Russian Chemical Reviews **78** (6) 495 – 511 (2009) © 2009 Russian Academy of Sciences and Turpion Ltd

DOI 10.1070/RC2009v078n06ABEH004032

Molecular modelling of receptors of physiologically active compounds for medicinal chemistry purposes †

I I Baskin, V A Palyulin, N S Zefirov

Contents

I. Introduction 495
II. Methods of molecular modelling of proteins 495
III. Molecular modelling of different types of biological receptors 502

Abstract. A general strategy of molecular modelling of biological receptors and their interaction with ligands is considered. Procedures for construction of protein 3D models, molecular docking, model quality assessment and determination of the free energy of protein binding with ligands are discussed. Methods of molecular design of novel drugs based on molecular models of biological targets, namely, virtual screening and *de novo* design are presented. Application of the approaches listed above to modelling of a number of pharmacologically significant receptors, analysis of receptor – ligand interactions and design of novel biologically active organic compounds is exemplified. The bibliography includes 314 references.

I. Introduction

Receptors of physiologically active compounds (PAC) usually represent partially glycosylated protein molecules present in cell membranes (or in cell organoid membranes). The biological function of such proteins is to transmit an external signal to a cell (or a cell organoid). Transmission occurs by binding molecules of a natural agonist (a neuromediator or a hormone). This causes specific conformational changes of the receptor molecule and is followed by intracellular (intraorganoid) signal transduction through a release of secondary messengers (in the case of metabotropic receptors) or initiation of transmembrane ionic currents (in the case of ionotropic receptors). It is receptors that are involved in control of main biological functions of organisms on the cell level.

Organic molecules introduced into an organism can also bind to the receptor protein macromolecules and thus cause conformational changes in them (in the case of full or partial agonists) or, *vice versa*, block them (in the case of antagonists), or change the intensity of intracellular response (in the case of modulators). Organic molecules control the biological function of an organism by affecting the transduction of intracellular signals. This underlies the possibility of treatment of various diseases with corresponding drugs. In addition to enzymes, PAC receptors belong to main types of biological targets for drugs. Therefore, knowledge of the 3D structures of PAC receptors and especially their binding sites with organic ligands is important for the development of novel drugs.

The main method of investigation of the 3D structure of biological macromolecules is X-ray analysis. Probably, in the near future these studies will also be carried out by NMR spectroscopy, but at present the application field of this method is limited to studies of small peptide molecules as well as segments or domains of complex protein molecules. The major limitation of X-ray analysis is the need to grow crystals of the objects under study. Crystals of small water-soluble proteins can be grown with relative ease, but crystals of proteins and protein complexes incorporated into cell membranes are very hard to grow. The last-mentioned type includes most PAC receptors localized in the cell membranes. The 3D structures of only a few such receptors have been experimentally investigated so far. That is why theoretical prediction of the 3D structure of PAC receptors and computer modelling of their interaction with molecules of biologically active organic compounds is a priority problem in medicinal chemistry and pharmacology.

II. Methods of molecular modelling of proteins

Molecular modelling of a PAC receptor usually involves construction of a crude model of the 3D molecular structure followed by model refinement, molecular dynamics studies of the dynamic properties of the receptor macromolecule and theoretical studies of protein – ligand interactions.

I I Baskin, V A Palyulin, N S Zefirov Department of Chemistry, M V Lomonosov Moscow State University, Leninskie Gory, 119991 Moscow, Russian Federation. Fax (7-495) 939 02 90, tel. (7-495) 939 35 57, e-mail: baskin@org.chem.msu.ru (I I Baskin), vap@org.chem.msu.ru (V A Palyulin), tel. (7-495) 939 16 20, e-mail: zefirov@org.chem.msu.ru (N S Zefirov)

Received 29 January 2009
Uspekhi Khimii **78** (6) 539 – 557 (2009); translated by A M Raevskiy

† This review completes the series of publications illustrating the main avenues of medicinal chemistry research [see *Russ. Chem. Rev.* **78** (5) (2009)].

1. Construction of protein 3D structures using comparative modelling

At present, the initial protein structure is most often constructed using comparative modelling, which is also called homology modelling. Here, the 3D model of the protein macromolecule is constructed starting from its amino acid sequence by transferring experimental information on the known template proteins to the proteins being modeled. This can be done either globally or locally.

On the global level, information transferred to the modelled protein includes the spatial structure of the template protein backbone specified by positions of C_α atoms (usually, but not necessarily, within such structurally conserved regions of the secondary structure of the protein molecule, as α-helices, β-layers, *etc.*). This approach is based on the idea that all native proteins are a result of long molecular evolution, which on the genetic level can be described as a sequence of mutations (with some complications due to exchange of genetic material by, *e.g.*, crossing-over). This means that the molecular evolution of proteins involved successive point substitution of amino acids and, probably, insertion or removal of small segments of amino acid sequences. To carry out the biological functions in a living organism, it is of crucial importance to conserve particular features of the 3D structure of the protein molecule.

For an organism to survive, all proteins should function correctly; therefore, the 3D structures of evolution-related native proteins should be similar at least on the global level. Hence, on the macrolevel, information on the 3D structures of native proteins can be transferred from one or more template proteins with experimentally determined 3D structures to the evolution-related native proteins whose molecular models should be constructed. The higher the similarity between the primary structures of the modeled protein and the template protein the closer the evolutionary relation between them and the better the 3D model. Proteins with similar amino acid sequences are always evolutionarily related and have similar 3D structures. However, the reverse is not always true because sometimes proteins with strongly different amino acid sequences have similar 3D structures. An example is provided by the family of G-protein coupled receptors (GPCR), which indicates a remote evolutionary relation between them. In such cases, other relevant criteria should be used.

Information transfer on the local level occurs while constructing side chains of all amino acids as well as for building protein loops. In these cases, the side-chain rotamer library and the protein loop database are used. This information is taken from the analysis of the 3D structures of all known proteins rather than the evolution-related proteins only.

Comparative modelling of proteins involves the following steps: 1) choice of the template protein; 2) amino acid sequence alignment for the modelled and template proteins; 3) modelling of the 3D structure of the protein under study; and 4) model quality assessment.[1] Usually, the first and second steps are carried out concurrently.

a. The choice of template protein and the amino acid sequence alignment

Comparative modelling begins with choosing an evolution-related template protein with the known 3D structure for the protein being modelled. This can be done by, *e.g.*, comparing the amino acid sequence of the modelled protein with those of all proteins with known 3D structures. The proteins for which the degree of similarity between the amino acid sequences is higher than a particular threshold value can be chosen as templates. In this case, the primary protein structures are compared using the sequence alignment procedure, which involves the search for such a mutual arrangement of the amino acid sequences in corresponding rows of a matrix, at which amino acids in each column exhibit identical or similar properties. If necessary, gaps are inserted; this is usually done if they correspond to the amino acids absent in the structurally conserved regions of the secondary structure, in particular α-helices and β-layers. There are two types of sequence alignment, namely, global (amino acid sequences of similar length are compared) and local (a sequence is compared with a segment of a longer sequence). Usually, global and local sequence alignment use algorithms based on the dynamic programming methods,[2,3] *viz.*, the Needleman–Wunsch algorithm[4] in the former and the Smith–Waterman algorithm[5] in the latter case. Both of them involve combinatorial minimization of the total penalty function for difference between the amino acids in the same matrix column. This function is calculated using preselected empirical substitution matrices[6] determining penalties for different types of amino acids. The most widely used types of substitution matrices include the PAM[7] and BLOSUM[8] matrix families, the latter being more preferable in homology modelling using the structures of evolutionary distinct proteins.[9] Sequence alignment can be either pairwise or multiple (involving three and more sequences). Usually, multiple alignment enables construction of more correct molecular models, even when using only the amino acid sequence of a single template protein.

The choice of the template protein and the sequence alignment often assume search in very large databases including all known amino acid sequences. This is usually done with such programmes, as FASTA[10] and BLAST[11,12] (for pair sequence alignment) and PSI-BLAST[12,13] (for multiple sequence alignment). However, it should be noted that the FASTA and BLAST programmes were designed to provide maximum performance necessary for operation with large databases and therefore the quality of sequence alignment is not always high. Because of this it is appropriate to perform sequence alignment using software designed for more thorough operation (*e.g.*, the CLUSTAL[14–16] and COFFEE[17,18] programme families) in the first step and then carry out molecular modelling.

An alternative method of search for the template protein and sequence alignment is called the threading method.[19,20] This is an efficient computational method of protein fold recognition[21] and prediction of protein structures.[22,23] It is based on the idea that the molecular evolution mainly occurs by creating mutations without changing the protein fold. The emergence of a new protein fold is a very rare event and therefore a relatively small number of protein folds (about 1500–1600) are known; they are responsible for a vast variety of protein structures formed in the course of evolutionary development. By all accounts, the emergence of much larger number of protein folds is impossible. Proteins experimentally studied to date by X-ray diffraction analysis and NMR spectroscopy belong to 1100 protein folds. Therefore, it is highly probable (probability is 70%–80%) that an arbitrarily chosen protein is characterized by a protein fold for which there is at least one compound with the known structure and, therefore, the

protein 3D structure can be predicted using comparative modelling.

In the threading method an amino acid sequence of the modelled protein is 'threaded' through all protein folds available in the library using dynamic programming methods.[2,3] Then, the protein fold, which best matches the sequence under study, is determined using a specially derived empirical energy function. The description of the protein fold includes characterization of its 3D structure; therefore, this procedure is also called the 1D – 3D-alignment. The empirical energy function used in this case is of statistical nature, *i.e.*, it is related by the Boltzmann relation to the occurrence probabilities of the given types of amino acids in certain structural environment (*e.g.*, outside or inside a protein globule and within each element of the secondary structure) and to the occurrence frequencies of these types of amino acids at a particular distance from one another.[24]

b. Modelling

Construction of the 3D structure of the protein molecule by comparative modelling is based on information obtained by sequence alignment of the modelled protein and one or more template proteins, on the 3D structures of the template proteins and on the statistics of distributions of local geometric parameters over the whole set of proteins with known 3D structures. At present, there are three main methods of modelling,[25] namely, 1) fragment assembly; 2) segment matching; and 3) transfer of statistical distributions of geometric parameters.

Fragment assembly. The fragment assembly technique is the historically first approach to constructing protein 3D structures by comparative modelling. The technique is based on the following empirical regularity: the structures of the proteins included in the closely related family (*e.g.*, in the family of mammalian serine proteases) comprise two types of segments.[26] The 3D structures of the first type of segments are strictly conserved within the family of proteins; therefore, they are often called 'structurally conserved regions'. Usually, they are built of stable elements of the secondary structure, in particular α-helices and β-sheets. Contrary to this, the 3D structures of the second type of elements vary significantly within the family of proteins. Usually, they correspond to loops in actual protein structures and therefore this type of segments is conventionally (within the context of this method of modelling) called loops.

The fragment assembly technique in the first stage involves assembly of structural fragments of the protein being modelled within the structurally conserved regions by transferring the atomic coordinates of the template protein backbones to the modelled protein. Then, a search for the structure in the loop database is carried out for each loop in this protein molecule. The database is formed using the results of analysis of all proteins with known 3D structures. The most appropriate among the loops selected, *i.e.*, those best matching the available protein fragments and the closest in homology to the corresponding amino acid sequences of the modelled loops, serve templates for specification of atomic coordinates of the protein backbone within the loops of the modelled protein. Then, the side-chain atomic coordinates are specified for all amino acids using the data retrieved from the rotamer library, which includes, for each type of amino acids and for each type of secondary structures, the values of torsion angles characterized by the highest occurrence frequency. A large number of modifications of this procedure are available, which differ in details.[27] The most widely known programme implementing this approach called COMPOSER[28] was a module of the SYBYL programme package.

Segment matching. Comparative modelling using the segment matching technique is based on the division of the modelled protein into small segments and modelling of each segment using particular segments of the protein molecule with known 3D structures as templates chosen upon local sequence alignment.[29]

Transfer of statistical distributions of geometric parameters. At present, the method of transfer of statistical distributions of geometric parameters is most often used in modelling protein 3D structures.[30] The technique can be considered as an extended method of distance geometry, which makes it possible to generate molecular conformations relevant to (possible) geometric constraints on interatomic distances.[31,32] A fundamental difference between the approach in question and the distance geometry method is that in the former case each geometric constraint is described by the probability density function of distribution of the corresponding geometric parameter parameterized using the set of proteins with known geometries rather than the exact minimum and maximum values of this parameter.[30] In the framework of this approach, two main types of the probability density function of the distribution of geometric parameters are used. These are the distances between C_α atoms (carbon atoms of the backbone) and internal geometric parameters (bond lengths, interatomic distances, bond angles and torsion angles) of the protein molecule. The distributions are approximated starting from the known geometry of the template proteins for the first type of geometric parameters and using a large set of proteins with known 3D structures for the second type.

By summing negative logarithms of the probability density functions for different geometric parameters one gets the molecular probability density function, which depends on the atomic coordinates in the protein molecule. This function is pseudoenergy in character and its value equals the sum of the contributions of the geometric parameters listed above. Each contribution is determined by the occurrence frequency of the corresponding parameter in the protein database with known geometries used in parameterization. All atomic coordinates of the modelled protein are determined by optimizing the molecular probability density function using nonlinear programming methods.[30] An example of software implementation of this method is provided by the programme MODELLER,[33–38] which was used for comparative modelling of a large number of PAC receptors. The comparative modelling mode of the programme Rosetta operates in a similar fashion.[39–41]

Accuracy of homology modelling. It is accepted that the accuracy of homology modelling depends on the degree of similarity between the amino acid sequences in the modelled protein and in the template protein. It is believed[25] that if the degree of similarity is higher than 50%, the corresponding models are quite reliable, possible errors usually concern only the side-chain conformations and the root-mean-square deviation from experimentally determined structures is ~1 Å. If the degree of similarity of the primary structures is in the range 30% – 50%, incorrect backbone conformations within loops appear; if this parameter is less than 30%, the folding type of the protein backbone also may

appear to be wrong and the root-mean-square deviation from experimental data may be larger than 3 Å.[25] However, these estimates should not be treated as a dogma, because often (in, *e.g.*, modelling of GPCR) a reasonable-quality model can be constructed at even lower degrees of similarity between the amino acid sequences.

c. Model quality assessment

Modelling of protein 3D structures is an ambiguous procedure because it is affected by a large number of factors including the choice of template proteins, details of the sequence alignment procedure, the choice of a random starting point for geometry optimization, *etc*. Therefore, the protein model quality assessment procedure is of great importance, because it allows one to select the best models, reveal errors in them and introduce necessary corrections to the modelling. At present, there are five main approaches to quality assessment of the 3D models of protein macromolecules, namely, comparison with results of biological experiments, the use of geometric criteria, the method of statistical potential, physicochemical approaches and statistical learning methods.

Comparison with results of biological tests. The most efficient and often used method of verification of protein 3D models is site-directed mutagenesis, which uses genetic engineering principles.[42, 43] In particular, it allows one to change the type of any amino acid in the protein molecule, which enables experimental investigations of the effect of this replacement on any protein function. In this case, the criterion for correctness of the 3D model is the possibility to explain the observed effect of the replacement of the amino acid on the protein function under study (*e.g.*, intracellular response under the action of a particular agonist localized in the cell membrane of the receptor of physiologically active compounds) using this model. In the latter case the most pronounced changes should occur in intracellular response upon replacement of amino acids in the nearest environment of the agonist molecule within the binding site.

Thus, using information obtained in the course of genetic engineering experiments on site-directed mutagenesis, one can carry out partial or even full mapping of the receptor molecule. This enables exact identification of the binding sites of different ligands and provides valuable information for the choice of correct 3D model. This principle underlies, in particular, iterative procedures for refinement of the model including a number of successive steps: 1) construction of the comparative model of a protein; 2) listing, according to the model constructed, of the amino acids forming the binding site; 3) replacement of these amino acids by, *e.g.*, alanine using site-directed mutagenesis; 4) measurements of functions of the mutant molecules; 5) comparison of experimental data with the results of prediction based on the protein 3D model; and 6) revision of the protein model in the case of mismatch (if necessary). This iterative procedure was used to construct a 3D model for the cholecycstokinine receptor.[44, 45]

However, one should not overestimate the potential of site-directed mutagenesis in verification of the quality of 3D models, because changes in the mutation-induced biological response can proceed by many different mechanisms, which can also not be related to position of a given residue in the active site of the protein. In particular, point mutations can cause global changes in the 3D structure of the protein molecule and influence the occurrence probability of conformational rearrangements necessary for functioning of the protein. In addition, mutations of the amino acids distant from the active site can influence the ligand access to the active site and the incorporation of the protein into cell membranes. Therefore, one should use additional criteria for correctness of models constructed.

The protein model quality can be assessed using biological experiments aimed at revealing glycosylation sites. In this case, the criterion for model correctness is the presence of the amino acids being glycosylated on the surface of the protein globule contacting the aqueous phase. This can also be achieved by elucidating palmitoylation sites, which should also be located on the protein surface, but they should be localized near the membrane edges.

Geometric criteria. Verification of the geometric criteria in the constructed protein 3D model involves calculations of all bond lengths and bond angles. Clearly, the deviation of at least one value from the standard values indicates local errors in the model. This is also indicated by the too short distance between two nonbonded atoms in the molecule. Many local errors can be eliminated in the course of subsequent model refinement during geometry optimization.

Calculations of the absolute stereochemical configurations of the C_α-atoms (carbon atoms of the protein backbone) and the C_β-atom of isoleucine side chain are a necessary step in model quality assessment. Clearly, the stereochemical configuration R of at least one of these centres indicates a severe error in the model. In addition, the stereochemical configurations of the atoms mentioned above should be determined repeatedly and in all stages of modelling, because an incorrect configuration can appear even in the final steps in the course of molecular geometry optimization by the molecular mechanics and molecular dynamics methods. This also holds for verification of the *trans*-configurations of all amide bonds in the protein backbone. Among the geometric criteria, one should also verify the out-of-plane deviation of atoms in all benzene rings in the phenylalanine and thyrosine residues, in heterocyclic rings in the tryptophan and histidine residues, in the guanidinium group of histidine, in all carboxyl and carbamide groups and in the environment of all amide bonds in the protein backbone.

In verifying geometric criteria one should analyze the distributions of the torsion angles φ and ψ in the protein backbone. For each type of amino acids, the values of these angles are used as coordinates in determining the position of corresponding amino acid in the Ramachandran plot. In this case, the role of criterion for correctness of the 3D model for the protein molecule is played by the number of amino acids in the forbidden regions in the Ramachandran plots. This parameter should not be much larger than the corresponding number in the experimentally determined 3D structures of the template proteins involved in comparative modelling.

The method of statistical potential. On the local level, the correctness of information transfer to the modeled proteins is controlled using the geometric criteria mentioned above. On the global level (see above), statistical potentials[24, 46–49] are used. In this case, a statistical potential is a function characterizing the relative‡ probability of a given 3D structure of the modelled protein. The probability is

‡ Relative to reference state with randomly arranged amino acids.

expressed in energy units using the Boltzmann relation. These functions are additive with respect to the contributions made by all structural features (*e.g.*, contacts between these types of amino acids or a particular structural environment of a given type of amino acids). Each contribution is determined by statistical analysis of the relative occurrence frequency of corresponding structural features in the database including all known protein 3D structures. Empirical functions used in the threading method (see above) also belong to statistical potentials.

Statistical potentials can be used in model quality assessment in a global or in a local mode. In the former case, the statistical potential is calculated for the entire protein molecule and then used to select the best model. This is just the case for the statistical potential DOPE [50, 51] used in the MODELLER programme.[33–38] In the local mode, the statistical potential is calculated for each amino acid residue and then used to plot the so-called 3D-profiles,[52] which characterize the dependence of the partial statistical potential on the number of the amino acid residue. The 3D-profiles can reveal particular, incorrectly modelled segments in the protein molecule. This methodology was implemented in a number of programmes and Web-services, in particular Verify3D [21, 52, 53] and ProSA.[46, 54] However, one should keep in mind that the methods mentioned above were parameterized for globular proteins. Therefore, in modelling membrane-bound receptors these methods should be used only to check the correctness of modelling of the extracellular and intracellular domains in aqueous environment.

Physicochemical approaches. Physicochemical methods of model quality assessment are based on the molecular mechanics calculations of the total energy of the protein molecule using the force fields with implicit inclusion of solvation effects. An example is provided by the EFF force field,[55, 56] which allows one to distinguish between the native conformation of the protein molecule and incorrectly constructed molecular models.

Statistical learning methods. Since there is a variety of methods for quality assessment of the 3D model for protein molecules, the choice of the most relevant techniques and their combination to provide higher quality of recognition of the correct molecular models compared to individual methods becomes topical. At present, this is usually done using modern statistical learning methods, such as artificial neural networks [57–59] and the support vector machines.[60–62] In a very early study, artificial neural networks were used to improve the performance of the GenTHREADER programme by combining the statistical potential used in the threading method, the scoring function describing the quality of the sequence alignment procedure, the length of the simulated amino acid sequence and some other characteristics.[63]

The support vector machines were used [51] to combine the statistical potential DOPE,[50, 51] three statistical potentials used in the system MODPIPE[§] and two scoring functions PSIPRED/DSSP [68–70] in order to derive a new scoring function, which permits best recognition of correct 3D model for protein molecules.

§ Automated system of comparative modelling of protein molecules based on the PSIBLAST programme [12, 13] for sequence alignment and on the MODELLER programme [33–38] for modelling of the 3D structure.[64–67]

2. *Ab initio* molecular modelling of proteins

A clearly seen drawback of comparative modelling methods is the impossibility of constructing 3D structure of the proteins provided no appropriate template protein.[71–75] Recently, template-free *ab initio* methods of molecular modelling have been used in these cases. The most widely used programme for *ab initio* molecular modelling is Rosetta,[76–80] which involves construction of a protein 3D model by assembly from microfragments using a stochastic procedure for minimization of the empirical energy function (in essence, it is similar to the statistical potentials mentioned above). The SimFold programme [81] operates similarly; here molecular modelling is accompanied by minimization of the same-name energy function, but this function is of physicochemical nature. One should keep in mind that *ab initio* modelling of large protein molecules gives less accurate models compared to those obtained by comparative modelling provided appropriate template proteins.

3. Model refinement and investigations of the dynamic properties of receptor macromolecules

Three-dimensional comparative modelling of the receptor molecule is with necessity followed by model refinement by optimizing the molecular geometry using the molecular mechanics [82, 83] with empirical force fields.[84] Often, the optimized geometry of the protein molecule is used as starting point in molecular dynamics simulation.[83, 85] In these cases, the molecule of a water-soluble protein (or a protein domain contacting water) is usually placed into a cube (sometimes, an octahedron or a sphere) filled with water molecules, thus simulating solvation; certain counterions are also added to obtain a zero total electric charge. The solvation effect is also included using continuum models for solvation (*e.g.*, the Born model). When simulating domains of PAC receptors incorporated into cell membranes, the model protein is embedded into a preliminarily prepared molecular model for lipid bilayer (for an example, see Ref. 86).

Molecular dynamics simulation is usually aimed at the following. First, simulation on short time scale allows one to refine the molecular geometry because the system can go out of shallow local energy minima, while simulation on long time scale enables monitoring of the effect of the medium (*e.g.*, membrane) on the protein structure. Second, molecular dynamics simulation on a particular time scale after attainment of equilibrium gives averaged values of geometric parameters and energy characteristics, which can be compared with experimental data. This particularly concerns investigations of protein–ligand complexes. Third, if the time scale of molecular dynamics simulation is longer than 5 to 10 ns, one can study the dynamic properties of macromolecules. As applied to simulation of PAC receptors, this means the possibility of studying conformational transitions in the protein macromolecule, which lead to receptor activation, which is necessary for search for potential drugs affecting the receptor fanction.

The molecular mechanics and molecular dynamics simulation of PAC receptors and their complexes with ligands is most often carried out with the following programmes and programme packages: AMBER,[87, 88] CHARMM,[89] NAMD,[90] GROMOS [91] and GROMACS.[92]

4. Critical assessment of methods for protein structure prediction

To obtain an unbiased assessment of which methods, software and computational servers make it possible to construct the most accurate protein 3D models, experiments on prediction of protein 3D structures starting with amino acid sequences have been carried out regularly since 1994. In the course of this experiment called Critical Assessment of Techniques for Protein Structure Prediction (CASP), research groups participate in a specific contest aimed at checkout of the predictive power of the approaches they have proposed. Each participant gets amino acid sequences of proteins with unknown 3D structures (only the organizers know them). All participants have to construct the 3D models of these proteins. The results of comparison of the model structures with the experimentally determined 3D structures are regularly (biannually) reported at the URL http://predictioncenter.org. In addition, detailed reports on the experiments are published in special issues of the journal '*Proteins*', which are also accessible from this server.

Thus, by studying the published literature one can obtain an objective picture of the potential of methods of protein 3D structure prediction, which helps any user to correctly choose the software for modelling of proteins of interest.

5. Modelling of ligand – receptor interaction

The last step in molecular modelling of PAC receptors is research on their interaction with ligands. This step is the most important step for drug design purposes. Analysis of ligand – receptor interactions is usually aimed at two main goals, *viz*., to describe in detail the mode of ligand binding to the binding site of the receptor and to use information thus obtained in search for new ligands for a given receptor.

a. Molecular docking

The main method for prediction of binding of a biological macromolecule to a ligand is molecular docking.[93–98] It allows the energetically most favourable arrangement of a ligand in the binding site of the biological macromolecule to be determined. At present, molecular docking and analysis of its results play a key role in the search for novel drugs,[99–104] thus being of great importance for medicinal chemistry. It should be emphasized that the results of docking of not only small organic molecules, but also small peptides and even whole protein molecules (the last-mentioned case is called protein – protein docking[105–109]) into biological targets are also important. Indeed, detailed knowledge of the structure of a complex between a receptor and a peptide ligand provides valuable information for the design of peptidomimetics (small organic molecules simulating the action of peptides),[110–115] while knowledge of the structure of protein – protein complexes facilitates the design of novel drugs whose biological action is due to inhibition of protein – protein interactions by small organic molecules.[116,117]

Initially, the aim of molecular docking was to find an optimum arrangement of a ligand molecule relative to the biological target (mainly, protein) molecule without considering, in explicit form, the possibility of structural rearrangements in each of them upon binding ('rigid' biological target – 'rigid' ligand docking). Here, the 'key' (ligand) – 'lock' (protein biological target) procedure is used.[118] The most efficient method for practical implementation of this technique is to describe the 'key' and 'lock' forms (by, *e.g*., the molecular surfaces or Fourier spherical harmonics[119–121]) and to search for such mutual arrangement of these forms in space, at which they can become complementary to each other.[122]

An alternative and more traditional method of docking is to search for the most energetically favourable 'key' – 'lock' configuration using stochastic optimization techniques (*e.g*., genetic algorithms and simulated annealing algorithms).[123] In the latter case, in order to improve the efficiency of calculations, the interaction energy is usually calculated by interpolating the values obtained from preliminary calculations at the latice nodes.[124] At present, almost all docking programmes include algorithms for taking into account conformational flexibility of small molecules ligands, while new versions of software also partially take into account protein flexibility at least by 'tuning' the side-chain amino acid conformations in the binding site.[125] Methods for inclusion of protein backbone flexibility in the course of docking were reported.[126–128] Usually, the key factor precluding the inclusion of the protein molecule flexibility is an abrupt decrease in the performance of software.

At present, the most widely used programmes for docking of small ligands are as follows: DOCK,[129–132] AUTODOCK,[123,133–135] GOLD,[136,137] Glide[138–140] and FlexX.[141–144] A number of comparative studies have been reported (see, *e.g*., Ref. 145). The most widely used programmes for protein – protein docking are RosettaDock,[146,147] ZDOCK[148,149] and PatchDock.[150]

b. Estimation of the free energy of binding and ligand ranking with respect to receptor affinity

Quantitative estimates of the free energy of protein – ligand interaction (in fact, binding constants) are valuable for the purposes of medicinal chemistry because it is this parameter that is responsible for the biological activity of a ligand. There are the nonempirical, semiempirical and empirical estimation techniques.

A nonempirical method of estimation of the free energy of protein – ligand interaction usually means a group of similar methods based on rigorous use of statistical thermodynamics formalism and differing in details of the integration procedure; these are the free energy perturbation technique, thermodynamic integration method and the slow growth method.[151–153] Basic to these methods is construction of a thermodynamic cycle including 'alchemical' stages of gradual transformation ('mutations') of a chemical structure to another one. Each 'mutation' involves a large number of intermediate 'non-physical' states. For each state, the trajectory-averaged internal energy of the system is determined by thorough molecular dynamics simulation and then used in the integration to calculate the free energy of 'mutation'. This procedure is rigorous and therefore the results thus obtained are in good agreement with experimental data. Nevertheless, its practical implementation is computationally too expensive due to the need for molecular dynamics simulation of a large number (at least several tens) of 'non-physical' states. As a consequence, practical implementation of the *ab initio* methods of free energy calculations for the purposes of medicinal chemistry is limited to investigations of the interaction of only one or a few candidate organic ligands with the biological target macromolecule.

Unlike the *ab initio* approaches, semiempirical methods do not involve molecular dynamics calculations of numer-

ous 'non-physical' states and therefore their computational cost is more than tenfold lower. One of the most widely used semiempirical methods is MM-PBSA (MM-GBSA),[154–157] which combines the molecular dynamics calculations of the average energy of protein–ligand interaction, estimation of conformational entropy of the complex by calculating normal vibrational frequencies, calculations of electrostatic component of the energy of solvation according to Poisson–Boltzmann[158] or using the generalized Born method,[159] and estimation of hydrophobic contribution to the energy of solvation using calculations of the solvent-accessible surface area.[160]

An alternative to the methods mentioned above is provided by a simpler, although less accurate, method of linear interaction energy (LIE).[161] It is based on estimation of the free energy of formation of a protein–ligand complex using an empirical relation including some fitting parameters which allow for the changes in the electrostatic and Lennard-Jones components of the internal energy upon complexation. The necessary contributions to the energy can be calculated using the molecular dynamics simulation.

Semiempirical methods can also be applied to small sets including a few tens of molecules; they are successfully used in the search for novel drugs. Nevertheless, the need for molecular dynamics simulation still makes semiempirical methods rather labor-consuming and therefore inappropriate for screening of large databases of organic compounds.

Very fast semiquantitative estimation of the free energy of receptor–ligand interaction or a simple ranking of ligands with respect to the receptor affinity can be done using empirical scoring functions.[162–165] To calculate them, one should only know the 3D structure of the protein–ligand complex, which can be determined upon completion of the docking procedure. Therefore, the scoring functions are usually calculated immediately after completion of this step. At present, there are three types of scoring functions, *viz*., molecular-mechanical, physicochemical and knowledge-based ones.

The molecular-mechanical scoring functions can be used for the determination of a correct position of a ligand within the receptor cavity in the course of docking and for approximate estimation of the free energy of complexation. These scoring functions represent the molecular-mechanical force fields augmented with means for estimation of the energy of solvation and conformational entropy. Examples are provided by the scoring functions in the programmes AUTODOCK,[123] GOLD (GoldScore function)[136,137] and the programme package Sybyl (D-Score function) (www.tripos.com).

The physicochemical (empirical) scoring functions include contributions of particular nature, that is, the contributions responsible for hydrophobic interaction, formation of hydrogen bonds and salt bridges, freezing of internal rotation in the course of complexation, *etc*. Coefficients at these contributions are usually chosen using multilinear regression analysis in such a manner that the experimental values of the free energies of complexation for the protein–ligand complexes with known 3D structures be reproducible. Examples are provided by the scoring functions Score1,[166] Score2,[167] Score,[168] X-Score,[169,170] and F-Score incorporated into the Sybyl programme package, as well as the functions ChemScore,[171,172] LigScore,[173] PLP and LUDI incorporated into the Cerius programme package[2] (http://accelrys.com/products/cerius2/) and HINT.[174]

The knowledge-based scoring functions obey the same principles as the statistical potentials mentioned above. Their calculation includes determination of the number of contacts between the ligand and protein atoms. The contribution of each type of contact is calculated from the corresponding relative occurrence frequencies in the database including known 3D structures of protein–ligand complexes after conversion to energy units using the Boltzmann relation. The numbers of contacts thus obtained can be used for ligand ranking with respect to protein affinity. In the families of related proteins, they correlate with the free energy of protein–ligand binding. Examples are provided by the scoring functions PMF,[175,176] DrugScore,[177] SmoG,[178–180] BHB,[181] and ITScore.[182,183]

Although a large number of scoring functions have been developed, none of them proved itself as reliable tool for estimation of protein–ligand binding constants or at least for ligand ranking with respect to protein affinity. Therefore, recently researchers have paid increasing attention to the approach based on the search for 'consensus' of predictions made using different scoring functions.[184–187] This method repeatedly proved itself as efficient. There are several strategies of search for 'consensus' for the scoring functions.[187] In addition, a pronounced effect is attained upon combining the scoring functions using statistical learning methods with experimental data.[188–190] These ideas also underlie a number of hybrid scoring functions, such as CSCORE[185] and GFScore.[191]

6. Molecular design of novel drugs

The value of the receptor models for medicinal chemistry purposes is that they can be used in the molecular design of novel drugs based on the knowledge of the 3D structures of biological targets.[104,192–202] To this end, ligands are docked into the molecular model of the receptor and then the ligands with the highest predicted receptor affinity are selected. Depending on the ligand source, virtual screening and *de novo* design are distinguished.

a. Virtual screening

In medicinal chemistry, the workflow of a virtual screening run involves 'sieving' of large databases of organic compounds through a set of computational 'sieves' in order to select potential drugs.[203–214] Usually, the databases include either catalogized data on commercially available organic compounds or virtual libraries of compounds, which can be synthesized with ease by, *e.g*., methods of combinatorial chemistry.[215] In the case of virtual screening[204,205,208] based on the knowledge of the 3D structures of biological targets, the main 'sieve' is a combination of docking to the molecular model of a receptor with the scoring procedure to select ligands with the highest affinity towards the biological target. As a rule, additional 'sieves' are also used, which makes it possible to reject, in the very early stage, the compounds with undesired pharmacokinetic properties and potential toxicants, as well as structures containing no pharmacophores necessary for binding to the desired biological target.

b. *De novo* design

In contrast to virtual screening using known ligand databases, *de novo* design[216–219] involves evaluation of the properties of novel ligands in the structure generation stage. New ligand structures are modelled within the molecular model for the binding site of the receptor.[220] In most

cases the *de novo* computational procedure is based on a genetic algorithm [220] (or 'particle swarm' method [221]) involving combinatorial optimization of the structures of organic compounds, which affords sets of ligands characterized by the highest predicted affinity to the receptor (as predicted by the scoring functions).

In the case of the design of novel drugs, an apparent drawback of virtual screening compared to *de novo* design is that the chemical compound libraries used in the screening procedure include only a very small fraction of the diversity of chemical structures. Therefore, in principle the *de novo* design methods make it possible to model ligand structures with much better target characteristics, which are not included in databases. A significant drawback of the *de novo* design is that chemical structures thus generated can appear to be synthetically difficult to access. Apparently, further development of this approach should be related to integration with methods of computer-assisted synthesis and automated assessment of synthetic accessibility of organic compounds.[222] The most widely known programmes for *de novo* design of novel drugs are LigBuilder,[223] LUDI,[224–226] the LeapFrog module incorporated into the Sybyl programme complex, ADAPT,[220] SPROUT (in combination with other two programmes, HIPPO for analysis of the receptor binding site and CAESA for assessment of synthetic accessibility of ligands generated),[227–229] PRO_LIGAND,[230,231] and AlleGrow (http://bostondenovo.com/Allegrow.htm).

III. Molecular modelling of different types of biological receptors

PAC receptors are complex structures comprising a number of domains and subunits, which can undergo global conformational transitions between different forms corresponding to the state of rest, activation and even desensitization. This is only one of many reasons (mention may also be made of multiplicity of binding sites, allosteric modulation, *etc*.) for great diversity of the types of influence of small organic molecules on the functioning of receptors. In particular, drug molecules can act as agonists (full, partial, neutral and inverse), co-agonists, antagonists (competitive and non-competitive), modulators (negative and positive), *etc*. Because of this, molecular modelling of different types of biological receptors and their complexes with organic ligands has some features, which will be briefly outlined in this Section. We will only dwell on those types of receptors which were modeled by the authors of this review.

1. G-protein coupled receptors

G-protein coupled receptors are biological targets for about 40% of modern drugs.[232] Functionally, this type of receptors is characterized by the following common features: all of them are proteins embedded into a cell membrane and transmit the signal into the cell by binding to agonist molecules 'fished' in the intercellular fluid and can undergo conformational rearrangements leading to interaction with G-protein molecule and formation of the cell response. In spite of very high pharmacological value of this type of receptors, the 3D structures have been determined by X-ray diffraction analysis very recently only for two (β_2-adrenergic receptor [233–235] and A_{2A}-adenosine receptor in a complex with an antagonist [236]) out of more than a thousand GPCR. This is the reason why methods of computer modelling have played and will play, in the foreseeable future, the leading role in studies of the 3D structures and functioning of GPCR and in the design of novel drugs relevant to these biological targets.[237–239]

Structurally, a common feature of all GPCR is an identical architecture of their transmembrane segments representing a bundle of seven transmembrane α-helices interconnected by three intracellular and three extracellular loops. Because of this the template used in the early studies on comparative modelling of GPCR was the bacteriorhodopsin molecule whose 3D structure was first determined by electron diffraction [240,241] and then by X-ray diffraction analysis.[242] It should be noted that initially only the transmembrane α-helices were simulated ignoring loops.

In the year 2000, the crystal structure of bovine rhodopsin,[243] a light-driven GPCR, was solved. Since GPCR are much more similar in structure and functioning to mammalian rhodopsin rather than bacteriorhodopsin; it is the rhodopsin structure that was chosen to be a template for comparative modelling. This is the reason for much higher accuracy of molecular models of receptors and for the possibility of modelling of most loops, which predetermined the use of these models for targeted search for novel drugs. In the year 2007, the determination of the crystal structure of the β_2-adrenergic receptor [233–235] made it possible to estimate the accuracy of the receptor models thus constructed. It was found that the transmembrane segment representing the most part of the ligand binding site can be modelled almost perfectly, but there are significant errors in modelling of the second extracellular loop, which sheathes the binding site.

In spite of identical 3D structure of the transmembrane segment of GPCR, there are significant differences between six classes of receptors belonging to this family. First of all, distinctions concern the amino acid sequences, the mode of receptor action and the structures of extracellular domains. Because of this, each class should be modelled in its own manner.

a. Type 1 GPCR

Type 1 GPCR include rhodopsin-like proteins characterized by a common mechanism of activation and signal transduction.[244] Owing to this, each transmembrane segment of the type 1 GPCR comprises one or more amino acid residues that are strictly conservative within this class and necessary for the mechanism to work. Such residues can be used for sequence alignment. A combination of sequence alignment using the structure of bovine rhodopsin as template is basic to most recent studies on molecular modelling of type 1 GPCR. Examples are provided by models of the 3D structures of adenosine (Fig. 1),[86,245–249] purine [250,251] and melatonin [252,253] receptors and their complexes with ligands.

There are many limitations on modelling of rhodopsin-like receptors using the rhodopsin structure as template. First of all, in the crystal the rhodopsin molecules are in the inactive state (they do not activate G-proteins and thus do not initiate intracellular response); as a result, comparative modelling leads to a conformation of the receptor molecule, which is also at rest. The molecular model thus constructed corresponds to the receptor conformation suitable for the formation of stable complexes with stabilizing inverse agonists. Based on the dose–response curve, the inverse agonists are usually identified as competitive antagonists; therefore, the models constructed can be used only in

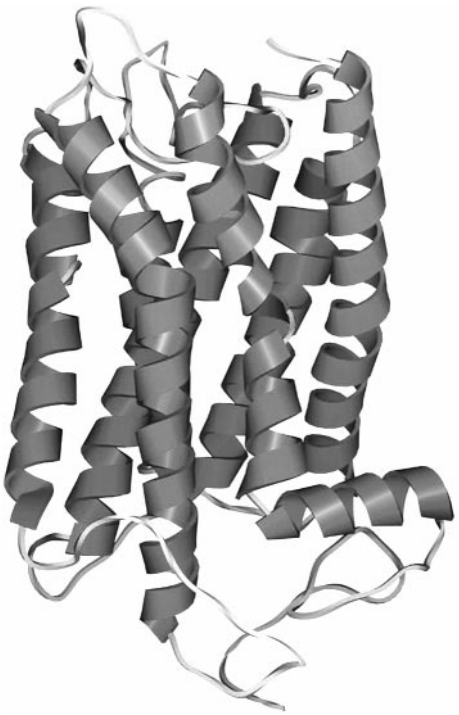


Figure 1. Molecular model for the A_{2B} adenosine receptor.

studies of receptor binding to antagonists and in targeted search for antagonists of this receptor.

If the aim of research efforts is to design agonists forming stable complexes with the activated form of the receptor, one of the following approaches is usually applied. First, there is no need to consider explicitly the conformational mobility of the receptor. One can model the binding of agonists and antagonists to the same 3D model and fix distinctions between agonists and antagonists only based on the type of the contacting segment in the binding site. Although the basic assumption that the receptor molecule has the same conformation when binding to all types of ligands is erroneous, this approach works quite correctly in practice. This is often explained using a model, according to which agonists initially bind to inactive form of the receptor and only then interact with particular centres of the binding site, thus initiating conformational rearrangements in the receptor molecule leading to its activation. Yet another explanation is the fact that the structure of the ligand binding site changes insignificantly upon receptor activation and therefore there should exist a strong correlation between the ligand affinity to both forms of the receptor. Nevertheless, the approach appeared to be inappropriate for solving some problems [254] because it is difficult to select potential agonists in the course of automated virtual screening or *de novo* design.

The second (more recent) approach is to construct individual 3D models for the inactive and active forms of the receptor. Since until recently detailed information on the activation mechanism of rhodopsin-like receptors was scarce, the 3D structure of the agonist-binding activated form of the receptor was modeled by rotating and displacing some transmembrane α-helices in the inactive form with allowance for possible activation mechanism to attain consistence with particular indirect experimental data. In particular, this approach was used to model [255] active forms of the rhodopsin-like receptors, which were used for docking of agonists in the course of virtual screening [255] and in 'inverse' screening, which permits indentification of possible biological targets for ligands.[256]

The third approach typical of recent studies is to construct a molecular model of metarhodopsin II (activated form of rhodopsin), which can be used as template for comparative modelling of agonist-binding activated form of the rhodopsin-like receptors. This was for the first time done by Nikiforovich *et al.*[257–259] who performed a search for optimum spatial arrangement of the transmembrane segments, which would be consistent with experimental data. Tikhonova *et al.*[260] reached the same goal by computer modelling of rhodopsin activation using specific versions of the molecular dynamics method and experimental values of particular interatomic distances. A kind of this approach is to construct a model for the activated form of the receptor from its inactive form by modelling the receptor activation stage using targeted molecular dynamics technique.[261]

Finally, the fourth method could probably involve the use of the 3D structure of metarhodopsin II, which was first determined by NMR spectroscopy [262] and then by X-ray analysis,[263] as a template. However, the accuracy of such models is too low.

A serious problem in modelling of rhodopsin-like receptors is prediction of the 3D structure of extracellular loops which immediately interact with bulky peptide ligands. Admissible molecular models can only be constructed by combining with numerous experimental data on site-directed mutagenesis and NMR spectroscopy data (see, *e.g.*, Ref. 45).

b. Type 2 GPCR

Unlike type 1 GPCR, which bind to relatively small ligand molecules and have binding sites located within the transmembrane segment, type 2 GPCR bind to large peptide hormone molecules as ligands (typical representatives are secretin and parathyroid hormone [264]). In addition to the transmembrane segment, type 2 GPCR also include the N-terminal domain, which also participates in ligand binding. The 3D models of the transmembrane segment of these receptors were formally constructed by comparative modelling using bovine rhodopsin as template;[256] however, the correctness of these models was not proved by experimental data. There is no similarity between the amino acid sequence of the N-terminal domain of the secretin receptor and the amino acid sequences of all other proteins with known 3D structures; therefore, the molecular model was constructed [265] by the *ab initio* method using experimental data.

c. Type 3 GPCR

Type 3 GPCR include metabotropic glutamate receptors [266, 267] and the $GABA_B$ receptor (GABA is γ-aminobutyric acid). These are dimeric structures whose monomeric fragments are built of the following three domains: N-terminal, transmembrane and 'cysteine-rich' (CRD). All agonists and competitive antagonists bind to the N-terminal domain (sometimes, it is analogized to venus Flytrap) because it has two hemispheres of a globule, which close as the agonist molecule appears between them. In turn, this causes a global conformational rearrangement of the entire dimer and subsequent signal transduction into the

cell. Non-competitive antagonists bind to the transmembrane domain.

Modelling of N-terminal domain. Pioneering theoretical studies on the 3D structures of the open and closed forms of the N-terminal domain of the $mGluR_1$ subtype and its interactions with ligands were carried out by comparative modelling using bacterial periplasmic proteins as templates.[268,269] Based on the results of modelling, it was hypothesized that agonists lie flat when binding to the protein and thus initiate the closure of the hemispheres of the globule, whereas antagonists are arranged across and preclude the closure. Since then the model for the open form (Fig. 2) was essentially refined by multiple sequence alignment[270] and the model for the closed form was refined using the threading method.[271] Based on the model for the open form of the N-terminal domain, significance of the Arg78 residue for ligand binding in the $mGluR_1$ subtype was hypothesized.[270] The hypothesis was immediately confirmed experimentally.[272] At the same time, taking account of the same model, the hypothesis of the antagonist binding mode (see above) casts some doubt.[273] The key results of modelling were then confirmed by X-ray diffraction analysis of the N-terminal domain of $mGluR_1$.[274] Then, using the $mGluR_1$ model as template, models for the N-terminal domains of all subtypes of metabotropic glutamate receptors were constructed;[275] this enabled comparative analysis of their binding sites.



Figure 2. Model for the open form of the N-terminal domain of $mGluR_1$.

Modelling of the transmembrane domain. Ligands binding to the transmembrane domain of the metabotropic glutamate receptor are very active and selective non-competitive antagonists. The structure of this domain was determined by comparative modelling using the rhodopsin 3D structure as template.[276–278] The models constructed were used in virtual-screening-based search for novel non-competitive antagonists.[277]

Modelling of the 'cysteine-rich' domain. The 3D models for the CRD of the metabotropic glutamate receptor were first constructed[279] by comparative modelling using the sequence alignment determined by the threading method.

d. Frizzled receptors

In addition to Smoothened receptors, frizzled receptors also called FZD receptors recently have been considered as type 6 GPCR¶ (Refs 280 and 281). The interaction of the FZD receptors with Wnt-proteins acting as ligands is basic to the Wnt–FZD signaling pathway, which plays an important role in embryogenesis and functioning of organisms. At the same time, violations in this signaling pathway are characteristic of the onset of malignant tumour.[282] Therefore, the FZD receptor–Wnt-protein binding site is a possible target for anticancer drugs.

The structure of the FZD receptors includes not only the transmembrane domain comprising seven α-helices, but also an extracellular CRD, which participates in the interaction with the Wnt-proteins.[283] The 3D structure of this domain was studied by X-ray diffraction analysis.[283] Owing to low solubility of the Wnt-proteins, the Wnt–FZD interactions have not been studied experimentally. This gave an impetus to molecular modelling of Wnt-proteins and their complexes with the FZD receptors.[284–286] Since it appeared to be impossible to find an appropriate template protein for comparative modelling, the 3D structure of the Wnt-protein was determined from *ab initio* calculations using the Rosetta programme.[76–80] Protein-protein docking was performed with the PatchDock server.[150] The molecular model for the Wnt–FZD complex was used in studies of the interaction of these proteins[284] and in the design of non-peptide inhibitors of this interaction.[285]

2. Ionotropic glutamate receptors

Ionotropic glutamate receptors are divided into three types with respect to the name of selective agonists, *i.e.*, 2-amino-3-(3-hydroxy-5-methylisoxazol-4-yl)propionic acid (AMPA receptors), *N*-methyl-D-aspartic acid (NMDA receptors) and kainic acid (kainate receptors).[4] These receptors are supramolecular complexes comprising a number of subunits which form an ion channel.[266,287,288] Passage of ionic current through this channel causes an intracellular response. Each subunit of the glutamate receptors is comprised of several domains, each having one or more binding sites to pharmacologically significant ligands.[266,287,288]

a. Simulation of extracellular agonist(co-agonist)-binding domains

Extracellular agonist(co-agonist)-binding domains include the binding sites of a natural agonist (glutamate) or a natural co-agonist (glycine or D-serine in the case of the NMDA receptor).[266,287,288] The glutamate- and glycine-binding sites of the NMDA receptor have similar 3D structures and are localized in different subunits of the receptor. Simultaneous binding to agonists and co-agonists causes conformational rearrangements resulting in receptor activation, opening of the ion channel and penetration of metal ions into the cell. On the contrary, competitive antagonists binding to the same sites preclude activation of the receptor.

Pioneering ideas of the 3D structure and activation mechanism of extracellular agonist-binding domains of the ionotropic glutamate receptors were based on homology with the amino acid sequences of some bacterial periplasmic binding proteins.[289] The first attempt to construct molecular models for the glutamate and glycine sites of the

¶ No data on models for the type 4 and type 5 GPCR are available as yet.

NMDA receptor by comparative modelling using the periplasmic protein LAOBP as template was reported by Laube *et al.*[290] In the year 2000, the results of X-ray diffraction study of an artificial protein based on the agonist-binding domain of the AMPA receptor in complexes with some ligands were reported.[291] This made it possible, due to relatively high homology between the amino acid sequences of the AMPA and NMDA receptors in the domain in question, to construct much better molecular models of the glutamate (Fig. 3)[292,293] and glycine (Fig. 4)[294,295] binding sites of the NMDA receptor. These models were used to explain the available structure–activity relationships for ligands of the glutamate site,[293] in constructing quantitative models for ligand binding to the glutamate site of the AMPA receptor[296] and to the glycine site of the NMDA receptor,[295] to develop the concept of 'selectivity fields',[297] and for virtual screening and creation of focussed libraries of potential ligands of the NMDA and AMPA receptors.[298]

b. Modelling of ion channel of the NMDA receptor

Molecular modelling of the ion channel of ionotropic glutamate receptors is based on architectural similarity to bacterial potassium ion channels.[299] The 3D model for the ion channel of the NMDA receptor (Fig. 5)[300] was constructed by comparative modelling with the structure of the potassium channel KcsA[301] as template and then used to study the interaction with non-competitive antagonist (channel blocker) MK-801 and to design its flexible analogues.[300] More recently, this model was refined and used for quantitative and qualitative modelling of the binding of channel blockers.[302]

c. Modelling of N-terminal modulator domains

A feature of ionotropic glutamate receptors is a large number of binding sites in the case of molecules modulators (both positive and negative) towards the NMDA receptor. Many pharmacologically significant organic compounds affecting the central nervous system have binding sites localized in the N-terminal domains of the NMDA receptor. This makes modelling of the 3D structures of these domains topical; such models can favour the design of novel drugs for treatment of neurodegenerative diseases. The first molecular models for domains were obtained[303] by comparative modelling using the 3D structure of the N-terminal domain of the metabotropic glutamate receptor $mGluR_1$[274]

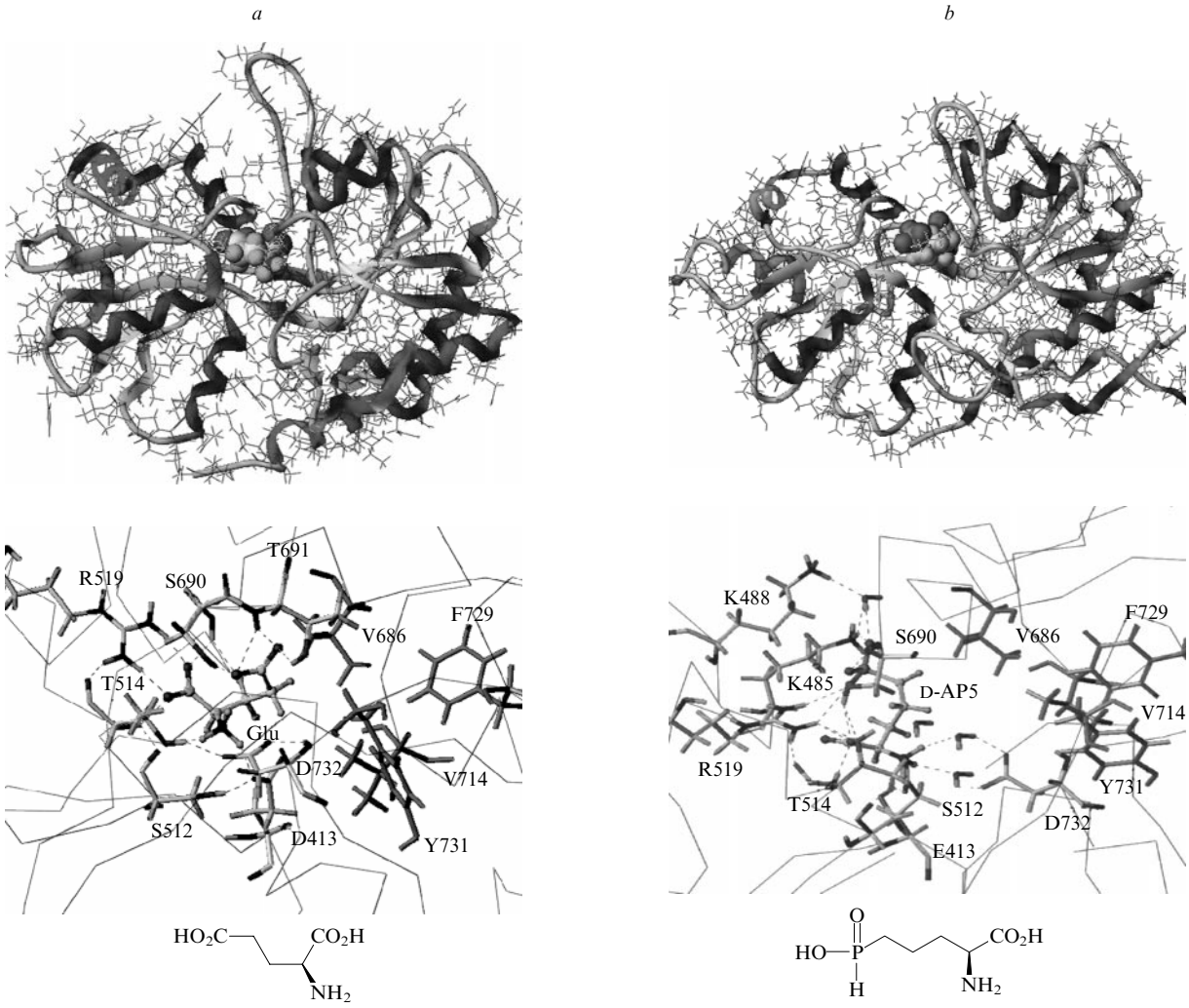


Figure 3. Complexes of agonist (glutamate) with the closed form (*a*) and antagonist (AP5) with the open form (*b*) of the glutamate-binding domain of the NMDA receptor.
Shown are the globule with ligand (above) and the binding site (below).

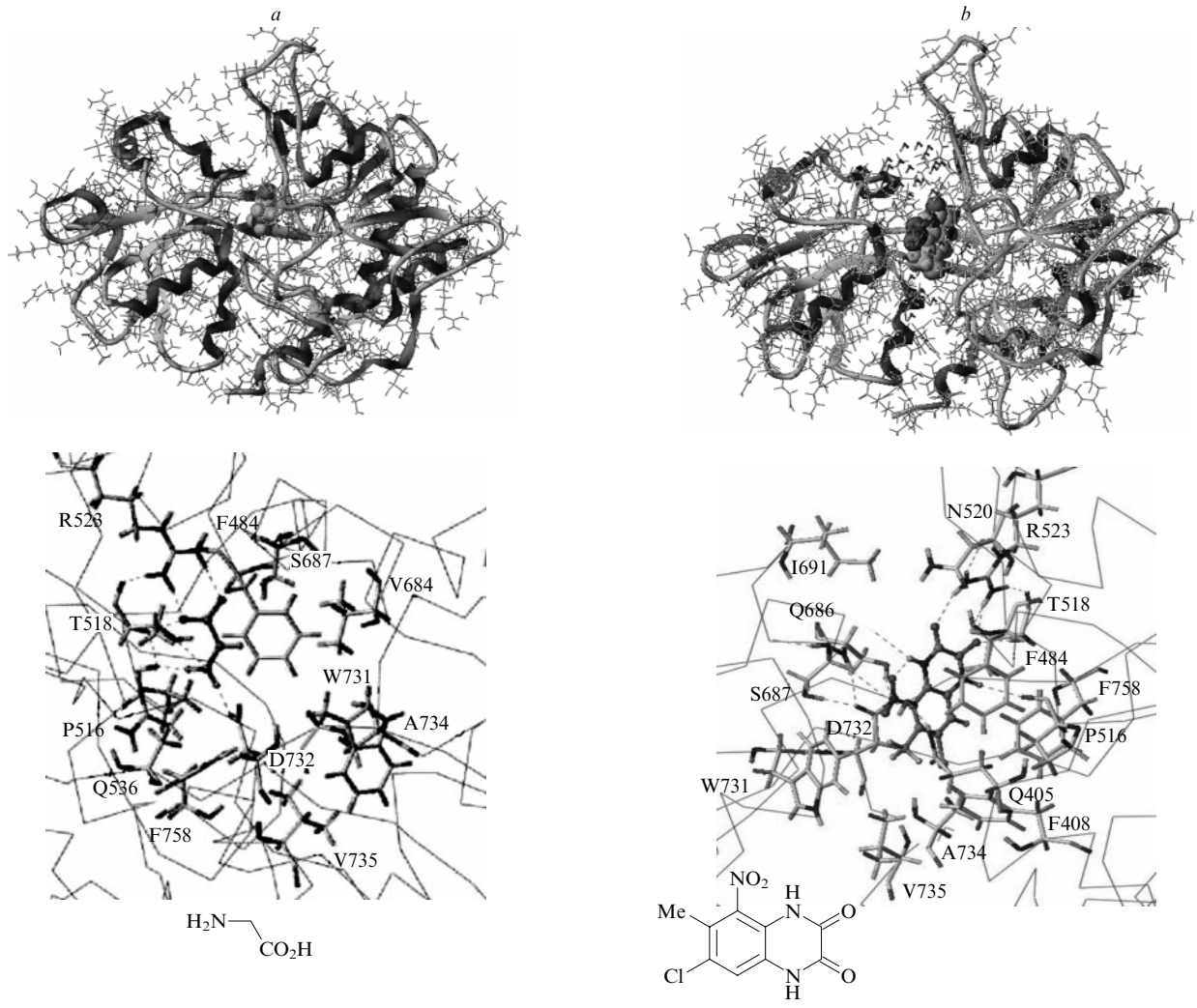


Figure 4. Complexes of agonist (glycine) with the closed form (*a*) and antagonist (7-chloro-6-methyl-5-nitroquinoxaline-2,3-dione) with the open form (*b*) of the glycine-binding domain of the NMDA receptor.
Shown are the globule with ligand (above) and the binding site (below).

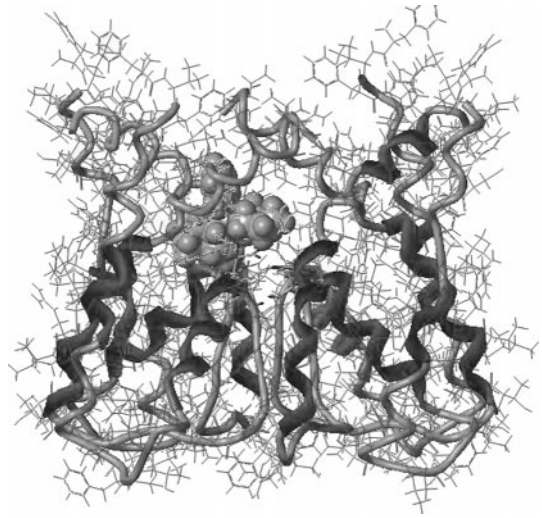


Figure 5. Model for the ion channel of the NMDA receptor with phencyclidine blocker.

as template. Fragments of the secondary structure of the target protein were predicted by the PSI-Pred programme [69] and sequence alignment was done by the threading method. The domain models thus constructed served as a basis for the assumption of binding modes of modulators of the NMDA receptor and for construction of quantitative models for binding of ifenprodil analogues.[303]

d. Modelling of binding of positive allosteric modulators to the AMPA receptor

Recently, particular attention in drug design has drawn to positive allosteric modulators of the AMPA receptor.[304, 305] An X-ray diffraction study revealed that their binding site is located between two subunits of the AMPA receptor. The action mechanism of positive modulators seems to consist in slowing down mutual rotation of the subunits, which is necessary for desensitization of the receptor (to close the ion channel, which was preliminarily open after binding to glutamate) and thus can enhance the magnitude and duration of the ionic current passing through the channel.[306] The binding site was used for modelling of the interaction of positive modulators with the AMPA receptor and in the design of novel drugs.[307, 308]

3. Receptors structurally similar to the nicotine receptor

The nicotine receptor and the subtypes A and C of the GABA receptor ($GABA_A$ and $GABA_C$, respectively), the subtype 3 of the serotonin receptors ($5\text{-}HT_3$) and the glycine receptor are similar in 3D structure to the nicotine receptor; they are sometimes called nicotinoid or cysteine-loop receptors. They represent ion-channel-forming pentameric supramolecular structures incorporated into the cell membrane. The interaction with agonist molecules causes changes in the receptor conformation and the opening of the ion channel. The results of X-ray diffraction studies of acetylcholine-binding protein [309] and its complexes with ligands [310, 311] are of great importance for investigations of the 3D structure of this family of receptors. These structures were used as templates to construct molecular models for the $GABA_A$ (see Ref. 312) (Fig. 6) and $GABA_C$ receptors [313, 314] by comparative modelling, to carry out theoretical analysis of the interaction of the receptors with agonist and competitive antagonist molecules and to propose a receptor activation mechanism. They can be used in virtual screening of novel potential active ligands.[313]



Figure 6. Model for the open form of the $GABA_A$ receptor.

* * *

At present, molecular modelling techniues play an important role in rational design of novel drugs. They are particularly important when choosing PAC receptors as biological targets for drugs. Taking into account improvements to the methodology of molecular modelling of proteins and protein–ligand interactions, significant improvement in performance of computational facilities and rapid accumulation of a great body of experimental data necessary to improve computational methods, one can expect an increasing significance of molecular modelling in medicinal chemistry. It is computational methods combining molecular modelling, bioinformatics, chemogenomics and chemoinformatics that will be in the vanguard of medicinal chemistry in the future.

References

1. M A Marti-Renom, A C Stuart, A Fiser, R Sanchez, F Melo, A Sali *Annu. Rev. Biophys. Biomol. Struct.* **29** 291 (2000)
2. R Giegerich, C Meyer, P Steffen *Sci. Comput. Progr.* **51** 215 (2004)
3. S R Eddy *Nat. Biotechnol.* **22** 909 (2004)
4. S B Needleman, C D Wunsch *J. Mol. Biol.* **48** 443 (1970)
5. T F Smith, M S Waterman *J. Mol. Biol.* **147** 195 (1981)
6. S F Altschul *J. Mol. Biol.* **219** 555 (1991)
7. M O Dayhoff, R Schwartz, B C Orcutt *Atlas of Protein Sequence and Structure* Vol. 5 (Suppl. 3) (Ed. M O Dayhoff) (Washington, DC: National Biomedical Research Foundation, 1978) p. 345
8. S Henikoff, J G Henikoff *Proc. Natl. Acad. Sci. USA* **89** 10915 (1992)
9. S R Eddy *Nat. Biotechnol.* **22** 1035 (2004)
10. D J Lipman, W R Pearson *Science* **227** 1435 (1985)
11. S F Altschul, W Gish, W Miller, E W Myers, D J Lipman *J. Mol. Biol.* **215** 403 (1990)
12. S F Altschul, T L Madden, A A Schaffer, J Zhang, Z Zhang, W Miller, D J Lipman *Nucleic Acids Res.* **25** 3389 (1997)
13. S F Altschul, D J Lipman *Proc. Natl. Acad. Sci. USA* **87** 5509 (1990)
14. D G Higgins *Methods Mol. Biol.* **25** 307 (1994)
15. J D Thompson, D G Higgins, T J Gibson *Nucl. Acid. Res.* **22** 4673 (1994)
16. M A Larkin, G Blackshields, N P Brown, R Chenna, P A McGettigan, H McWilliam, F Valentin, I M Wallace, A Wilm, R Lopez, J D Thompson, T J Gibson, D G Higgins *Bioinformatics* **23** 2947 (2007)
17. C Notredame, L Holm, D G Higgins *Bioinformatics* **14** 407 (1998)
18. C Notredame, D G Higgins, J Heringa *J. Mol. Biol.* **302** 205 (2000)
19. D T Jones, W R Taylor, J M Thornton *Nature (London)* **358** 86 (1992)
20. D T Jones, C Hadley *Bioinformatics: Sequence, Structure and Databanks* (Eds D G Higgins, W R Taylor) (Heidelberg: Springer, 2000) p. 3
21. J U Bowie, R Luthy, D Eisenberg *Science* **253** 164 (1991)
22. D T Jones *Methods Mol. Biol.* **143** 131 (2000)
23. D T Jones *Curr. Opin. Struct. Biol.* **10** 371 (2000)
24. S H Bryant, C E Lawrence *Proteins* **16** 92 (1993)
25. D Baker, A Sali *Science* **294** 93 (2001)
26. J Greer *J. Mol. Biol.* **153** 1027 (1981)
27. B Wallner, A Elofsson *Protein Sci.* **14** 1315 (2005)
28. C M Topham, P Thomas, J P Overington, M S Johnson, F Eisenmenger, T L Blundell *Biochem. Soc. Symp.* **57** 1 (1990)
29. M Levitt *J. Mol. Biol.* **226** 507 (1992)
30. A Sali, T L Blundell *J. Mol. Biol.* **234** 779 (1993)
31. T F Havel, G M Crippen, I D Kuntz, J M Blaney *J. Theor. Biol.* **104** 383 (1983)
32. T F Havel, I D Kuntz, G M Crippen *J. Theor. Biol.* **104** 359 (1983)
33. A Sali, L Potterton, F Yuan, H van Vlijmen, M Karplus *Proteins* **23** 318 (1995)
34. R Sanchez, A Sali *Proteins* (Suppl. 1) 50 (1997)
35. R Sanchez, A Sali *Methods Mol. Biol.* **143** 97 (2000)
36. A Fiser, A Sali *Methods Enzymol.* **374** 461 (2003)
37. N Eswar, B Webb, M A Marti-Renom, M S Madhusudhan, D Eramian, M Y Shen, U Pieper, A Sali *Curr. Protoc. Protein Sci., Ch. 2* Unit 2.9 (2007)
38. N Eswar, D Eramian, B Webb, M Y Shen, A Sali *Methods Mol. Biol.* **426** 145 (2008)
39. D Chivian, D E Kim, L Malmstrom, P Bradley, T Robertson, P Murphy, C E Strauss, R Bonneau, C A Rohl, D Baker *Proteins* **53** (Suppl. 6) 524 (2003)
40. C A Rohl, C E Strauss, D Chivian, D Baker *Proteins* **55** 656 (2004)

41. D Chivian, D Baker *Nucleic Acids Res*. **34** e112 (2006)
42. E P Garvey, C R Matthews *Biotechnology* **14** 37 (1990)
43. P N Bryan *Methods Mol. Biol*. **40** 271 (1995)
44. E Archer, B Maigret, C Escrieut, L Pradayrol, D Fourmy *Trends Pharmacol. Sci*. **24** 36 (2003)
45. E Archer-Lahlou, I Tikhonova, C Escrieut, M Dufresne, C Seva, L Pradayrol, L Moroder, B Maigret, D Fourmy *J. Med. Chem*. **48** 180 (2005)
46. M J Sippl *Proteins* **17** 355 (1993)
47. M J Sippl *J. Comput. Aided Mol. Des*. **7** 473 (1993)
48. K Park, M Vendruscolo, E Domany *Proteins* **40** 237 (2000)
49. P Narang, K Bhushan, S Bose, B Jayaram *J. Biomol. Struct. Dyn*. **23** 385 (2006)
50. M Y Shen, A Sali *Protein Sci*. **15** 2507 (2006)
51. D Eramian, M Y Shen, D Devos, F Melo, A Sali, M A Marti-Renom *Protein Sci*. **15** 1653 (2006)
52. D Eisenberg, J U Bowie, R Luthy, S Choe *Faraday Discuss*. 25 (1992)
53. R Luthy, J U Bowie, D Eisenberg *Nature (London)* **356** 83 (1992)
54. M Wiederstein, M J Sippl *Nucleic Acids Res*. **35** W407 (2007)
55. T Lazaridis, M Karplus *J. Mol. Biol*. **288** 477 (1999)
56. T Lazaridis, M Karplus *Proteins* **35** 133 (1999)
57. J Gasteiger, J Zupan *Angew. Chem., Int. Ed. Engl*. **105** 503 (1993)
58. N M Galbershtam, I I Baskin, V A Palyulin, N S Zefirov *Usp. Khim*. **72** 706 (2003) [*Russ. Chem. Rev*. **72** 629 (2003)]
59. I I Baskin, V A Palyulin, N S Zefirov *Ross. Khim. Zh*. **50** 86 (2006) [a]
60. V Vapnik *Statistical Learning Theory* (New York: Wiley-Interscience, 1998)
61. V N Vapnik *IEEE Trans. Neural Networks* **10** 988 (1999)
62. I Ovidiu, in *Reviews in Computational Chemistry* (Eds K B Lipkowitz, T R Cundari, D R Boyd) (New York: Wiley-VCH, 2007) P. 291
63. L J McGuffin, D T Jones *Bioinformatics* **19** 874 (2003)
64. U Pieper, N Eswar, H Braberg, M S Madhusudhan, F P Davis, A C Stuart, N Mirkovic, A Rossi, M A Marti-Renom, A Fiser, B Webb, D Greenblatt, C C Huang, T E Ferrin, A Sali *Nucleic Acids Res*. **32** D217 (2004)
65. U Pieper, N Eswar, F P Davis, H Braberg, M S Madhusudhan, A Rossi, M Marti-Renom, R Karchin, B M Webb, D Eramian, M Y Shen, L Kelly, F Melo, A Sali *Nucleic Acids Res*. **34** D291 (2006)
66. U Pieper, N Eswar, A C Stuart, V A Ilyin, A Sali *Nucleic Acids Res*. **30** 255 (2002)
67. R Sanchez, U Pieper, N Mirkovic, P I de Bakker, E Wittenstein, A Sali *Nucleic Acids Res*. **28** 250 (2000)
68. W Kabsch, C Sander *Biopolymers* **22** 2577 (1983)
69. L J McGuffin, K Bryson, D T Jones *Bioinformatics* **16** 404 (2000)
70. N Colloc'h, C Etchebest, E Thoreau, B Henrissat, J P Mornon *Protein Eng*. **6** 377 (1993)
71. A Kolinski, J Skolnick *Proteins* **18** 338 (1994)
72. J U Bowie, D Eisenberg *Proc. Natl. Acad. Sci. USA* **91** 4436 (1994)
73. K Yue, K A Dill *Protein Sci*. **5** 254 (1996)
74. R Srinivasan, G D Rose *Proteins* **22** 81 (1995)
75. D Chivian, T Robertson, R Bonneau, D Baker *Methods Biochem. Anal*. **44** 547 (2003)
76. K T Simons, C Kooperberg, E Huang, D Baker *J. Mol. Biol*. **268** 209 (1997)
77. K T Simons, R Bonneau, I Ruczinski, D Baker *Proteins* (Suppl. 3) 171 (1999)
78. K T Simons, I Ruczinski, C Kooperberg, B A Fox, C Bystroff, D Baker *Proteins* **34** 82 (1999)
79. R Bonneau, J Tsai, I Ruczinski, D Chivian, C Rohl, C E Strauss, D Baker *Proteins* (Suppl. 5) 119 (2001)
80. R Bonneau, C E Strauss, C A Rohl, D Chivian, P Bradley, L Malmstrom, T Robertson, D Baker *J. Mol. Biol*. **322** 65 (2002)
81. Y Fujitsuka, G Chikenji, S Takada *Proteins* **62** 381 (2006)
82. U Burkert, N Allinger *Molecular Mechanics* (Washington, DC: American Chemical Society, 1982)
83. T Schlick *Molecular Modeling and Simulation: An Interdisciplinary Guide* (New York: Springer, 2002)
84. A D Mackerell Jr *J. Comput. Chem*. **25** 1584 (2004)
85. J M Haile *Molecular Dynamics Simulation* (New York: Wiley, 1997)
86. A A Ivanov, I I Baskin, V A Palyulin, L Piccagli, P G Baraldi, N S Zefirov *J. Med. Chem*. **48** 6813 (2005)
87. D A Case, T E Cheatham III, T Darden, H Gohlke, R Luo, K M Merz Jr , A Onufriev, C Simmerling, B Wang, R J Woods *J. Comput. Chem*. **26** 1668 (2005)
88. J W Ponder, D A Case *Adv. Protein Chem*. **66** 27 (2003)
89. B R Brooks, R E Bruccoleri, B D Olafson, D J States, S Swaminathan, M Karplus *J. Comput. Chem*. **4** 187 (1983)
90. J C Phillips, R Braun, W Wang, J Gumbart, E Tajkhorshid, E Villa, C Chipot, R D Skeel, L Kale, K Schulten *J. Comput. Chem*. **26** 1781 (2005)
91. M Christen, P H Hunenberger, D Bakowies, R Baron, R Burgi, D P Geerke, T N Heinz, M A Kastenholz, V Krautler, C Oostenbrink, C Peter, D Trzesniak, W F van Gunsteren *J. Comput. Chem*. **26** 1719 (2005)
92. D Van Der Spoel, E Lindahl, B Hess, G Groenhof, A E Mark, H J Berendsen *J. Comput. Chem*. **26** 1701 (2005)
93. T Lengauer, M Rarey *Curr. Opin. Struct. Biol*. **6** 402 (1996)
94. I Halperin, B Ma, H Wolfson, R Nussinov *Proteins* **47** 409 (2002)
95. R D Taylor, P J Jewsbury, J W Essex *J. Comput. Aided Mol. Des*. **16** 151 (2002)
96. R T Kroemer *Biochem. Soc. Trans*. **31** 980 (2003)
97. V Mohan, A C Gibbs, M D Cummings, E P Jaeger, R L DesJarlais *Curr. Pharm. Des*. **11** 323 (2005)
98. S F Sousa, P A Fernandes, M J Ramos *Proteins* **65** 15 (2006)
99. D A Gschwend, A C Good, I D Kuntz *J. Mol. Recognit*. **9** 175 (1996)
100. R Abagyan, M Totrov *Curr. Opin. Chem. Biol*. **5** 375 (2001)
101. B K Shoichet, S L McGovern, B Wei, J J Irwin *Curr. Opin. Chem. Biol*. **6** 439 (2002)
102. J C Alvarez *Curr. Opin. Chem. Biol*. **8** 365 (2004)
103. D B Kitchen, H Decornez, J R Furr, J Bajorath *Nat. Rev. Drug Discov*. **3** 935 (2004)
104. R T Kroemer *Curr. Protein Pept. Sci*. **8** 312 (2007)
105. M J Sternberg, H A Gabb, R M Jackson *Curr. Opin. Struct. Biol*. **8** 250 (1998)
106. G R Smith, M J Sternberg *Curr. Opin. Struct. Biol*. **12** 28 (2002)
107. J J Gray *Curr. Opin. Struct. Biol*. **16** 183 (2006)
108. A M Bonvin *Curr. Opin. Struct. Biol*. **16** 194 (2006)
109. D W Ritchie *Curr. Protein Pept. Sci*. **9** 1 (2008)
110. T Kieber-Emmons, R Murali, M I Greene *Curr. Opin. Biotechnol*. **8** 435 (1997)
111. M G Bursavich, D H Rich *J. Med. Chem*. **45** 541 (2002)
112. J M Ahn, N A Boyle, M T MacDonald, K D Janda *Mini Rev. Med. Chem*. **2** 463 (2002)
113. J J Perez, F Corcho, O Llorens *Curr. Med. Chem*. **9** 2209 (2002)
114. B M Rajesh, J Iqbal *Curr. Pharm. Biotechnol*. **7** 247 (2006)
115. J Vagner, H Qu, V J Hruby *Curr. Opin. Chem. Biol*. **12** 292 (2008)
116. S Betzi, A Restouin, S Opi, S T Arold, I Parrot, F Guerlesquin, X Morelli, Y Collette *Proc. Natl. Acad. Sci. USA* **104** 19256 (2007)
117. S R Stauffer *Curr. Top. Med. Chem*. **7** 961 (2007)
118. W L Jorgensen *Science* **254** 954 (1991)
119. W Cai, X Shao, B Maigret *J. Mol. Graph. Model*. **20** 313 (2002)
120. R J Morris, R J Najmanovich, A Kahraman, J M Thornton *Bioinformatics* **21** 2347 (2005)
121. A Kahraman, R J Morris, R A Laskowski, J M Thornton *J. Mol. Biol*. **368** 283 (2007)
122. B K Shoichet, I D Kuntz, D L Bodian *J. Comput. Chem*. **13** 380 (1992)
123. G M Morris, D S Goodsell, R S Halliday, R Huey, W E Hart, R K Belew, A J Olson *J. Comput. Chem*. **19** 1639 (1998)
124. E C Meng, B K Shoichet, I D Kuntz *J. Comput. Chem*. **13** 505 (1992)
125. Y Zhao, M F Sanner *J. Comput. Aided Mol. Des*. **22** 673 (2008)
126. S Chaudhury, J J Gray *J. Mol. Biol*. **381** 1068 (2008)
127. N Kamiya, Y Yonezawa, H Nakamura, J Higo *Proteins* **70** 41 (2008)

128. N Andrusier, E Mashiach, R Nussinov, H J Wolfson *Proteins* **73** 271 (2008)
129. I D Kuntz *Science* **257** 1078 (1992)
130. I D Kuntz, E C Meng, B K Shoichet *Acc. Chem. Res.* **27** 117 (1994)
131. E C Meng, I D Kuntz *Drug Inform. J.* **28** 733 (1994)
132. D T Moustakas, P T Lang, S Pegg, E Pettersen, I D Kuntz, N Brooijmans, R C Rizzo *J. Comput. Aided Mol. Des.* **20** 601 (2006)
133. D S Goodsell, A J Olson *Proteins* **8** 195 (1990)
134. D S Goodsell, G M Morris, A J Olson *J. Mol. Recognit.* **9** 1 (1996)
135. G M Morris, D S Goodsell, R Huey, A J Olson *J. Comput. Aided Mol. Des.* **10** 293 (1996)
136. M L Verdonk, J C Cole, M J Hartshorn, C W Murray, R D Taylor *Proteins* **52** 609 (2003)
137. J C Cole, J W M Nissink, R Taylor *Virtual Screening in Drug Discovery* (Eds B Shoichet, J Alvarez) (Boca Raton, FL: Taylor Francis, CRC Press, 2005)
138. R A Friesner, J L Banks, R B Murphy, T A Halgren, J J Klicic, D T Mainz, M P Repasky, E H Knoll, M Shelley, J K Perry, D E Shaw, P Francis, P S Shenkin *J. Med. Chem.* **47** 1739 (2004)
139. T A Halgren, R B Murphy, R A Friesner, H S Beard, L L Frye, W T Pollard, J L Banks *J. Med. Chem.* **47** 1750 (2004)
140. M P Repasky, M Shelley, R A Friesner *Curr. Protoc. Bioinformat.* (Ch. 8) Unit 8.12 (2007)
141. H Claussen, M Gastreich, V Apelt, J Greene, S A Hindle, C Lemmen *Curr. Drug. Discov. Technol.* **1** 49 (2004)
142. S S Cross *J. Chem. Inf. Model.* **45** 993 (2005)
143. B Kramer, M Rarey, T Lengauer *Proteins* **37** 228 (1999)
144. I Schellhammer, M Rarey *Proteins* **57** 504 (2004)
145. J C Cole, C W Murray, J W Nissink, R D Taylor, R Taylor *Proteins* **60** 325 (2005)
146. S Lyskov, J J Gray *Nucl. Acids Res.* **36** W233 (2008)
147. C Wang, O Schueler-Furman, I Andre, N London, S J Fleishman, P Bradley, B Qian, D Baker *Proteins* **69** 758 (2007)
148. R Chen, L Li, Z Weng *Proteins* **52** 80 (2003)
149. K Wiehe, B Pierce, W W Tong, H Hwang, J Mintseris, Z Weng *Proteins* **69** 719 (2007)
150. D Schneidman-Duhovny, Y Inbar, V Polak, M Shatsky, I Halperin, H Benyamini, A Barzilai, O Dror, N Haspel, R Nussinov, H J Wolfson *Proteins* **52** 107 (2003)
151. M A McCarrick, P Kollman *Methods Enzymol.* **241** 370 (1994)
152. W Wang, O Donini, C M Reyes, P A Kollman *Annu. Rev. Biophys. Biomol. Struct.* **30** 211 (2001)
153. H J Woo, B Roux *Proc. Natl. Acad. Sci. USA* **102** 6825 (2005)
154. J Srinivasan, T E Cheatham III, P Cieplak, P A Kollman, D A Case *J. Am. Chem. Soc.* **120** 9401 (1998)
155. I Massova, P A Kollman *J. Am. Chem. Soc.* **121** 8133 (1999)
156. A Weis, K Katebzadeh, P Soderhjelm, I Nilsson, U Ryde *J. Med. Chem.* **49** 6596 (2006)
157. F Fogolari, A Brigo, H Molinari *Biophys. J.* **85** 159 (2003)
158. B Honig, A Nicholls *Science* **268** 1144 (1995)
159. B Jayaram, D Sprous, D L Beveridge *J. Phys. Chem. B* **102** 9571 (1998)
160. D Sitkoff, K A Sharp, B Honig *J. Phys. Chem.* **98** 1978 (1994)
161. J Aqvist, J Marelius *Combin. Chem. High Throughput Screen.* **4** 613 (2001)
162. R Wang, Y Lu, S Wang *J. Med. Chem.* **46** 2287 (2003)
163. R Wang, Y Lu, X Fang, S Wang *J. Chem. Inf. Comput. Sci.* **44** 2114 (2004)
164. X D Li, T J Hou, X J Xu *Acta Phys. Chim. Sinica* **21** 504 (2005)
165. A N Jain *Curr. Protein Pept. Sci.* **7** 407 (2006)
166. H J Böhm *J. Comput. Aided Mol. Des.* **8** 243 (1994)
167. H J Böhm *J. Comput. Aided Mol. Des.* **12** 309 (1998)
168. R Wang, L Liu, L Lai, Y Tang *J. Mol. Model.* **4** 379 (1998)
169. R Wang, L Lai, S Wang *J. Comput. Aided Mol. Des.* **16** 11 (2002)
170. C Obiol-Pardo, J Rubio-Martinez *J. Chem. Inf. Model.* **47** 134 (2007)
171. M D Eldridge, C W Murray, T R Auton, G V Paolini, R P Mee *J. Comput. Aided Mol. Des.* **11** 425 (1997)
172. C W Murray, T R Auton, M D Eldridge *J. Comput. Aided Mol. Des.* **12** 503 (1998)
173. A Krammer, P D Kirchhoff, X Jiang, C M Venkatachalam, M Waldman *J. Mol. Graph. Model.* **23** 395 (2005)
174. P Cozzini, M Fornabaio, A Marabotti, D J Abraham, G E Kellogg, A Mozzarelli *J. Med. Chem.* **45** 2469 (2002)
175. I Muegge, Y C Martin, P J Hajduk, S W Fesik *J. Med. Chem.* **42** 2498 (1999)
176. I Muegge *J. Med. Chem.* **49** 5895 (2006)
177. H F Velec, H Gohlke, G Klebe *J. Med. Chem.* **48** 6296 (2005)
178. R S DeWitte, A V Ishchenko, E I Shakhnovich *J. Am. Chem. Soc.* **119** 4608 (1997)
179. R S DeWitte, E I Shakhnovich *J. Am. Chem. Soc.* **118** 11733 (1996)
180. A V Ishchenko, E I Shakhnovich *J. Med. Chem.* **45** 2770 (2002)
181. M Feher, E Deretey, S Roy *J. Chem. Inf. Comput. Sci.* **43** 1316 (2003)
182. S Y Huang, X Zou *J. Comput. Chem.* **27** 1866 (2006)
183. S Y Huang, X Zou *J. Comput. Chem.* **27** 1876 (2006)
184. P S Charifson, J J Corkery, M A Murcko, W P Walters *J. Med. Chem.* **42** 5100 (1999)
185. R D Clark, A Strizhev, J M Leonard, J F Blake, J B Matthew *J. Mol. Graph. Model.* **20** 281 (2002)
186. M Feher *Drug Discov. Today* **11** 421 (2006)
187. A Oda, K Tsuchida, T Takakura, N Yamaotsu, S Hirono *J. Chem. Inf. Model.* **46** 380 (2006)
188. A E Klon, M Glick, J W Davies *J. Med. Chem.* **47** 4356 (2004)
189. R Teramoto, H Fukunishi *J. Chem. Inf. Model.* **47** 526 (2007)
190. S Costanzi, I G Tikhonova, M Ohno, E J Roh, B V Joshi, A O Colson, D Houston, S Maddileti, T K Harden, K A Jacobson *J. Med. Chem.* **50** 3229 (2007)
191. S Betzi, K Suhre, B Chetrit, F Guerlesquin, X Morelli *J. Chem. Inf. Model.* **46** 1704 (2006)
192. J Greer, J W Erickson, J J Baldwin, M D Varney *J. Med. Chem.* **37** 1035 (1994)
193. P M Colman *Curr. Opin. Struct. Biol.* **4** 868 (1994)
194. C L M J Verlinde, W G J Hol *Structure* **2** 577 (1994)
195. P J Whittle, T L Blundell *Annu. Rev. Biophys. Biomol. Struct.* **23** 349 (1994)
196. T L Blundell *Nature (London)* **384** 23 (1996)
197. T J Marrone, J M Briggs, J A McCammon *Annu. Rev. Pharmacol. Toxicol.* **37** 71 (1997)
198. K Gubernator, H J Böhm *Structure-Based Ligand Design, Methods and Principles in Medicinal Chemistry* (Weinheim: Wiley-VCH, 1998)
199. M Isiguro *Tanpakushitsu Kakusan Koso (Protein, Nucleic Acid, Enzyme)* **45** 880 (2000)
200. G Klebe *J. Mol. Med.* **78** 269 (2000)
201. A C Anderson *Chem. Biol.* **10** 787 (2003)
202. M M Daile, C Mait, P A Holt, J M Maguire, J B Meier, M C Miller, L Petraccone, J O Trent *Exp. Mol. Pathol.* **86** 141 (2009)
203. X Barril, R E Hubbard, S D Morley *Mini Rev. Med. Chem.* **4** 779 (2004)
204. S Ghosh, A Nie, J An, Z Huang *Curr. Opin. Chem. Biol.* **10** 194 (2006)
205. A Good *Curr. Drug Discov. Develop.* **4** 301 (2001)
206. A N Jain *Curr. Drug Discov. Develop.* **7** 396 (2004)
207. T Langer, R D Hoffmann *Curr. Pharm. Des.* **7** 509 (2001)
208. P D Lyne *Drug Discov. Today* **7** 1047 (2002)
209. C McInnes *Curr. Opin. Chem. Biol.* **11** 494 (2007)
210. I Muegge, S Oloff *Drug Discov. Today: Technolog.* **3** 405 (2006)
211. W P Walters, M T Stahl, M A Murcko *Drug Discov. Today* **3** 160 (1998)
212. A S Reddy, S P Pati, P P Kumar, H N Pradeep, G N Sastry *Curr. Protein Pept. Sci.* **8** 329 (2007)
213. G Schneider, H J Böhm *Drug Discov. Today* **7** 64 (2002)
214. B Waszkowycz *Curr. Drug Discov. Develop.* **5** 407 (2002)
215. D V S Green *Prog. Med. Chem.* **41** 61 (2003)
216. P M Dean *Mol. Biotechnol.* **37** 237 (2007)
217. T Honma *Med. Res. Rev.* **23** 606 (2003)
218. H Mauser, W Guba *Curr. Drug Discov. Develop.* **11** 365 (2008)

219. G Schneider, U Fechner *Nat. Rev. Drug Discov.* **4** 649 (2005)
220. S C H Pegg, J J Haresco, I D Kuntz *J. Comput. Aided Mol. Des.* **15** 911 (2001)
221. M Hartenfeller, E Proschak, A Schuller, G Schneider *Chem. Biol. Drug Des.* **72** 16 (2008)
222. J Gasteiger *J. Comput. Aided Mol. Des.* **21** 307 (2007)
223. R Wang, Y Gao, L Lai *J. Mol. Model.* **6** 498 (2000)
224. H J Böhm *J. Comput. Aided Mol. Des.* **6** 593 (1992)
225. H J Böhm *J. Comput. Aided Mol. Des.* **6** 61 (1992)
226. H J Böhm *J. Comput. Aided Mol. Des.* **8** 623 (1994)
227. V Gillet, A P Johnson, P Mata, S Sike, P Williams *J. Comput. Aided Mol. Des.* **7** 127 (1993)
228. V J Gillet, W Newell, P Mata, G Myatt, S Sike, Z Zsoldos, A P Johnson *J. Chem. Inf. Comput. Sci.* **34** 207 (1994)
229. V J Gillet, G Myatt, Z Zsoldos, A P Johnson *Persp. Drug Discov. Des.* **3** 34 (1995)
230. D E Clark, D Frenkel, S A Levy, J Li, C W Murray, B Robson, B Waszkowycz, D R Westhead *J. Comput. Aided Mol. Des.* **9** 13 (1995)
231. D R Westhead, D E Clark, D Frenkel, J Li, C W Murray, B Robson, B Waszkowycz *J. Comput. Aided Mol. Des.* **9** 139 (1995)
232. D Filmore *Mod. Drug Discov.* **7** 24 (2004)
233. V Cherezov, D M Rosenbaum, M A Hanson, S G Rasmussen, F S Thian, T S Kobilka, H J Choi, P Kuhn, W I Weis, B K Kobilka, R C Stevens *Science* **318** 1258 (2007)
234. S G Rasmussen, H J Choi, D M Rosenbaum, T S Kobilka, F S Thian, P C Edwards, M Burghammer, V R Ratnala, R Sanishvili, R F Fischetti, G F Schertler, W I Weis, B K Kobilka *Nature (London)* **450** 383 (2007)
235. D M Rosenbaum, V Cherezov, M A Hanson, S G Rasmussen, F S Thian, T S Kobilka, H J Choi, X J Yao, W I Weis, R C Stevens, B K Kobilka *Science* **318** 1266 (2007)
236. V P Jaakola, M T Griffith, M A Hanson, V Cherezov, E Y Chien, J R Lane, A P Ijzerman, R C Stevens *Science* **322** 1211 (2008)
237. M Ashton, M H Charlton, M K Schwarz, R J Thomas, M Whittaker *Combin. Chem. High Throughput Screen.* **7** 441 (2004)
238. P Jimonet, R Jäger *Curr. Drug Discov. Develop.* **7** 325 (2004)
239. S Radestock, T Weil, S Renner *J. Chem. Inf. Model.* **48** 1104 (2008)
240. N Grigorieff, T A Ceska, K H Downing, J M Baldwin, R Henderson *J. Mol. Biol.* **259** 393 (1996)
241. Y Kimura, D G Vassylyev, A Miyazawa, A Kidera, M Matsushima, K Mitsuoka, K Murata, T Hirai, Y Fujiyoshi *Nature (London)* **389** 206 (1997)
242. E Pebay-Peyroula, G Rummel, J P Rosenbusch, E M Landau *Science* **277** 1676 (1997)
243. K Palczewski, T Kumasaka, T Hori, C A Behnke, H Motoshima, B A Fox, I Le Trong, D C Teller, T Okada, R E Stenkamp, M Yamamoto, M Miyano *Science* **289** 739 (2000)
244. L A Rubenstein, R G Lanzara *J. Mol. Struct. (THEOCHEM)* **430** 57 (1998)
245. A A Ivanov, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **386** 548 (2002) [b]
246. A A Ivanov, I I Baskin, V A Palyulin, N S Zefirov *Vestn. Mosk. Univ., Ser. 2, Khim.* **43** 231 (2002) [c]
247. A A Ivanov, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **289** 404 (2003) [b]
248. A A Ivanov, V A Palyulin, N S Zefirov *J. Mol. Graph. Model.* **25** 740 (2007)
249. A A Ivanov, B Wang, A M Klutz, V L Chen, Z G Gao, K A Jacobson *J. Med. Chem.* **51** 2088 (2008)
250. A A Ivanov, H Ko, L Cosyn, S Maddileti, P Besada, I Fricks, S Costanzi, T K Harden, S V Calenbergh, K A Jacobson *J. Med. Chem.* **50** 1166 (2007)
251. A A Ivanov, S Costanzi, K A Jacobson *J. Comput. Aided Mol. Des.* **20** 417 (2006)
252. A A Ivanov, A E Voronkov, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **394** 822 (2004) [b]
253. A E Voronkov, A A Ivanov, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **403** 409 (2005) [b]
254. C Bissantz *J. Recept. Signal Transduct. Res.* **23** 123 (2003)
255. C Bissantz, P Bernard, M Hibert, D Rognan *Proteins* **50** 5 (2003)
256. C Bissantz, A Logean, D Rognan *J. Chem. Inf. Comput. Sci.* **44** 1162 (2004)
257. G V Nikiforovich, G R Marshall *Biochemistry* **42** 9110 (2003)
258. G V Nikiforovich, G R Marshall *Biochem. Biophys. Res. Commun.* **345** 430 (2006)
259. G V Nikiforovich, C M Taylor, G R Marshall *Biochemistry* **46** 4734 (2007)
260. I G Tikhonova, R B Best, S Engel, M C Gershengorn, G Hummer, S Costanzi *J. Am. Chem. Soc.* **130** 10141 (2008)
261. E Marco, M Foucaud, I Langer, C Escrieut, I G Tikhonova, D Fourmy *J. Biol. Chem.* **282** 28779 (2007)
262. G Choi, J Landin, J F Galan, R R Birge, A D Albert, P L Yeagle *Biochemistry* **41** 7318 (2002)
263. D Salom, D T Lodowski, R E Stenkamp, I Le Trong, M Golczak, B Jastrzebska, T Harris, J A Ballesteros, K Palczewski *Proc. Natl. Acad. Sci. USA* **103** 16123 (2006)
264. A J Harmar *Genome Biol.* **2** R3013 (2001)
265. W R Taylor, R E Munro, K Petersen, R P Bywater *Comput. Biol. Chem.* **27** 103 (2003)
266. H Brauner-Osborne, J Egebjerg, E O Nielsen, U Madsen, P Krogsgaard-Larsen *J. Med. Chem.* **43** 2609 (2000)
267. H Brauner-Osborne, P Wellendorph, A A Jensen *Curr. Drug Targ.* **8** 169 (2007)
268. G Costantino, R Pellicciari *J. Med. Chem.* **39** 3998 (1996)
269. G Costantino, A Macchiarulo, R Pellicciari *J. Med. Chem.* **42** 5390 (1999)
270. I I Baskin, M S Belenikin, G Kostantino, V A Palyulin, R Pellichari, N S Zefirov *Dokl. Akad. Nauk* **374** 347 (2000) [b]
271. M S Belenikin, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **383** 643 (2002) [b]
272. A A Jensen, P O Sheppard, P J O'Hara, P Krogsgaard-Larsen, H Brauner-Osborne *Eur. J. Pharmacol.* **397** 247 (2000)
273. M S Belenikin, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **384** 113 (2002) [b]
274. N Kunishima, Y Shimada, Y Tsuji, T Sato, M Yamamoto, T Kumasaka, S Nakanishi, H Jingami, K Morikawa *Nature (London)* **407** 971 (2000)
275. M S Belenikin, I I Baskin, G Kostantino, V A Palyulin, R Pellichari, N S Zefirov *Dokl. Akad. Nauk* **386** 261 (2002) [b]
276. M S Belenikin, G Kostantino, V A Palyulin, R Pellichari, N S Zefirov *Dokl. Akad. Nauk* **393** 827 (2003)
277. T Noeske, A Jirgensons, I Starchenkovs, S Renner, I Jaunzeme, D Trifanova, M Hechenberger, T Bauer, V Kauss, C G Parsons, G Schneider, T Weil *ChemMedChem* **2** 1763 (2007)
278. M Vanejevs, C Jatzke, S Renner, S Muller, M Hechenberger, T Bauer, A Klochkova, I Pyatkin, D Kazyulkin, E Aksenova, S Shulepin, O Timonina, A Haasis, A Gutcaits, C G Parsons, V Kauss, T Weil *J. Med. Chem.* **51** 634 (2008)
279. M S Belenikin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **394** 259 (2004) [b]
280. H C Huang, P S Klein *Genom. Biol.* **5** 234 (2004)
281. C C Malbon *Front. Biosci.* **9** 1048 (2004)
282. X He *Dev. Cell* **4** 791 (2003)
283. C E Dann, J C Hsieh, A Rattner, D Sharma, J Nathans, D J Leahy *Nature (London)* **412** 86 (2001)
284. A E Voronkov, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **412** 262 (2007) [b]
285. A E Voronkov, I I Baskin, V A Palyulin, N S Zefirov *J. Mol. Graph. Model.* **26** 1179 (2008)
286. A E Voronkov, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **419** 696 (2008) [b]
287. R Dingledine, K Borges, D Bowie, S F Traynelis *Pharmacol. Rev.* **51** 7 (1999)
288. P E Chen, D J Wyllie *Br. J. Pharmacol.* **147** 839 (2006)

289. P J O'Hara, P O Sheppard, H Thogersen, D Venezia, B A Haldeman, V McGrane, K M Houamed, C Thomsen, T L Gilbert, E R Mulvihill *Neuron* **11** 41 (1993)
290. B Laube, H Hirai, M Sturgess, H Betz, J Kuhse *Neuron* **18** 493 (1997)
291. N Armstrong, E Gouaux *Neuron* **28** 165 (2000)
292. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **382** 550 (2002) [b]
293. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov, S O Bachurin *J. Med. Chem.* **45** 3836 (2002)
294. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **382** 840 (2002) [b]
295. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *J. Med. Chem.* **46** 1609 (2003)
296. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **389** 123 (2003) [b]
297. I I Baskin, I G Tikhonova, V A Palyulin, N S Zefirov *J. Med. Chem.* **46** 4063 (2003)
298. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *Izv. Akad. Nauk, Ser. Khim.* 1282 (2004) [d]
299. T Kuner, P H Seeburg, H R Guy *Trends Neurosci.* **26** 27 (2003)
300. S Bachurin, S Tkachenko, I Baskin, N Lermontova, T Mukhina, L Petrova, A Ustinov, A Proshin, V Grigoriev, N Lukoyanov, V Palyulin, N Zefirov *Ann. NY Acad. Sci.* **939** 219 (2001)
301. D A Doyle, J M Cabral, R A Pfuetzner, A Kuo, J M Gulbis, S L Cohen, B T Chait, R MacKinnon *Science* **280** 69 (1998)
302. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **396** 551 (2004) [b]
303. I G Tikhonova, I I Baskin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **397** 549 (2004) [b]
304. A Arai, M Kessler, P Xiao, J Ambros-Ingerson, G Rogers, G Lynch *Brain Res.* **638** 343 (1994)
305. K A Yamada *Neurosci. Lett.* **249** 119 (1998)
306. Y Sun, R Olson, M Horning, N Armstrong, M Mayer, E Gouaux *Nature (London)* **417** 245 (2002)
307. I G Tikhonova, M I Lavrov, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **399** 268 (2004) [b]
308. P. RF 2333211 (2008); *Byull. Izobret.* (25) (2008)
309. K Brejc, W J van Dijk, R V Klaassen, M Schuurmans, J van Der Oost, A B Smit, T K Sixma *Nature (London)* **411** 269 (2001)
310. P H Celie, S E van Rossum-Fikkert, W J van Dijk, K Brejc, A B Smit, T K Sixma *Neuron* **41** 907 (2004)
311. Y Bourne, T T Talley, S B Hansen, P Taylor, P Marchot *EMBO J.* **24** 1512 (2005)
312. V I Chupakhin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **408** 693 (2006) [b]
313. D I Osolodkin, V I Chupakhin, V A Palyulin, N S Zefirov *Dokl. Akad. Nauk* **412** 404 (2007) [b]
314. D I Osolodkin, V I Chupakhin, V A Palyulin, N S Zefirov *J. Mol. Graph. Model.* **27** 813 (2009)

[a] — *Mendeleev Chem. J. (Engl. Transl.)*
[b] — *Dokl. Biochem. Biophys. (Engl. Transl.)*
[c] — *Moscow Univ. Chem. Bull. (Engl. Transl.)*
[d] — *Russ. Chem. Bull., Int. Ed. (Engl. Transl.)*