantum chemical parameters used for predicting the activity of neodymium carboxylates in isoprene polymerization.[331]

A review[332] addresses the results of studying palladium-based bimetallic catalysts, most often, by the molecular design of the ligand–catalyst complex. The selectivities of various target structures complementary to the ligand activity were determined, together with the effects of deformation of the complex, ensemble effects and support effects; the discussion of reaction mechanisms determined by the structure of the complex was summarized for selective hydrogenation, CO oxidation, methane combustion, electrocatalytic reactions, reduction of CO_2, hydrogen production, biomass upgrading and fine chemical synthesis. The structure–property principles were summarized for minimizing the catalyst amount.

8.2. 3D printing as a method for structuring of catalysts for targeted synthesis

The morphology of heterogeneous catalysts can affect their characteristics. However, the standard methods for catalyst manufacturing such as extrusion or pelletizing offer few opportunities for adjustment of the catalyst structure and activity for determining the reaction pathway.

The catalyst modelling followed by 3D printing of the ready structure according to the model is little addressed in the literature. Meanwhile, 3D printing allows most precise determination of the structure and the sequence of compo-

nents that form the ready product, *e.g.*, a heterogeneous catalyst. In recent years, particularly the 3D printing has been actively used to produce catalysts with controlled topology. The structure of the catalytic material can be optimized for any type of reactor including microreactors.[333, 334]

In the vigorously increasing number of publications devoted to additive manufacturing, particularly 3D printing, one can distinguish [36] two types of 3D-printed products:

(1) printing of inert materials of various topology for conducting reactions or for biomedical purposes. The role of design of the 3D-printed porous framework and structure – function relationships for synthetic metal, polymer and ceramic biomaterials are discussed by Kelly *et al.*;[335]

(2) printing of reactants and catalysts with different degrees of binding — composites or pure reactants.

For targeted catalytic synthesis, it is of interest to consider the relationship between the structure of the printed catalyst and the reaction pathway (products). As an example, consider catalytically active copper particles regularly distributed in an alumina matrix (Ref. 336): the solid support (heat-treated Al_2O_3) provides chemical, thermal and mechanical stabilization of immobilized catalytic particles; this, in turn, reduces contamination of the reaction product with the catalyst (copper), which is important for the use of these catalysts in medicinal chemistry. One more example is provided by palladium particles in a 3D printed supported solid catalyst for the Suzuki – Miyaura cross-coupling reactions.[337] The catalyst consists of polypropylene as the support and palladium nanoparticles on silica as a catalytically active component.

Formation of a regular catalyst structure with a specified spacing is quite important for catalysis, especially in relation to the catalytic properties of MOFs. The 3D printing of MOFs is described in papers dealing with gas separation and storage [338, 339] and in a number of reviews.[340 – 342] However, sintering (analogous to laser printing) during 3D printing does not facilitate the formation of a regular matrix of the catalyst. A way to avoid sintering by varying the composition of ink (analogous to the ink printing) was developed. For example, Grande *et al.*[343] reported a 3D printed catalyst manufactured using a non-aqueous ink formulation containing hydroxypropyl cellulose and boehmite [γ-AlO(OH)] with highly porous regular MOF UTSA-16 as the active component.[344]

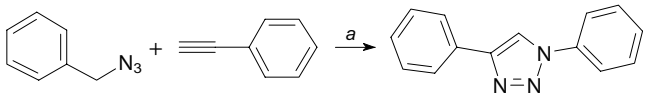
An example of another application field of 3D printing is catalytic destruction of lignocellulose in corn cobs with xylanase (a hemicellulase enzyme). This process is unfavourable for industry due to xylanase low stability, high consumption and difficulty of disposal in the homogeneous version of the reaction. The structural programmability of xylanase 3D printing on calcium alginate [345] allows manufacturing of the active component as particles of any shape: spheres, hooks, quadrangles or hexagons. Xylanase immobilized in this way recovers 60% of the catalytic activity after being used seven times, which improves the process efficiency.

The kinetic aspect of the effect of 3D printing on the hydrolysis of cellulose was implemented [346] as a multisectional magnetic stirrer 3D printed from acrylic acid (acid regulator of the reaction) and 1,6-hexanediol diacrylate (hydrophobic binding sites). The reaction hydrodynamics: tilt of the stirrer blades, speed and component ratio for control of the catalytic activity were preliminarily simulated on a computer.

8.3. Modelling of the active component by 3D printing

A precise spatial positioning of chemically reactive objects and endowing them with the desired degree and sort of functionality (which is characterized most unambiguously with the term 'molecular design') as applied to 3D-printed products is the most promising area of target-oriented catalysis.[347, 348] The expression 'direct 3D printing of catalytically active structures' was illustrated by Manzano *et al.*,[349] who used this expression as the title for their paper, in relation to photopolymerization of bifunctional monomers in a stereolithographic 3D printer. The vinyl group was used to accomplish a 3D assembly of the polymer. During printing, functional groups (carboxyl, amine) were attached to catalyze the Mannich reaction, aldol condensation and cycloaddition. The Mannich reaction was carried out in the spectrophotometer cell made of the catalytic material. The cyloaddition reaction of benzyl azide with phenylacetylene was conducted in a microfluidic system (click reaction) in the 3D-printed catalytically active device of 3 cm size. Scheme 12 shows the Huisgen cycloaddition reaction catalyzed by 3D-printed copper carboxylate on the polymer (AL is the Ames Laboratory in which the experiment was carried out).[349]

Scheme 12



(*a*) AL-$(COOH)_2$ – Cu, PPh_3

One more example of the cross-linking of the catalytic site and the matrix *via* the stereolithographic printing are MOF-based catalysts with Zr porphyrin complexes as the catalytically active sites for 3D controllable photopolymerization.[350]

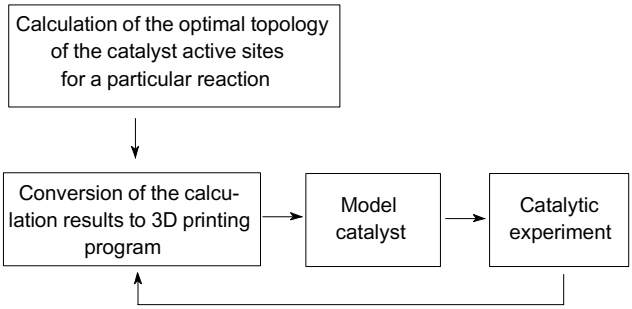
Certainly, a catalyst can also be modelled empirically by successively bringing the structure to the required activity using a feedback. An example of empirical modelling is structuring of 3D-printed zeolite bearing a catalyst for methanol conversion to dimethyl ether.[351] The feedback tuning of the zeolite porosity and the density control of the acid sites of the catalyst, silicoaluminophosphate (SAPO-34), resulted in 96% selectivity to the target product.

Preliminary computer simulation accelerates the fabrication of a catalyst with the required activity. Using quantum chemical calculations, it is possible to determine configuration of the catalyst active sites with a sufficient accuracy, for example, for structures based on 3d elements.[352]

Using the modelling strategy described in Sections 2 and 3 of the present review, it is possible to obtain a fairly evident diagram for application of 3D printing in the computer modelling of catalysis (Scheme 13).

Experimental testing can provide a feedback for tuning the object of 3D printing. It is evident that all examples presented in this Section would have given a markedly more predictable results if the geometric and physicochemical parameters of the 3D printed object and the ligand – catalyst complex had been predicted before. However, these studies

Scheme 13



are only starting to appear. For example, the density functional theory (DFT) calculations and molecular dynamics (MD) simulations of the oxygen adsorption on nanopores of modified graphene oxide-incorporated 3D-printed polymers [353] showed the O_2 transfer dynamics that completely corresponded to the experimental results obtained by IR, Raman and X-ray photoelectron spectroscopy.

9. Conclusion

Thus, this review considers molecular modelling methods for the subsequent targeted synthesis of organic compounds. The inventory of computational methods for the use in various fields of modern chemistry — organic synthesis, catalysis, biocatalysis, *etc* — is presented. The integration of the principles of preliminary modelling for the synthesis of target compounds led to the following conclusions.

The efficiency of targeted synthesis is determined by the knowledge of the structure and activity of the desired product. This knowledge substantially reduces the necessity of checking compounds to be synthesized to obtain the most active one. The resource intensity of the synthesis and analysis (especially biological activity analysis) of the products simultaneously decreases. The promising model structure formed on the basis of computational methods is validated exclusively by the subsequent synthesis and testing of the desired activity.

The modelling of the product composition is based on

— statistically treated available data on the ligand activity, where ligand is a molecule with the desired activity,

— caslculated information on the ligand–target complex structure, where target is a biological macromolecule or a catalyst,

— calculation of the reaction profile determining the composition of products of the targeted synthesis. The best effect is attained upon combination of the preliminary selection of the model by simpler methods (statistical calculations, classical docking) with the subsequent refinement and extensions to classical docking, quantum chemistry methods and molecular spectroscopy.

The relationship between the objects and methods of molecular modelling considered in this review is illustrated in Fig. 21.

Examples of the preparation of compounds with a specified activity that showed the desired properties in the validating experiment demonstrate the efficiency of computational methods in elucidation the structure–property relationships for the products of organic synthesis in the relatively small number of selected areas addressed in the review. However, there are grounds to believe that the introduction of preliminary predictive modelling as the first stage of targeted synthesis is a fairly universal method for optimizing the synthesis.

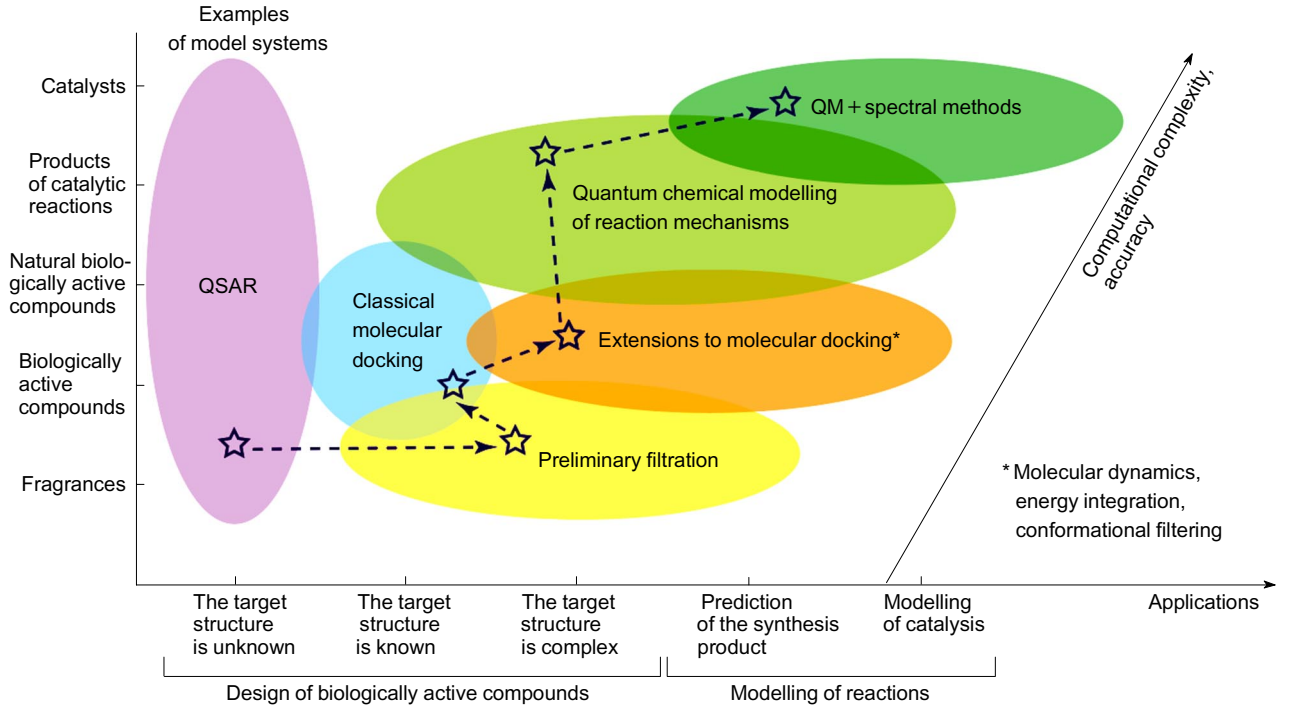


Figure 21. Relationship between the molecular modelling objects and methods.

10. List of acronyms and designations

AChE — acetylcholinesterase,
ADME — absorption, distribution, metabolism, excretion,
ALK — anaplastic lymphoma kinase,
AMPA — α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid,
AmpC — beta-lactamase,
ANN — artificial neural networks,
AR — androgen receptor,
AUC — area under the ROC curve,
BACE1 — beta-secretase 1,
BBU — bicyclic bis-ureas,
BCR/ABL — chimeric tyrosine kinase formed upon fusion of the Abelson tyrosine kinase (ABL) gene and breakpoint cluster region (BCR) gene,
CDK — cyclin-dependent kinases,
$3CL^{pro}$ — 3C-like protease (M^{pro}),
CL — central limit theorem,
CoMFA — comparative molecular field analysis,
D3R GC — D3R grand challenge (a competition on blind prediction of protein – ligand complexes and ligand ranking in terms of the target affinity),
DFT — density-functional theory,
DMCA — 3,4-dimethoxycinnamic acid,
ED — ensemble docking,
Eg5 — kinesin-5,
FEP — free energy perturbation,
FXa — blood coagulation factor Xa,
ΔG — change in the Gibbs free energy,
GPCR— G protein-coupled receptors,
hFPPS — human farnesyl pyrophosphate synthase,
HIV — human immunodeficiency virus,
5HT1A — serotonin 1A receptor,
HTS — high-throughput screening,
IC_{50} — inhibitor concentration suppressing 50% of the enzyme activity,
IFD — induced fit docking,
IR — infrared spectroscopy,
IVS — inverse virtual screening,
K_d — dissociation constant,
KEAP1 — Kelch like ECH associated protein 1,
$KRAS^{G12C}$ — mutant of a protooncogene, a member of the Ras protein family,
Ligand — chemical compound involved in the intermolecular complex with the target,
MD — molecular dynamics,
MDM2 — E3 ubiquitin-protein ligase,
MEP — molecular electrostatic potential (3D), spatial distribution of the charge generated by the molecule,
MERS — Middle East respiratory syndrome,
MM/GBSA — molecular mechanics/generalized Born surface area,
Mnk1 — MAP kinase-activated protein kinase
MOF — metal organic frameworks,
MT — methyltransferase,
MurA — UDP-N-acetylglucosamine 1-carboxyvinyltransferase,
OPLS — optimized potentials for liquid simulations,
p53 — protein p53, a transcription factor regulating the cell cycle,
PDE — phosphodiesterase,
PKC — protein kinase C,
PLA2 — phospholipase A2,
PLS — partial least squares of latent variables,
Pose (in the ligand – target system) — the spatial position of the ligand relative to the target,
PrP — platelet-rich plasma,
QM — quantum mechanics,
QSAR — quantitative structure – activity relationship,
QSPR — quantitative structure – property relationship,
RMSD — root mean square deviation. A measure of correctness of superimposition of two structures,
ROC — receiver operating characteristic (dependence of the number of correctly classified positive objects on the number of incorrectly classified negative objects),
ROC AUC — area confined by the ROC curve and the axis of the false positive rate. The higher AUC, the better the classifier,
SAM — S-adenosyl methionine,
SARS-CoV-2 — severe acute respiratory syndrome coronavirus-2,
SBDD — structure-based drug design,
SGK1 — serine/threonine-protein kinase,
TGF-β — transforming growth factor beta,
Therapeutic target — protein participating in a particular pathophysiological process and being of interest for the targeted drug action.
TS — transition state,
VS — virtual screening; molecular docking of each compound from the library into a definite therapeutic target (protein) in order to calculate the structure and strength (energy) of the protein – compound complex.

References

1. M.H.J.Seifert, K.Wolf, D.Vitt. *BIOSILICO*, **1**, 143 (2003)
2. D.E.Clark. In *Annual Reports in Computational Chemistry*. Vol. 1. (Ed. D.Spellmeyer). (Elsevier, 2005). P. 133
3. A.Tropsha. In *Annual Reports in Computational Chemistry*. Vol. 2. (Ed. D.Spellmeyer). (Elsevier, 2006). P. 113
4. M.I.Skvortsova, I.V.Stankevich, V.A.Palyulin, N.S.Zefirov. *Russ. Chem. Rev.*, **75**, 961 (2006)
5. O.A.Raevsky. *Russ. Chem. Rev.*, **68**, 505 (1999)
6. D.E.Mager. *Adv. Drug Deliver. Rev.*, **58**, 1326 (2006)
7. Y.Tang, W.Zhu, K.Chen, H.Jiang. *Drug Discov. Today: Technol.*, **3**, 307 (2006)
8. S.Kar, K.Roy. *Expert Opin. Drug Dis.*, **8**, 245 (2013)
9. Open-access antimicrobial screening program, https://www.co-add.org/, Accessed 30.11.2020
10. T.Grützner, C.Schnider, D.Zollinger, B.C.Seyfang, N.Künzle. *Chem. Eng. Technol.*, **39**, 1835 (2016)
11. S.H.Unger. *J. Pharm. Sci.*, **76**, 269 (1987)
12. H.L.Hall, L.B.Kier, In *Topological Indices and Related Descriptors in QSAR and QSPR*. (Eds A.T.B.James Devillers). (Amsterdam: Gordon and Breach Science Publishers, 1999). P. 307
13. T.I.Madzhinov, I.I.Baskin, A.A.Varnek. *VVedenie v Khemoinformtiku. Khimicheskoe Prostranstvo i Virtual'nyi Skrining. (Introduction to Chemoinformatics. Chemical Space and Virtual Screening)*. Pt. 2. (Kazan: Kazan State University, 2019)

14. E.N.Muratov, J.Bajorath, R.P.Sheridan, I.V.Tetko, D.Filimonov, V.Poroikov, T.I.Oprea, I.I.Baskin, A.Varnek, A.Roitberg, O.Isayev, S.Curtalolo, D.Fourches, Y.Cohen, A.Aspuru-Guzik, D.A.Winkler, D.Agrafiotis, A.Cherkasov, A.Tropsha. *Chem. Soc. Rev.*, **49**, 3525 (2020)
15. O.A.Raevsky. *Modelirovanie Sootnoshenii 'Struktura – Svoistvo' (Structure – Property Relationship Modelling)*. (Moscow: KDU Dobrosvet, 2015)
16. E.I.Prokhorov, I.V.Svitan'ko, A.L.Zakharenko, M.V.Sukhanova, A.V.Bekker, A.V.Perevoznikov, M.I.Kumskov. *Pattern Recognit. Image Anal.*, **26**, 427 (2016)
17. P.C.S.Costa, J.S.Evangelista, I.Leal, P.C.M.L.Miranda. *Mathematics*, **9**, 60 (2020)
18. A.Tropsha. *Fundamentals of QSAR Modeling: Basic Concepts and Applications*; https://ntp.niehs.nih.gov/iccvam/meetings/commprac-2016/tropsha-508.pdf, Accessed 28.01.2021
19. D.S.Goodsell, A.J.Olson. *Proteins: Struct., Funct., Genet.*, **8**, 195 (1990)
20. I.D.Kuntz. *Science*, **257**, 1078 (1992)
21. N.S.Pagadala, K.Syed, J.Tuszynski. *Biophys. Rev.*, **9**, 91 (2017)
22. M.Medvedev, O.Stroganov, A.Dmitrienko, M.Panova, I.Svitanko, F.Novikov, G.Chilov. *J. Chem. Inf. Model.*, (2021) (in the press)
23. H.S.Muniz, A.S.Nascimento. *PLOS ONE*, **12**, e0174336 (2017)
24. V.Salmaso, S.Moro. *Front. Pharmacol.*, **9**, 923 (2018)
25. S.Kumar, B.Ma, C.-J.Tsai, N.Sinha, R.Nussinov. *Protein Sci.*, **9**, 10 (2008)
26. F.Fratev, S.Sirimulla. *Sci. Rep.*, **9** (2019)
27. N.Hansen, W.F.van Gunsteren. *J. Chem. Theory Comput.*, **10**, 2632 (2014)
28. D.Wu. *J. Chem. Phys.*, **133**, 244116 (2010)
29. A.A.Zeifman, V.V.Stroylov, F.N.Novikov, O.V.Stroganov, V.Kulkov, G.G.Chilov. *J. Chem. Theor. Comput.*, **9**, 1093 (2013)
30. Y-Q.Su, Y.Wang, J-X.Liu, I.A.W.Filot, K.Alexopoulos, L.Zhang, V.Muravev, B.Zijlstra, D.G.Vlachos, E.J.M.Hensen. *ACS Catalysis*, **9** (4), 3289 (2019)
31. V.M.Chernyshev, E.A.Denisova, D.B.Eremin, V.P.Ananikov. *Chem. Sci.*, **11**, 6957 (2020)
32. A.S.Kashin, V.P.Ananikov. *Nat. Rev. Chem.*, **3**, 624 (2019)
33. J.V.Burykina, N.S.Shlapakov, *E.G.*Gordeev, B.König, V.P.Ananikov. *Chem. Sci.*, **11**, 10061 (2020)
34. K.S.Egorova, A.S.Galushko, V.P.Ananikov. *Angew. Chem., Int. Ed.*, 59, 22296 (2020)
35. E.O.Pentsak, *E.G.*Gordeev, V.P.Ananikov. *Dokl. Phys. Chem.*, **493**, 95 (2020)
36. E.G.Gordeev, V.P.Ananikov. *Russ. Chem. Rev.*, **89** (2020)
37. J.Verma, V.Khedkar, E.Coutinho. *Curr. Top. Med. Chem.*, **10**, 95 (2010)
38. C.Hansch, P.P.Maloney, T.Fujita, R.M.Muir. *Nature*, **194**, 178 (1962)
39. A.Cherkasov, E.N.Muratov, D.Fourches, A.Varnek, I.I.Baskin, M.Cronin, J.Dearden, P.Gramatica, Y.C.Martin, R.Todeschini, V.Consonni, V.E.Kuz'min, R.Cramer, R.Benigni, C.Yang, J.Rathman, L.Terfloth, J.Gasteiger, A.Richard, A.Tropsha. *J. Med. Chem.*, **57**, 4977 (2014)
40. R.D.Cramer. *J. Comput. Aided Mol. Des.*, **26**, 35 (2012)
41. J.Polanski. *Curr. Med. Chem.*, **16**, 3243 (2009)
42. R.Todeschini, V.Consonni. *Molecular Descriptors for Chemoinformatics*. (Wiley, 2009)
43. R.D.Cramer, D.E.Patterson, J.D.Bunce. *JACS*, **110**, 5959 (1988)
44. S.Wold, C.Albano, W.J.Dunn, K.Esbensen, S.Hellberg, E.Johansson, W.Lindberg, M.Sjöström. *Analysis*, **12**, 477 (1984)
45. G.Klebe, U.Abraham, T.Mietzner. *J. Med. Chem.*, **37**, 4130 (1994)
46. P.J.Goodford. *J. Med. Chem.*, **28**, 849 (1985)
47. M.Pastor, G.Cruciani, I.McLay, S.Pickett, S.Clementi. *J. Med. Chem.*, **43**, 3233 (2000)
48. Á.Durán, I.Zamora, M.Pastor. *J. Chem. Inf. Model.*, **49**, 2129 (2009)
49. C.M.R.Low, J.G.Vinter. *J. Med. Chem.*, **51**, 565 (2008)
50. S.L.Dixon, A.M.Smondyrev, S.N.Rao. *Chem. Biol. Drug Des.*, **67**, 370 (2006)
51. A.R.Ortiz, M.T.Pisabarro, F.Gago, R.C.Wade. *J. Med. Chem.*, **38**, 2681 (1995)
52. H.Gohlke, G.Klebe. *J. Med. Chem.*, **45**, 4153 (2002)
53. A.M.Doweyko. *J. Med. Chem.*, **31**, 1396 (1988)
54. Z.Simon, I.Badilescu, T.Racovitan. *J. Theor. Biol.*, **66**, 485 (1977)
55. R.Varela, W.P.Walters, B.B.Goldman, A.N.Jain. *J. Med. Chem.*, **55**, 8926 (2012)
56. D.Mendez, A.Gaulton, A.P.Bento, J.Chambers, M.De Veij, E.Félix, M.P.Magariños, J.F.Mosquera, P.Mutowo, M.Nowotka, M.Gordillo-Marañón, F.Hunter, L.Junco, G.Mugumbate, M.Rodriguez-Lopez, F.Atkinson, N.Bosc, C.J.Radoux, A.Segura-Cabrera, A.Hersey, A.R.Leach. *Nucleic Acids Res.*, **47**, D930 (2019)
57. S.Kim, J.Chen, T.Cheng, A.Gindulyte, J.He, S.He, Q.Li, B.A.Shoemaker, P.A.Thiessen, B.Yu, L.Zaslavsky, J.Zhang, E.E.Bolton. *Nucleic Acids Res.*, **49**, D1388 (2021)
58. V.V.Poroikov, D.A.Filimonov, T.A.Gloriozova, A.A.Lagunin, D.S.Druzhilovskiy, A.V.Rudik, L.A.Stolbov, A.V.Dmitriev, O.A.Tarasova, S.M.Ivanov, P.V.Pogodin. *Russ. Chem. Bull.*, **68**, 2143 (2020)
59. GUSAR, http://www.pharmaexpert.ru, Accessed 30.11.2020
60. A.Lagunin, S.Ivanov, A.Rudik, D.Filimonov, V.Poroikov. *Bioinformatics*, **29**, 2062 (2013)
61. D.A.Filimonov, A.A.Lagunin, T.A.Gloriozova, A.V.Rudik, D.S.Druzhilovskii, P.V.Pogodin, V.V.Poroikov. *Chem. Heterocycl. Compd.*, **50**, 444 (2014)
62. http://www.way2drug.com, Accessed 25.01.2021
63. A.A.Nikitina, A.A.Orlov, L.I.Kozlovskaya, V.A.Palyulin, D.I.Osolodkin. *Database*, **2019** (2019)
64. A.K.Halder, M.N.D.S.Cordeiro. *Int. J. Mol. Sci.*, **20**, 4191 (2019)
65. E.J.Ariens, A.M.Simonis, J. M. van Rossum. In *Drug-receptor Interaction: Interaction of One or More Drugs with One Receptor System*. (Ed. E.J.Ariens). (Academic Press, 1964)
66. V.E.Golender, A.B.Rozenblit. *Logical and Combinatorial Algorithms for Drug Design*. (Research Studies Press, 1983)
67. R.L.DesJarlais, R.P.Sheridan, J.S.Dixon, I.D.Kuntz, R.Venkataraghavan. *J. Med. Chem.*, **29**, 2149 (1986)
68. N.Colloc'h, J.-P.Mornon. *J. Mol. Graphics*, **8**, 133 (1990)
69. C.Chothia, J.Janin. *Nature*, **256**, 705 (1975)
70. S.J.Wodak, J.Janin. *J. Mol. Biol.*, **124**, 323 (1978)
71. I.D.Kuntz, J.M.Blaney, S.J.Oatley, R.Langridge, T.E.Ferrin. *J. Mol. Biol.*, **161**, 269 (1982)

72. E.E.Hodgkin, W.G.Richards. *Int. J. Quantum Chem*, **32**, 105 (1987)
73. P.Sjoberg, J.S.Murray, T.Brinck, P.Evans, P.Politzer. *J. Mol. Graphics*, **8**, 81 (1990)
74. A.M.Richard. *J. Comput. Chem.*, **12**, 959 (1991)
75. P.Politzer, D.G.Truhlar. *Chemical Applications of Atomic and Molecular Electrostatic Potentials*. (Boston, MA: Springer, 1981)
76. E.Scrocco, J.Tomasi. *Top. Curr. Chem.*, **42**, 95 (1973)
77. M.A.B.dos Santos, L.F.S.de Oliveira, A.F.de Figueiredo, F.dos Santos Gil, M.de Souza Farias, H.R.Bitencourt, J.R.B.Lobato, R.D.de P.Farreira, S.S.da S.Alves, E.L.C.de Aquino, J.Ciríaco-Pinheiro. *Molecular Electrostatic Potential and Chemometric Techniques as Tools to Design Bioactive Compounds*. In *Cheminformatics and its Applications*. (Eds A.Stefaniu, A.Rasul, G.Hussain). (IntechOpen, 2020); https://doi.org/10.5772/intechopen.89113
78. R.Bonaccorsi, E.Scrocco, J.Tomasi. *Theor. Chim. Acta*, **21**, 17 (1971)
79. P.K.Weiner, P.A.Kollman. *J. Comput. Chem.*, **2**, 287 (1981)
80. E.I.Prokhorov, A.V.Bekker, A.V.Perevoznikov, I.V.Svitanko, A.L.Zakharenko, M.V.Sukhanova, M.I.Kumskov. In. *Prilozheniya Metoda Evolyutsionnogo Otbora Deskriptorov v Matematicheskom Modelirovanii Zavisimosti Biologicheskoi Aktivnosti Soedineniya ot ego Structury (Applications of the Method of Evolutionary Selection of Descriptors in Mathematical Modeling of the Dependence of the Biological Activity of a Compound on its Structure)*. (Moscow: MAKS Press, 2012). P. 3
81. B.B.Masek, A.Merchant, J.B.Matthew. *J. Med. Chem.*, **36**, 1230 (1993)
82. I.V.Svitan'ko, M.I.Kumskov, I.L.Zyryanov, I.A.Suslov. *Mendeleev Commun.*, **4**, 161 (1994)
83. M.I.Kumskov, I.L.Zyryanov, I.V.Svitan'ko. *Pattern Recognit. Image Anal.*, **3**, 477 (1995)
84. T.Ghafourian, J.C.Dearden. *Il Farmaco*, **59**, 473 (2004)
85. T.Ghafourian, J.C.Dearden. *J. Pharm. Pharmacol.*, **52**, 603 (2000)
86. F.W.Biegler-König, R.F.W.Bader, T.-H.Tang. *J. Comput. Chem.*, **3**, 317 (1982)
87. J.L.Derissen, P.H.Smit. *Acta Crystallogr. Sect. A: Found. Crystallogr.*, **34**, 842 (1978)
88. U.C.Singh, P.A.Kollman. *J. Comput. Chem.*, **5**, 129 (1984)
89. S.K.Kearsley, G.M.Smith. *Tetrahedron Comp. Method.*, **3**, 615 (1990)
90. D.A.Shulga, O.I.Titov, S.A.Pisarev, V.A.Palyulin. *SAR QSAR Environ. Res.*, **29**, 21 (2018)
91. O.I.Titov, D.A.Shulga, V.A.Palyulin. *J. Chem. Theory Comput.*, **15**, 1159 (2019)
92. O.I.Titov, A.R.Shaimardanov, V.A.Palyulin. *Mendeleev Commun.*, **30**, 741 (2020)
93. M.Genva, T.Kenne Kemene, M.Deleu, L.Lins, M.-L.Fauconnier. *Int. J. Mol. Sci.*, **20**, 3018 (2019).
94. F.Luan, H.T.Liu, Y.Y.Wen, X.Y.Zhang. *Flavour Frag. J.*, **23**, 232 (2008)
95. C.Rojas, P.R.Duchowicz, P.Tripaldi, R.P.Diez. *Chemometr. Intell. Lab.*, **140**, 126 (2015)
96. F.Vella. *Biochem. Ed.*, **21**, 165 (1993)
97. J.Solms, L.Vuataz, R.H.Egli. *Experientia*, **21**, 692 (1965)
98. S.S.Schiffman, T.B.Clark, J.Gagnon. *Physiol. Behav.*, **28**, 457 (1982)
99. W.H.De Camp. *Chirality*, **1**, 2 (1989)
100. A.J.Hutt, S.C.Tan. *Drugs*, **52**, 1 (1996)
101. S.Wnendt, K.Zwingenberger. *Nature*, **385**, 303 (1997)
102. J.A.Bajgrowicz, I.Frank. *Tetrahedron: Asymmetry*, **12**, 2049 (2001)
103. G.Buchbauer, P.Lebada, L.Wiesinger, P.Weiss-Greiler, P.Wolschann. *Chirality*, **9**, 380 (1997)
104. T.A.Cleland, A.Borthakur. *Front. Comput. Neurosc.*, **14**, 579143 (2020)
105. C.S.Sell. *Angew. Chem., Int. Ed.*, **45**, 6254 (2006)
106. T.Salthammer, F.Monegel, N.Schulz, E.Uhde, S.Grimme, J.Seibert, U.Hohm, W.U.Palm. *Chem.–Eur. J.*, **27**, 1046 (2021)
107. P.Mombaerts, F.Wang, C.Dulac, S.K.Chao, A.Nemes, M.Mendelsohn, J.Edmondson, R.Axel. *Cell*, **87**, 675 (1996)
108. Z.Zou. *Science*, **311**, 1477 (2006)
109. A.Kovatcheva, A.Golbraikh, S.Oloff, Y.-D.Xiao, W.Zheng, P.Wolschann, G.Buchbauer, A.Tropsha. *J. Chem. Inf. Comp. Sci.*, **44**, 582 (2004)
110. I.Bersuker, A.Dimoglo, M.Y.Gorbachov, M.Koltsa, P.Vlad. *Nouv. J. Chim.*, **9**, 211 (1985)
111. P.Politzer, J.S.Murray. In *Reviews in Computational Chemistry*. (Eds K.B.Lipkowitz, D.B.Boyd). (Hoboken, NJ, Wiley, 2007). P. 273
112. I.V.Svitanko, D.A.Devetyarov, D.E.Tcheboukov, M.S.Dolmat, A.M.Zakharov, S.S.Grigor'eva, V.T.Chichua, L.A.Ponomareva, M.I.Kumskov. *Mendeleev Commun.*, **17**, 90 (2007)
113. L.A.Kheifits, V.M.Dashunin. *Dushistye Veshchestva i Drugie Produkty dlya Parfyumerii (Fragrance Compounds and Other Products for Perfumery)*. (Moscow: Chemistry, 1994)
114. I.V.Svitanko. Doctoral Thesis in Chemical Sciences, N.D.Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Moscow, 2018
115. A.V.Val'dman, A.Kostandov, A.V.Martynikhin. *Bull. Exp. Biol. Med.*, **95**, 62 (1983)
116. N.N.Makhova, L.I.Belen'kii, G.A.Gazieva, I.L.Dalinger, L.S.Konstantinova, V.V.Kuznetsov, A.N.Kravchenko, M.M.Krayushkin, O.A.Rakitin, A.M.Starosotnikov, L.L.Fershtat, S.A.Shevelev, V.Z.Shirinian, V.N.Yarovenko. *Russ. Chem. Rev.*, **89**, 55 (2020)
117. V.V.Baranov, Y.A.Barsegyan, Y.A.Strelenko, V.A.Karnoukhova, A.N.Kravchenko. *Mendeleev Commun.*, **30**, 479 (2020)
118. V.V.Baranov, A.B.Ivanova, N.G.Kolotyrkina, A.N.Kravchenko. *Mendeleev Commun.*, **29**, 94 (2019)
119. V.A.Eres'ko, L.V.Epishina, O.V.Lebedev, M.V. Povstyanoi, L.I.Khmel'nitskii, S.S.Novikov. *Russ. Chem. Bull.*, **29**, 1138 (1980)
120. L.V.Anikina, G.A.Gazieva, A.N.Kravchenko. *Russ. Chem. Bull.*, **69**, 563 (2020)
121. I.V.Svitan'ko, I.L.Zyryanov, M.I.Kumskov, L.I.Khmel'nitskii, L.I.Suvorova, A.N.Kravchenko, T.B.Markova, O.V.Lebedev, G.A.Orekhova, S.V.Belova. *Mendeleev Commun.*, **5**, 49 (1995)
122. G.A.Gazieva, Y.B.Vikharev, L.V.Anikina, T.B.Karpova, A.N.Kravchenko, E.A.Permyakov, I.V.Svitanko. *Mendeleev Commun.*, **23**, 202 (2013)
123. *3D QSAR in Drug Design. Theory Methods and Applications*. Vol. 1. (Ed. Hugo Kubinyi). (Netherlands: Springer, 1994)
124. M.Damale, S.Harke, F.Kalam Khan, D.Shinde, J.Sangshetti. *Mini-Rev. Med. Chem.*, **14**, 35 (2014)
125. N.Vahedi, M.Mohammadhosseini, M.Nekoei. *Curr. Anal. Chem.*, **16**, 1088 (2020)

126. A.Tropsha. In *Burger's Medicinal Chemistry and Drug Discovery*. (Ed. D.J.Abraham). (Hoboken, NJ: Wiley, 2003); https://doi.org/10.1002/0471266949.bmc002.pub2
127. S.K.Burley, C.Bhikadiya, C.Bi, S.Bittrich, L.Chen, G.V.Crichlow, C.H. Christie, K.Dalenberg, L.Costanzo, J.M.Duarte, S.Dutta, Z.Feng, S.Ganesan, D.S.Goodsell, S.Ghosh, R.Green, V.Guranović, D.Guzenko, B.Hudson, C.Lawson, Y.Liang, R.Lowe, H.Namkoong, E.Peisach, I.Persikova, C.Randle, A.Rose, Y.Rose, A.Sali, J.Segura, M.Sekharan, C.Shao, Y.-P.Tao, M.Voigt, J.Westbrook, J.Young, C.Zardecki, M.Zhuravleva. *Nucleic Acids Res*., **49**, D437 (2021)
128. O.V.Stroganov, F.N.Novikov, M.G.Medvedev, A.O.Dmitrienko, I.Gerasimov, I.V.Svitanko, G.G.Chilov. *J. Comput. Aided Mol. Des*., **34**, 121 (2020)
129. J.J.Irwin, B.K.Shoichet. *J. Med. Chem*., **59**, 4103 (2016)
130. A.Sethi, K.Joshi, K.Sasikala, M.Alvala. In *Drug Discovery and Development — New Advances*. (Eds V.Gaitonde, P.Karmakar, A.Trivedi). (IntechOpen, India, 2020)
131. X.C.Yan, J.M.Sanders, Y.-D.Gao, M.Tudor, A.M.Haidle, D.J.Klein, A.Converso, C.A.Lesburg, Y.Zang, H.B.Wood. *J. Chem. Inf. Model*., **60**, 4144 (2020)
132. N.van Hilten, F.Chevillard, P.Kolb. *J. Chem. Inf. Model*., **59**, 644 (2019)
133. C.Gorgulla, A.Boeszoermenyi, Z.-F.Wang, P.D.Fischer, P.W.Coote, K.M.Padmanabha Das, Y.S.Malets, D.S.Radchenko, Y.S.Moroz, D.A.Scott, K.Fackeldey, M.Hoffmann, I.Iavniuk, G.Wagner, H.Arthanari. *Nature*, **580**, 663 (2020)
134. J.Lyu, S.Wang, T.E.Balius, I.Singh, A.Levit, Y.S.Moroz, M.J.O'Meara, T.Che, E.Algaa, K.Tolmachova, A.A.Tolmachev, B.K.Shoichet, B.L.Roth, J.J.Irwin. *Nature (Lodnon)*, **566**, 224 (2019)
135. I.Casciuc, D.Horvath, A.Gryniukova, K.A.Tolmachova, O.V.Vasylchenko, P.Borysko, Y.S.Moroz, J.Bajorath, A.Varnek. *Eur. J. Med. Chem*., **165**, 258 (2019)
136. X.Xu, M.Huang, X.Zou. *Biophys. Rep*., **4**, 1 (2018)
137. R.Acharya, S.Chacko, P.Bose, A.Lapenna, S.P.Pattanayak. *Sci. Rep*., **9** (2019)
138. M.Mantipally, M.R.Gangireddy, R.Gundla, V.N.Badavath, S.R.Mandha, V.C.Maddipati. *Biorg. Med. Chem. Lett*., **29**, 2248 (2019)
139. C.B.M.Platania, G.Giurdanella, L.Di Paola, G.M.Leggio, F.Drago, S.Salomone, C.Bucolo. *Biochem. Pharmacol*., **138**, 130 (2017)
140. W.Zhao, M.Xiong, X.Yuan, M.Li, H.Sun, Y.Xu. *J. Chem. Inf. Model*., **60**, 3265 (2020)
141. P.Bhatia, V.Sharma, O.Alam, A.Manaithiya, P.Alam, Kahksha, M.T.Alam, M.Imran. *Eur. J. Med. Chem*., **204**, 112640 (2020).
142. D.Bartuzi, A.Kaczor, K.Targowska-Duda, D.Matosiuk. *Molecules*, **22**, 340 (2017)
143. H.Liu, M.Wang, X.Zhang, J.Ma, G.Lu. *Appl. Catal., B*, **237**, 563 (2018)
144. I.Doytchinova, M.Atanasova, I.Valkova, G.Stavrakov, I.Philipova, Z.Zhivkova, D.Zheleva-Dimitrova, S.Konstantinov, I.Dimitrov. *J. Enzym. Inhib. Med. Chem*., **33**, 768 (2018)
145. F.Türkan, M.H.Calimli, G.S.Kanberoglu, M.Karaman. *J. Biomol. Struct. Dyn*., 1 (2020)
146. F.Ballante, A.Rudling, A.Zeifman, A.Luttens, D.D.Vo, J.J.Irwin, J.Kihlberg, J.Brea, M.I.Loza, J.Carlsson. *J. Med. Chem*., **63**, 613 (2020)
147. D.R.Weiss, J.Karpiak, X.-P.Huang, M.F.Sassano, J.Lyu, B.L.Roth, B.K.Shoichet. *J. Med. Chem*., **61**, 6830 (2018)
148. Y.Zou, Y.Hu, S.Ge, Y.Zheng, Y.Li, W.Liu, W.Guo, Y.Zhang, Q.Xu, Y.Lai. *Eur. J. Med. Chem*., **184**, 111750 (2019)
149. X.Hu, Y.-Q.Zhang, O.W.Lee, L.Liu, M.Tang, K.Lai, M.B.Boxer, M.D.Hall, M.Shen. *J. Comput. Aided Mol. Des*., **33**, 405 (2019)
150. W.Zhou, M.Duan, W.Fu, J.Pang, Q.Tang, H.Sun, L.Xu, S.Chang, D.Li, T.Hou. *Genomics, Proteom. Bioinform*., **16**, 416 (2018)
151. R.M.Stein, H.J.Kang, J.D.McCorvy, G.C.Glatfelter, A.J.Jones, T.Che, S.Slocum, Xi-Ping Huang, O.Savych, Y.S.Moroz, B.Stauch, L.C.Johansson, V.Cherezov, T.Kenakin, J.J.Irwin, B.K.Shoichet, B.L.Roth, M.L.Dubocovich. *Nature*, **579**, 609-614 (2020); https://www.nature.com/articles/s41586-020-2027-0
152. A.A.Mian, A.Rafiei, I.Haberbosch, A.Zeifman, I.Titov, V.Stroylov, A.Metodieva, O.Stroganov, F.Novikov, B.Brill, G.Chilov, D.Hoelzer, O.G.Ottmann, M.Ruthardt. *Leukemia*, **29**, 1104 (2015)
153. E.Ivanova, V.Tatarskiy, M.Yastrebova, A.Khamidullina, A.Shunaev, A.Kalinina, A.Zeifman, F.Novikov, Y.Dutikova, G.Chilov, A.Shtil. *Int. J. Oncol*., **55**, 289 (2019)
154. K.Lee, L.Goodman, C.Fourie, S.Schenk, B.Leitch, J.M.Montgomery. In *Advances in Protein Chemistry and Structural Biology*. Vol. 103. (Ed. R.Donev). (Elsevier, 2016). P. 203
155. E.V.Radchenko, D.S.Karlov, M.I.Lavrov, V.A.Palyulin. *Mendeleev Commun*., **27**, 623 (2017)
156. M.I.Lavrov, D.S.Karlov, V.A.Palyulin, V.V.Grigoriev, V.L.Zamoyski, G.E.Brkich, N.V.Pyatigorskaya, M.E.Zapolskiy. *Mendeleev Commun*., **28**, 311 (2018)
157. A.A.Nazarova, K.N.Sedenkova, D.S.Karlov, M.I.Lavrov, Y.K.Grishin, T.S.Kuznetsova, V.L.Zamoyski, V.V.Grigoriev, E.B.Averina, V.A.Palyulin. *MedChemComm*, **10**, 1615 (2019)
158. M.I.Lavrov, D.S.Karlov, T.A.Voronina, V.V.Grigoriev, A.A.Ustyugov, S.O.Bachurin, V.A.Palyulin. *Mol. Neurobiol*., **57**, 191 (2020)
159. M.I.Lavrov, O.V.Stroganov, V.L.Zamoyski, V.V.Grigoriev, M.E.Zapolskiy, S.V.Sysolyatin, V.V.Malykhin, V.N.Surmachev, V.A.Palyulin. *Mendeleev Commun*., **30**, 156 (2020)
160. M.I.Lavrov, P.N.Veremeeva, D.S.Karlov, V.L.Zamoyski, V.V.Grigoriev, V.A.Palyulin. *Mendeleev Commun*., **29**, 619 (2019)
161. S.Beura, P.Chetti. *J. Mol. Struct*., **1216**, 128271 (2020)
162. L.Zhang, F.Lai, X.Chen, Z.Xiao. *J. Mol. Graphics Modell*., **99**, 107628 (2020)
163. A.Ghaleb, A.Aouidate, H.B.E.Ayouchia, M.Aarjane, H.Anane, S.-E.Stiriba. *J. Biomol. Struct. Dyn*., 1 (2020)
164. S.Daoud, M.O.Taha. *J. Mol. Graphics Modell*., **99**, 107615 (2020)
165. Y.Rasouli, A.Davood. *Curr. Comput.–Aided Drug Des*., **14**, 35 (2018)
166. E.I.Prokhorov, A.V.Bekker, A.V.Perevoznikov, M.I.Kumskov, I.V.Svitanko. *Mendeleev Commun*., **25**, 214 (2015)
167. S.Khamouli, S.Belaidi, M.Ouassaf, T.Lanez, S.Belaaouad, S.Chtita. *J. Biomol. Struct. Dyn*., 1 (2020)
168. P.Shirvani, A.Fassihi. *Mol. Simulat*., **46**, 1265 (2020)
169. V.Yele, M.A.Azam, S.Jupudi. *Chem. Pap*., **74**, 4567 (2020)

170. R.Chayrov, N.A.Parisis, M.V.Chatziathanasiadou, E.Vrontaki, K.Moschovou, G.Melagraki, H.Sbirkova-Dimitrova, B.Shivachev, M.Schmidtke, Y.Mitrev, M.Sticha, T.Mavromoustakos, A.G.Tzakos, I.Stankova. *Molecules*, **25**, 3989 (2020)
171. W.Tan, H.Mei, L.Chao, T.Liu, X.Pan, M.Shu, L.Yang. *J. Comput. Aided Mol. Des.*, **27**, 1067 (2013)
172. P.Durai, Y.-J.Ko, C.-H.Pan, K.Park. *BMC Bioinformatics*, **21**, 309 (2020)
173. M.P.Baumgartner, D.A.Evans. *J. Comput. Aided Mol. Des.*, **32**, 45 (2018)
174. R.Friesner, R.Abel, D.M.Troast, S.B.Rafi, S.Mondal, P.Ghanakota, T.Day, N.Boyles, S.Jerome, S.Dixon, F.Ranalli, M.Grisewood, K.Borrelli, D.Sindhikara, R.Murphy, E.Miller. *ChemRxiv*, (2020); https://doi.org/10.26434/chemrxiv.11983845.v1
175. A.Bolia, S.B.Ozkan. *J. Chem. Inf. Model.*, **56**, 734 (2016)
176. Z.Ye, M.P.Baumgartner, B.M.Wingert, C.J.Camacho. *J. Comput. Aided Mol. Des.*, **30**, 695 (2016)
177. D.S.Paul, N.Gautham. *J. Mol. Graphics Modell.*, **74**, 89 (2017)
178. C.Grebner, J.Iegre, J.Ulander, K.Edman, A.Hogner, C.Tyrchan. *J. Chem. Inf. Model.*, **56**, 774 (2016)
179. W.Sherman, T.Day, M.P.Jacobson, R.A.Friesner, R.Farid. *J. Med. Chem.*, **49**, 534 (2006)
180. Y.Niu, X.Yao, H.Ji. *RSC Adv.*, **9**, 12441 (2019)
181. X.Tang, Z.Wang, T.Lei, W.Zhou, S.Chang, D.Li. *PhysChemChemPhys*, **20**, 5591 (2018)
182. K.Rožman, S.Lešnik, B.Brus, M.Hrast, M.Sova, D.Patin, H.Barreteau, J.Konc, D.Janežič, S.Gobec. *Biorg. Med. Chem. Lett.*, **27**, 944 (2017)
183. R.K.Mishra, M.R.Clutter, G.T.Blyth, E.M.Kosciuczuk, A.Z.Blackburn, E.M.Beauchamp, G.E.Schiltz, L.C.Platanias. *Chem. Biol. Drug Des.*, **94**, 1813 (2019)
184. S.Motta, L.Bonati. *J. Chem. Inf. Model.*, **57**, 1563 (2017)
185. R.E.Amaro, J.Baudry, J.Chodera, Ö.Demir, J.A.McCammon, Y.Miao, J.C.Smith. *Biophys. J.*, **114**, 2271 (2018)
186. W.Evangelista, R.L.Weir, S.R.Ellingson, J.B.Harris, K.Kapoor, J.C.Smith, J.Baudry. *Biorg. Med. Chem.*, **24**, 4928 (2016)
187. A.Bhattarai, J.Wang, Y.Miao. *BBA-Gen Subjects*, **1864**, 129615 (2020)
188. X.Li, X.X.Zhang, Y.X.Lin, X.M.Xu, L.Li, J.B.Yang. *Chem. Biodivers.*, **16**, 7, e1900170 (2019)
189. V.K.Sharma, P.P.Nandekar, A.Sangamwar, H.Pérez-Sánchez, S.M.Agarwal. *RSC Adv.*, **6**, 65725 (2016)
190. H.A.Velazquez, D.Riccardi, Z.Xiao, L.D.Quarles, C.R.Yates, J.Baudry, J.C.Smith. *Chem. Biol. Drug Des.*, **91**, 491 (2018)
191. K.Kapoor, N.McGill, C.B.Peterson, H.V.Meyers, M.N.Blackburn, J.Baudry. *J. Chem. Inf. Model.*, **56**, 535 (2016)
192. S.R.Ellingson, J.C.Smith, J.Baudry. *Mol. Simulat.*, **40**, 848 (2014)
193. O.Slater, M.Kontoyianni. *Expert Opin. Drug Dis.*, **14**, 619 (2019)
194. I.G.Tikhonova, I.I.Baskin, V.A.Palyulin, N.S.Zefirov. *Dokl. Biochem. Biophys.*, **382**, 67 (2002)
195. 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)
196. N.K.Broomhead, M.E.Soliman. *Cell. Biochem. Biophys.*, **75**, 15 (2017)
197. K.Chen, M.J.Mizianty, J.Gao, L.Kurgan. *Structure*, **19**, 613 (2011)
198. T.Ban, M.Ohue, Y.Akiyama. *Comput. Biol. Chem.*, **73**, 139 (2018)
199. I.Zanyatkin, Y.Stroylova, S.Tishina, V.Stroylov, A.Melnikova, T.Haertle, V.Muronetz. *Phytother. Res.*, **31**, 1046 (2017)
200. V.S.Stroylov, I.V.Svitanko. *Mendeleev Commun.*, **30**, 419 (2020)
201. B.-L.Lin, S.Zhao, Y.Wu, J.Tong, L.Xu. *ChemRxiv* (2020)
202. M.-H.Lin, D.C.Moses, C.-H.Hsieh, S.-C.Cheng, Y.-H.Chen, C.-Y.Sun, C.-Y.Chou. *Antivir. Res.*, **150**, 155 (2018)
203. R.B.Murphy, M.P.Repasky, J.R.Greenwood, I.Tubert-Brohman, S.Jerome, R.Annabhimoju, N.A.Boyles, C.D.Schmitz, R.Abel, R.Farid, R.A.Friesner. *J. Med. Chem.*, **59**, 4364 (2016)
204. A.Ganesan, M.L.Coote, K.Barakat. *Drug Discov. Today*, **22**, 249 (2017)
205. H.Guterres, W.Im. *J. Chem. Inf. Model.*, **60**, 2189 (2020)
206. H.Zhao, A.Caflisch. *Eur. J. Med. Chem.*, **91**, 4 (2015)
207. X.Liu, D.Shi, S.Zhou, H.Liu, H.Liu, X.Yao. *Expert Opin. Drug Discov.*, **13**, 23 (2017)
208. A.A.T.Naqvi, T.Mohammad, G.M.Hasan, M.I.Hassan. *Curr. Top. Med. Chem.*, **18**, 1755 (2018)
209. S.A.Hollingsworth, R.O.Dror. *Neuron*, **99**, 1129 (2018)
210. P.Sneha, C.George, P.Doss. *Adv.Protein Chem. Struct. Biol.*, **102**, 181 (2016)
211. Y.Lee, R.Lazim, S.J.Y.Macalino, S.Choi. *Curr. Opin. Struct. Biol.*, **55**, 147 (2019)
212. S.Aci-Sèche, S.Ziada, A.Braka, R.Arora, P.Bonnet. *Future Med. Chem.*, **8**, 545 (2016)
213. G.Zhang, Y.Ren. *Molecules*, **23**, 2924 (2018)
214. T.Mohammad, S.Siddiqui, A.Shamsi, M.F.Alajmi, A.Hussain, A.Islam, F.Ahmad, M.I.Hassan. *Molecules*, **25**, 823 (2020)
215. C.Godoy-Castillo, N.Bravo-Acuña, G.Arriagada, F.Faunes, R.León, J.Soto-Delgado. *J. Biomol. Struct. Dynamics*, 1 (2020)
216. X.Kong, H.Sun, P.Pan, F.Zhu, S.Chang, L.Xu, Y.Li, T.Hou. *PhusChemChemPhys*, **20**, 4851 (2018)
217. R.W.Zwanzig. *J. Chem. Phys.*, **22**, 1420 (1954)
218. P.Wirnsberger, A.J.Ballard, G.Papamakarios, S.Abercrombie, S.Racanière, A.Pritzel, D.Jimenez Rezende, C.Blundell. *J. Chem. Phys.*, **153**, 144112 (2020)
219. T.Senapathi, M.Suruzhon, C.B.Barnett, J.Essex, K.J.Naidoo. *J. Chem. Inf. Model.*, **60**, 5290 (2020)
220. H.Chen, J.D.C.Maia, B.K.Radak, D.J.Hardy, W.Cai, C.Chipot, E.Tajkhorshid. *J. Chem. Inf. Model.*, **60**, 5301 (2020)
221. W.Jespers, M.Esguerra, J.Åqvist, H.Gutiérrez-de-Terán. *J. Cheminform.*, **11** 26 (2019)
222. P.Bauer, A.Barrozo, M.Purg, B.A.Amrein, M.Esguerra, P.B.Wilson, D.T.Major, J.Åqvist, S.C.L.Kamerlin. *SoftwareX*, **7**, 388 (2018)
223. X.He, S.Liu, T.-S.Lee, B.Ji, V.H.Man, D.M.York, J.Wang. *ACS Omega*, **5**, 4611 (2020)
224. A.Y.Shih, M.Hack, T.Mirzadegan. *J. Chem. Inf. Model.*, **60**, 5287 (2020)
225. K.Wallraven, F.L.Holmelin, A.Glas, S.Hennig, A.I.Frolov, T.N.Grossmann. *Chem. Sci.*, **11**, 2269 (2020)

226. C.E.M.Schindler, H.Baumann, A.Blum, D.Böse, H.-P.Buchstaller, L.Burgdorf, D.Cappel, E.Chekler, P.Czodrowski, D.Dorsch, M.K.I.Eguida, B.Follows, T.Fuchß, U.Grädler, J.Gunera, T.Johnson, C.Jorand Lebrun, S.Karra, M.Klein, T.Knehans, L.Koetzner, M.Krier, M.Leiendecker, B.Leuthner, L.Li, I.Mochalkin, D.Musil, C.Neagu, F.Rippmann, K.Schiemann, R.Schulz, T.Steinbrecher, E.-M.Tanzer, A.Unzue Lopez, A.Viacava Follis, A.Wegener, D.Kuhn. *J. Chem. Inf. Model.*, **60**, 5457 (2020)
227. K.Roos, C.Wu, W.Damm, M.Reboul, J.M.Stevenson, C.Lu, M.K.Dahlgren, S.Mondal, W.Chen, L.Wang, R.Abel, R.A.Friesner, E.D.Harder. *J. Chem. Theory Comput.*, **15**, 1863 (2019)
228. M.Ciordia, L.Pérez-Benito, F.Delgado, A.A.Trabanco, G.Tresadern. *J. Chem. Inf. Model.*, **56**, 1856 (2016)
229. Z.Cournia, B.Allen, W.Sherman. *J. Chem. Inf. Model.*, **57**, 2911 (2017)
230. Z.Li, X.Li, Y.-Y.Huang, Y.Wu, R.Liu, L.Zhou, Y.Lin, D.Wu, L.Zhang, H.Liu, X.Xu, K.Yu, Y.Zhang, J.Cui, C.-G.Zhan, X.Wang, H.-B.Luo. *Proc. Natl Acad. Sci. USA*, **117**, 27381 (2020)
231. Z.Li, Y.Huang, Y.Wu, J.Chen, D.Wu, C.-G.Zhan, H.-B.Luo. *J. Med. Chem.*, **62**, 2099 (2019)
232. N.Singh, W.Li. *Int. J. Mol. Sci.*, **21**, 4765 (2020)
233. J.Mortier, A.Friberg, V.Badock, D.Moosmayer, J.Schroeder, P.Steigemann, F.Siegel, S.Gradl, M.Bauser, R.C.Hillig, H.Briem, K.Eis, B.Bader, D.Nguyen, C.D.Christ. *ChemMedChem*, **15**, 827 (2020)
234. S.Chinnasamy, G.Selvaraj, C.Selvaraj, A.C.Kaushik, S.Kaliamurthi, A.Khan, S.K.Singh, D.-Q.Wei. *Int. J. Biol. Macromol.*, **144**, 53 (2020)
235. R.P.Fisher. *Transcription*, **10**, 47 (2019)
236. C.Sánchez-Martínez, M.J.Lallena, S.G.Sanfeliciano, A.de Dios. *Biorg. Med. Chem. Lett.*, **29**, 126637 (2019)
237. P.V.Rusina, I.Y.Titov, M.V.Panova, V.S.Stroylov, Y.R.Abdyusheva, E.Y.Murlatova, I.V.Svitanko, F.N.Novikov. *Mendeleev Commun.*, **30**, 430 (2020)
238. J.Seligmann, C.Twelves. *Future Med. Chem.*, **5**, 339 (2013)
239. D.V.Tsyganov, M.N.Semenova, L.D.Konyushkin, V.I.Ushkarov, M.M.Raihstat, V.V.Semenov. *Mendeleev Commun.*, **29**, 163 (2019)
240. A.Pandit, K.Yadav, R.B.Reddy, S.Sengupta, R.Sharma, V.Chelvam. *J. Mol. Struct.*, **1223**, 129204 (2021)
241. S.Mirzaei, F.Eisvand, F.Hadizadeh, F.Mosafa, R.Ghodsi. *Iran. J. Basic. Med. Sci.*, **23** (2020)
242. E.T.Warda, I.A.Shehata, M.B.El-Ashmawy, N.S.El-Gohary. *Biorg. Med. Chem.*, **28**, 115674 (2020)
243. V.S.Stroylov, I.V.Svitanko, A.S.Maksimenko, V.P.Kislyi, M.N.Semenova, V.V.Semenov. *Biorg. Med. Chem. Lett.*, **30**, 127608 (2020)
244. Y.B.Ruiz-Blanco, E.Sanchez-Garcia. *J. Chem. Theory Comput.*, **16**, 1396 (2020)
245. A.D.Wade, A.Rizzi, Y.Wang, D.J.Huggins. *J. Chem. Inf. Model.*, **59**, 2776 (2019)
246. E.A.Rifai, V.Ferrario, J.Pleiss, D.P.Geerke. *J. Chem. Theory Comput.*, **16**, 1300 (2020)
247. A.D.Wade, D.J.Huggins. *J. Chem. Theory Comput.*, **15**, 6504 (2019)
248. N.D.Austin, N.V.Sahinidis, D.W.Trahan. *Chem. Eng. Res. Des.*, **116**, 2 (2016)
249. A.I.Papadopoulos, I.Tsivintzelis, P.Linke, P.Seferlis. *Computer-Aided Molecular Design: Fundamentals, Methods, and Applications*. In Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. (Ed. J.Reedijk). (Waltham, MA: Elsevier, 2018); https://doi.org/10.1016/B978-0-12-409547-2.14342-2
250. D.Mucs, R.A.Bryce. *Expert Opin. Drug Dis.*, **8**, 263 (2013)
251. C.Gertig, K.Leonhard, A.Bardow. *Curr. Opin. Chem. Eng.*, **27**, 89 (2020)
252. M.G.Khrenova, A.V.Nemukhin, V.G.Tsirelson. *Mendeleev Commun.*, **30**, 583 (2020)
253. M.G.Khrenova, V.G.Tsirelson. *Mendeleev Commun.*, **29**, 492 (2019)
254. N.M.Ankudinov, D.S.Perekalin. *Mendeleev Commun.*, **29**, 263 (2019)
255. A.A.Zeifman, F.N.Novikov, V.S.Stroylov, O.V.Stroganov, I.V.Svitanko, G.G.Chilov. *Dalton Trans.*, **44**, 17795 (2015)
256. N.Miyaura, A.Suzuki. *Chem. Rev.*, **95**, 2457 (1995)
257. S.Kozuch, J.M.L.Martin. *ACS Catalysis*, **1**, 246 (2011)
258. S.-L.Zhang, L.Huang, L.-J.Sun. *Dalton Trans.*, **44**, 4613 (2015)
259. F.Proutiere, F.Schoenebeck. *Angew. Chem., Int. Ed.*, **50**, 8192 (2011)
260. E.O.Pentsak, V.P.Ananikov. *Eur. J. Org. Chem.*, **2019**, 4239 (2019)
261. Y.Amar, A.M.Schweidtmann, P.Deutsch, L.Cao, A.Lapkin. *Chem. Sci.*, **10**, 6697 (2019)
262. J.J.Henle, A.F.Zahrt, B.T.Rose, W.T.Darrow, Y.Wang, S.E.Denmark. *JACS*, **142**, 11578 (2020)
263. A.F.Zahrt, S.V.Athavale, S.E.Denmark. *Chem. Rev.*, **120**, 1620 (2020)
264. D.A.Shulga, A.R.Shaimardanov, V.A.Palyulin. *Mendeleev Commun.*, **30**, 741 (2020)
265. H.-U.Blaser. *Rendiconti Lincei*, **18**, 281 (2007)
266. J.M.Hawkins, T.J.N.Watson. *Angew. Chem., Int. Ed.*, **43**, 3224 (2004)
267. J.P.Reid, R.S.J.Proctor, M.S.Sigman, R.J.Phipps. *JACS*, **141**, 19178 (2019)
268. L.Zhu, H.Mohamed, H.Yuan, J.Zhang. *Catal. Sci. Tech.*, **9**, 6482 (2019)
269. A.F.Zahrt, J.J.Henle, B.T.Rose, Y.Wang, W.T.Darrow, S.E.Denmark. *Science*, **363**, eaau5631 (2019)
270. F.Teng, T.Yu, Y.Peng, W.Hu, H.Hu, Y.He, S.Luo, Q.Zhu. *JACS*, **143**, 7, 2722 (2021)
271. A.V.Kolos, D.S.Perekalin. *Mendeleev Commun.*, **31**, 1 (2021)
272. T.Piou, T.Rovis. *Acc. Chem. Res.*, **51**, 170 (2018)
273. P.S.Schulz, N.Müller, A.Bösmann, P.Wasserscheid. *Angew. Chem., Int. Ed.*, **46**, 1293 (2007)
274. A.A.Zeifman, V.S.Stroylov, I.Y.Titov, F.N.Novikov, O.V.Stroganov, I.V.Svitanko, G.G.Chilov. *Mendeleev Commun.*, **25**, 269 (2015)
275. J.Wåhlander, M.Amedjkouh, D.Balcells. *Eur. J. Org. Chem.*, **2019**, 442 (2019)
276. N.Tokunaga, Y.Otomaru, K.Okamoto, K.Ueyama, R.Shintani, T.Hayashi. *JACS*, **126**, 13584 (2004)
277. K.Okamoto, T.Hayashi, V.H.Rawal. *Org. Lett.*, **10**, 4387 (2008)
278. V.P.Ananikov, E.A.Khokhlova, M.P.Egorov, A.M.Sakharov, S.G.Zlotin, A.V.Kucherov, L.M.Kustov, M.L.Gening, N.E.Nifantiev. *Mendeleev Commun.*, **25**, 75 (2015)

279. M.North, P.Villuendas. *Org. Lett.*, **12**, 2378 (2010)
280. S.V.Kochetkov, A.S.Kucherenko, S.G.Zlotin. *Mendeleev Commun.*, **25**, 168 (2015)
281. V.Wagner, P.S.Schulz, P.Wasserscheid. *J. Mol. Liq.*, **192**, 177 (2014)
282. J.Canivet, E.Bernoud, J.Bonnefoy, A.Legrand, T.K.Todorova, E.A.Quadrelli, C.Mellot-Draznieks. *Chem. Sci.*, **11**, 8800 (2020)
283. C.Velez, B.Doherty, O.Acevedo. *Int. J. Mol. Sci.*, **21**, 1190 (2020)
284. B.L.Yano. *Toxicol. Sci.*, **65**, 288 (2002)
285. H.J.Kim, M.W.Ruszczycky, S.-h.Choi, Y.-n.Liu, H.-w.Liu. *Nature (London)*, **473**, 109 (2011)
286. C.D.Fage, E.A.Isiorho, Y.Liu, D.T.Wagner, H.-w.Liu, A.T.Keatinge-Clay. *Nature Chemical Biology*, **11**, 256 (2015)
287. H.J.Kim, M.W.Ruszczycky, H.-w.Liu. *Curr. Opin. Chem. Biol.*, **16**, 124 (2012)
288. H.A.Kirst. *J. Antibiot.*, **63**, 101 (2010)
289. N.Preiswerk, T.Beck, J.D.Schulz, P.Milovnik, C.Mayer, J.B.Siegel, D.Baker, D.Hilvert. *Proc. Natl Acad. Sci. USA*, **111**, 8013 (2014)
290. E.G.Gordeev, V.P.Ananikov. *PLOS ONE*, **10**, e0119984 (2015)
291. B.A.Hess, L.Smentek. *Org. Biomol. Chem.*, **10**, 7503 (2012)
292. A.Patel, Z.Chen, Z.Yang, O.Gutiérrez, H.-w.Liu, K.N.Houk, D.A.Singleton. *JACS*, **138**, 3631 (2016)
293. A.E.Reed, R.B.Weinstock, F.Weinhold. *J. Chem. Phys.*, **83**, 735 (1985)
294. M.G.Medvedev, A.A.Zeifman, F.N.Novikov, I.S.Bushmarinov, O.V.Stroganov, I.Y.Titov, G.G.Chilov, I.V.Svitanko. *JACS*, **139**, 3942 (2017)
295. C.Zhang, X.Wang, Y.Chen, Z.He, P.Yu, Y.Liang. *J. Org. Chem.*, **85**, 9440 (2020)
296. D.McLeod, M.K.Thøgersen, N.I.Jessen, K.A.Jørgensen, C.S.Jamieson, X.-S.Xue, K.N.Houk, F.Liu, R.Hoffmann. *Acc. Chem. Res.*, **52**, 3488 (2019)
297. B.R.Lichman, S.E.O'Connor, H.Kries. *Chem. – Eur. J.*, **25**, 6864 (2019)
298. B.Zhang, K.B.Wang, W.Wang, X.Wang, F.Liu, J.Zhu, J.Shi, L.Y.Li, H.Han, K.Xu, H.Y.Qiao, X.Zhang, R.H.Jiao, K.N.Houk, Y.Liang, R.X.Tan, H.M.Ge. *Nature*, **568**, 122 (2019)
299. N.Y.Kuznetsov, V.I.Malishev, M.G.Medvedev, Y.N.Bubnov. *Mendeleev Commun.*, **29**, 190 (2019)
300. B.-s.Jeon, M.W.Ruszczycky, W.K.Russell, G.-M.Lin, N.Kim, S.-h.Choi, S.-A.Wang, Y.-n.Liu, J.W.Patrick, D.H.Russell, H.-w.Liu. *Proc. Natl Acad. Sci. USA*, **114**, 10408 (2017)
301. P.Yu, A.Patel, K.N.Houk. *JACS*, **137**, 13518 (2015)
302. D.Hilvert. *Annu. Rev. Biochem*, **82**, 447 (2013)
303. G.Kiss, N.Çelebi-Ölçüm, R.Moretti, D.Baker, K.N.Houk. *Angew. Chem., Int. Ed.*, **52**, 5700 (2013)
304. Y.Zheng, W.Thiel. *J. Org. Chem.*, **82**, 13563 (2017)
305. D.Tan, C.S.Jamieson, M.Ohashi, M.-C.Tang, K.N.Houk, Y.Tang. *JACS*, **141**, 769 (2019)
306. L.Pauling. *Nature (London)*, **161**, 707 (1948)
307. A.A.Korlyukov, Y.V.Nelyubina. *Russ. Chem. Rev.*, **88**, 677 (2019)
308. B.Amul, S.Muthu, M.Raja, S.Sevvanthi. *J. Mol. Struct.*, **1210**, 128040 (2020)
309. L.Banci, F.Camponeschi, S.Ciofi-Baffoni, M.Piccioli. *J. Biol. Inorg. Chem.*, **23**, 665 (2018)
310. V.I.Polshakov, E.A.Batuev, A.B.Mantsyzov. *Russ. Chem. Rev.*, **88**, 59 (2019)
311. N.A.Besley. *Acc. Chem. Res.*, **53**, 1306 (2020)
312. V.A.Semenov, Y.Y.Rusakov, D.O.Samultsev, L.B.Krivdin. *Mendeleev Commun.*, **29**, 315 (2019)
313. V.A.Semenov, D.O.Samultsev, I.L.Rusakova, L.B.Krivdin. *J. Phys. Chem. A*, **123**, 4908 (2019)
314. V.A.Semenov, D.O.Samultsev, L.B.Krivdin. *Mendeleev Commun.*, **30**, 476 (2020)
315. D.B.Longley, D.P.Harkin, P.G.Johnston. *Nat. Rev. Cancer*, **3**, 330 (2003)
316. N.Vogt, D.N.Ksenafontov, D.Savelev, A.N.Rykov. *Mendeleev Commun.*, **30**, 660 (2020)
317. E.A.Denisova, D.B.Eremin, E.G.Gordeev, A.M.Tsedilin, V.P.Ananikov. *Inorg. Chem.*, **58**, 12218 (2019)
318. D.A.Boiko, E.O.Pentsak, V.A.Cherepanova, V.P.Ananikov. *Sci. Data*, **7**, 101 (2020)
319. A.Rassolov, G.Bragina, G.Baeva, I.Mashkovsky, A.Y.Stakheev. *Kinet. Catal.*, **61**, 869 (2020)
320. A.Rassolov, G.Bragina, G.Baeva, N.Smirnova, A.Kazakov, I.Mashkovsky, A.Y.Stakheev. *Kinet. Catal.*, **60**, 642 (2019)
321. A.V.Rassolov, D.S.Krivoruchenko, M.G.Medvedev, I.S.Mashkovsky, A.Y.Stakheev, I.V.Svitanko. *Mendeleev Commun.*, **27**, 615 (2017)
322. J.P.Troutman, H.Li, A.M.Haddix, B.A.Kienzle, G.Henkelman, S.M.Humphrey, C.J.Werth. *ACS Catalysis*, **10**, 7979 (2020)
323. Y.Wang, B.Wang, L.Ling, R.Zhang, M.Fan. *Chem. Eng. Sci.*, **218**, 115549 (2020)
324. A.V.Rassolov, G.N.Baeva, I.S.Mashkovsky, A.Y.Stakheev. *Mendeleev Commun.*, **28**, 538 (2018)
325. S.Han, K.Shin, G.Henkelman, C.B.Mullins. *ACS Catal.*, **9**, 4360 (2019)
326. W.Jiao, C.Chen, W.You, G.Chen, S.Xue, J.Zhang, J.Liu, Y.Feng, P.Wang, Y.Wang, H.Wen, R.Che. *Appl. Catal., B*, **262**, 118298 (2020)
327. Q.Jin, Y.He, M.Miao, C.Guan, Y.Du, J.Feng, D.Li. *Appl. Catal., A*, **500**, 3 (2015)
328. I.S.Mashkovsky, P.V.Markov, G.O.Bragina, G.N.Baeva, A.V.Rassolov, A.V.Bukhtiyarov, I.P.Prosvirin, V.I.Bukhtiyarov, A.Y.Stakheev. *Mendeleev Commun.*, **28**, 152 (2018)
329. V.L.Cruz, J.Martinez, J.Martinez-Salazar, J.Ramos, M.L.Reyes, A.Toro-Labbe, S.Gutierrez-Oliva. *Polymer*, **48**, 7672 (2007)
330. W.Yang, J.Yi, Z.Ma, W.-H.Sun. *Catal. Commun.*, **101**, 40 (2017)
331. S.Xiang, J.Wang. *Chin. J. Chem. Eng.*, **28**, 1805 (2020)
332. J.Fan, H.Du, Y.Zhao, Q.Wang, Y.Liu, D.Li, J.Feng. *ACSCatal.*, **10**, 13560 (2020)
333. C.Parra-Cabrera, C.Achille, S.Kuhn, R.Ameloot. *Chem. Soc. Rev.*, **47**, 209 (2018)
334. S.Rossi, A.Puglisi, L.M.Raimondi, M.Benaglia. *Catalysts*, **10**, 109 (2020)
335. C.N.Kelly, A.T.Miller, S.J.Hollister, R.E.Guldberg, K.Gall. *Adv. Healthcare Mater.*, **7**, 1701095 (2018)
336. C.R.Tubío, J.Azuaje, L.Escalante, A.Coelho, F.Guitián, E.Sotelo, A.Gil. *J. Catal.*, **334**, 110 (2016)
337. E.Bulatov, E.Lahtinen, L.Kivijärvi, E.Hey-Hawkins, M.Haukka. *ChemCatChem*, **12**, 4831 (2020)
338. J.Dhainaut, M.Bonneau, R.Ueoka, K.Kanamori, S.Furukawa. *ACS Appl. Mater. Inter.*, **12**, 10983 (2020)

339. M.Bible, M.Sefa, J.A.Fedchak, J.Scherschligt, B.Natarajan, Z.Ahmed, M.R.Hartings. *3D Print. Additive Manufact.*, **5**, 63 (2018)
340. S.Hock, M.Rose. *Chem. Ing. Tech.*, **92**, 525 (2020)
341. G.J.H.Lim, Y.Wu, B.B.Shah, J.J.Koh, C.K.Liu, D.Zhao, A.K.Cheetham, J.Wang, J.Ding. *ACS Mater. Lett.*, **1**, 147 (2019)
342. H.Thakkar, S.Eastman, Q.Al-Naddaf, A.A.Rownaghi, F.Rezaei. *ACS Appl. Mater. Inter.*, **9**, 35908 (2017)
343. C.A.Grande, R.Blom, V.Middelkoop, D.Matras, A.Vamvakeros, S.D.M.Jacques, A.M.Beale, M.Di Michiel, K.A.Andreassen, A.M.Bouzga. *Chem. Eng. J.*, **402**, 126166 (2020)
344. L.Song, X.Zhang, C.Chen, X.Liu, N.Zhang. *Microporous Mesoporous Mater.*, **241**, 36 (2017)
345. W.Jiang, R.Pei, S.-F.Zhou. *Chem. Eng. J.*, **400**, 125920 (2020)
346. J.S.Manzano, H.Wang, L.Qi, I.I.Slowing. *Appl. Catal., A*, **605**, 117794 (2020)
347. M.R.Hartings, Z.Ahmed. *Nat. Rev. Chem.*, **3**, 305 (2019)
348. C.Hurt, M.Brandt, S.S.Priya, T.Bhatelia, J.Patel, P.Selvakannan, S.Bhargava. *Catal. Sci. Tech.*, **7**, 3421 (2017)
349. J.S.Manzano, Z.B.Weinstein, A.D.Sadow, I.I.Slowing. *ACS Catal.*, **7**, 7567 (2017)
350. L.Zhang, X.Shi, Z.Zhang, R.P.Kuchel, R.Namivandi-Zangeneh, N.Corrigan, K.Jung, K.Liang, C.Boyer. *Angew. Chem., Int. Ed.*, **60**, 5489 (2021)
351. F.Magzoub, X.Li, S.Lawson, F.Rezaei, A.A.Rownaghi. *Fuel*, **280**, 118628 (2020)
352. K.D.Vogiatzis, M.V.Polynski, J.K.Kirkland, J.Townsend, A.Hashemi, C.Liu, E.A.Pidko. *Chem. Rev.*, **119**, 2453 (2019)
353. N.A.Tran, S.Jang, S.-W.Joo. *Appl. Surf. Sci.*, **497**, 143745 (2019)