DOI: 10.1070/RCR4610

Molecular modelling and free-energy calculations of protein – ligand binding

Viktor B. Luzhkov

Institute of Problems of Chemical Physics, Russian Academy of Sciences
prosp. Akademika Semenova 1, 142432 Chernogolovka, Moscow Region, Russian Federation

Computational modelling of the structure and thermodynamics of microscopic models provides a valuable tool for research on the molecular mechanisms of biochemical processes and on the reactivity of physiologically active compounds. The review concerns methods of free-energy calculations of equilibrium systems, including various techniques developed to improve the accuracy and efficiency of computations. Calculations of the binding thermodynamics for model protein – ligand complexes are illustrated taking bacteriophage T4 lysozyme and β-trypsin as model proteins. The role and potential of molecular modelling methods in drug design are discussed.
The bibliography includes 237 references.

Contents

I. Introduction 211
II. Methods for free-energy calculations of equilibrium systems 212
III. Construction and analysis of microscopic models 221
IV. Search for new drugs and molecular modelling 225
V. Conclusion 227

I. Introduction

Targeted search for and design of new drugs and biomaterials is an important avenue of research in modern chemistry. A sufficiently complete and realistic description of processes involving biologically active compounds (BAC) can be obtained from *in silico* analysis of atom-molecular models of biochemical systems[1–4] which usually include a functionally significant fragment, namely, an attacking molecule and a biological target in local reaction environment (aqueous salt solution, biomembrane). Computer modelling of interactions between BAC and biomolecules is a rapidly progressing area in rational drug design.[1–8]

Binding to biomolecules plays a key role in molecular mechanisms of drug action. In some cases, biological action of a substance or its intermediate can be a result of chemical reactions in a living system, which involve dissociation and formation of covalent bonds. However, the activity of many modern drugs is due to direct blocking or signal regulation of biochemical processes through formation of noncovalent complexes with biotargets. Molecular recognition and selective binding processes are or paramount importance for living organisms and underlie various cell functions. In the human body, there are a few hundreds of targets of different nature, including G protein-coupled receptors, membrane ion channels and transport proteins, enzymes, *etc.* (see, *e.g.*, Refs 9 and 10). Target proteins have complex structures and contain multiple sites that are potentially accessible for binding small molecules (ligands). There are two types of binding, orthosteric and allosteric.[11] In the former case, the ligand directly blocks the active site of an enzyme. In the latter case, the ligand binds outside the active site and the functional properties of the target protein change due to perturbation of the free-energy surface upon complexation. In both cases, the apo form with the bound ligand or cofactor can be structurally different from the free protein. Important parameters also include the strength and selectivity of binding. Often, drugs have multiple effects on the human body, probably, due to interactions with several biotargets. From the standpoint of search for new active molecules it is important to monitor the lack of binding to known antitargets that may be responsible for toxic effects (*e.g.*, potassium ion channel hERG and cytochromes P450, see Ref. 10).

Calculations of the properties of molecular systems taking account of thermal motion of particles occupy a leading position among computer chemistry meth-

V.B.Luzhkov. Doctor of Chemical Sciences, Principal Researcher, Head of Research Group of Molecular Modelling of Biochemical Processes at the IPCP RAS.
Telephone: +7(496)522 – 1570, e-mail: vbl@icp.ac.ru
Current research interests: computer modelling of biologically active compounds.

Received 30 October 2015
Translation: A.M.Raevskiy

© 2017 Russian Academy of Sciences and Turpion Ltd.

ods.[1,2,4,12–25] These studies are described using the term 'molecular simulation'† defined as 'the numerical determination of the thermodynamic, energetic, structural, and dynamical properties of a mathematical model of a molecular assembly on a digital computer'.[15] Calculations of thermodynamic properties are key to the description of the reactivity of molecules, because the thermodynamic potentials, including the free energy, govern the directionality of chemical processes and equilibration in the system. Biochemical systems are characterized by occurrence of complex multiscale interactions of different nature (electrostatic, van der Waals, hydrophobic, *etc.*) Reliable description of the energetics of model systems is obtained using molecular mechanics and quantum mechanical methods. Modern scientific equipment allows one to investigate the dynamics of biomolecules in solution ($10^4 - 5\times10^5$ atoms) on the time scale of down to millisecond range.[21,23–25] Also, the available approaches make it possible to effectively overcome difficulties due to the wide range of characteristic times of the dynamics of biomolecules in solution ($10^{-12} - 10^2$ s). Computer simulations of microscopic systems give a reliable description of physicochemical properties, enable detailed examination of molecular mechanisms of biochemical processes and subsequent design of new molecules or optimization of the chemical structures of known active compounds.

Rapid progress in computer hardware and elaboration of new theoretical approaches and calculation algorithms predetermined intensive development of research on the molecular modelling of proteins, nucleic acids, biomembranes and BAC in the last twenty-five years. Achievements in molecular modelling of biochemical processes were marked by awarding the Nobel Prize in Chemistry of 2013 to M.Karplus, M.Levitt and A.Warshel 'for the development of multiscale models for complex chemical systems'.[23–25]

In this review we will consider main methods for free-energy calculations of equilibrium systems, outline modern methods for computer modelling of ligand binding to proteins and discuss the place and role of these approaches in drug design.‡ The material will be illustrated by the results of studies of complexes of bacteriophage T4 lysozyme and β-trypsin with various ligands.

Notations are as follows:

ADME — absorption, distribution, metabolism, excretion;
BAR — Bennett acceptance ratio method;
VS — virtual screening;
FEP — free energy perturbation;
DA — double annihilation;
LIE — linear interaction energies;
MD — molecular dynamics;
MC — Monte Carlo;
MM — molecular mechanics;
MS — molecular simulation;
GB — generalized Born formula;
RE — replica exchange;
PB — Poisson – Boltzmann;
PDB — Protein Data Bank;
PBC — periodic boundary conditions;
SA — surface area;
PMF — potential of mean force;
SBC — spherical boundary conditions;
SBDD — structure-based drug design;
FF — force field;
TI — thermodynamic integration;
TPT — thermodynamic perturbation theory.

II. Methods for free-energy calculations of equilibrium systems

From general considerations of statistical mechanics it follows that thermodynamic parameters of an equilibrium system can be determined if a complete description of energy levels is available.[26–31] On going from microscopic to macroscopic description a key role is played by the partition function Q (or the statistical integral in the case of classical treatment) of the system. A particular form of the function Q depends on the nature and statistical distribution of particles in the system under study. In most cases, numerical calculations of thermodynamic parameters of microscopic systems are carried out using the Boltzmann statistics, which is based on the assumption that the number of accessible energy levels much exceeds the number of particles in the system. This condition is met for atom-molecular ensembles at high temperatures[28] including those characteristic of biochemical processes. For instance, the statistical integral of the Gibbs canonical NVT ensemble (N identical noninteracting particles in the volume V at constant temperature T) has the form

$$Q(N,V,T) = \frac{1}{(h^{3N}N!)}\int\int \mathrm{e}^{-\beta H(\mathbf{p},\mathbf{r})}\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{r} \qquad (1)$$

where h is the Planck constant, $\beta = 1/kT$, k is the Boltzmann constant, $H(\mathbf{p},\mathbf{r})$ is the Hamiltonian of the system, $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_N)$ are momenta, and $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ are Cartesian coordinates of particles in the system. The Hamiltonian $H(\mathbf{p},\mathbf{r})$ is the sum of kinetic energy $K(\mathbf{p})$ and potential energy $U(\mathbf{r})$. The statistical integral is related to the Helmholtz free energy, F, of the system as follows:

$$F = -kT\ln Q \qquad (2)$$

Other thermodynamic functions, that is, energy E, entropy S, pressure P and chemical potential μ, can be determined from the potential F by differentiation.[28] The Boltzmann factor $\mathrm{e}^{-\beta H(\mathrm{p},\mathrm{r})}$ specifies the probability density of particles to occupy some volume of the phase space with the coordinates $\mathbf{p}$ and $\mathbf{r}$. The average value of an arbitrary function $\Phi(\mathbf{p},\mathbf{r})$ over the phase space is given by

$$\langle\Phi\rangle = \frac{\int\int \Phi(\mathbf{p},\mathbf{r})\mathrm{e}^{-\beta H(\mathbf{p},\mathbf{r})}\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{r}}{\int\int \mathrm{e}^{-\beta H(\mathbf{p},\mathbf{r})}\mathrm{d}\mathbf{p}\,\mathrm{d}\mathbf{r}} \qquad (3)$$

According to Gibbs' postulate, the average mechanical energy of canonical ensemble calculated using expression (3) equals the thermodynamic energy and the width of the mechanical energy distribution $H(\varepsilon)$ in the energy coordinates (ε) decreases as $\sqrt{N}$ and becomes infinitely small at

† The authors of Russian-language publications on theoretical chemistry and computer chemistry mainly use a broader term 'molecular modelling'. The field we have specified is part of 'molecular modelling'. In the text below it is termed 'molecular simulation'.

‡ Because of huge number of publications on the title topic we only selectively outline the key results and advances made in the field.

large N.[27, 28] Detailed information on the free energy surfaces is obtained from computer simulations of microscopic systems. This makes it possible to calculate the thermodynamic functions, equilibrium constants and direction of chemical processes.

Rigorous description of the potential energy surface $U(\mathbf{r})$ of a system can be obtained using quantum mechanical methods in the Born – Oppenheimer approximation.[14, 16, 19] However, quantum mechanical calculations of biomolecules in solution are computationally expensive, so in most cases the potential energies of microscopic models are determined by the force field method where the contributions to $U(\mathbf{r})$ are specified analytically. The force field (FF) of a molecular ensemble is written as the sum of contributions from groups of bound atoms and contributions of pair interactions between nonbonded atoms; the FF can be represented in different manner and the FF potentials can also be specified and parameterized in different manner.[1, 16, 17, 19, 32, 33] A base form of expression for the potential and kinetic energies of atom-molecular ensemble in the FF method has the form [17]

$$U(\mathbf{r}) = \sum_{k}^{N_{\text{bond}}} K_{bk}\Delta b_k^2 + \sum_{i}^{N_{\text{angle}}} K_{\theta l}\Delta\theta_l^2 + \qquad (4)$$

$$+ \sum_{m}^{N_{\text{tors}}} K_{\varphi m}\left[1 + \cos(n\varphi_m - \delta_m)\right] + \sum_{i<j}^{N_{\text{pair}}}\left(\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right)$$

$$K(\mathbf{p}) = \sum_{k}^{N_{\text{atom}}} \frac{\mathbf{p}_k^2}{2m_k}$$

The contribution of the energy of intramolecular interactions in expression (4) includes the sums of the harmonic potentials for a total of N_{bond} chemical bonds and N_{angle} bond angles and as well as the sum of the sinusoidal potentials for N_{tors} torsion angles, the energies of the corresponding potentials being calculated relative to the 'bottom' of the potential curves. The energy of nonbonding interactions is given by the sum of N_{pair} Lennard – Jones and Coulomb pair potentials acting between atoms i and j separated by the distance r_{ij}. The atoms in the system have the mass m and bear a fixed charge q. The force field parameters are determined using experimental data and results of quantum mechanical calculations.[1, 19]

The point atomic charges q are calculated using empirical schemes or quantum mechanically (also with inclusion of solvation effects).[1, 19] Equation (4) gives the electrostatic contribution for gas-phase interactions; however, the Coulomb potentials are sometimes corrected using a distance-dependent dielectric constant.[1, 19] The charges q are often corrected for polarization of molecules in condensed medium.[19, 32] Some problems require taking account of electronic polarization in explicit form,[1, 32] and the corresponding energies are included in the recent versions of the force fields of proteins, water and ligands.[33] The polarization contribution is difficult-to-model because of its cooperative nature; however, such calculations are performed increasingly widely.[33] Usually, the FF method ignores quantum effects for vibrational and rotational contributions; however, there are a few studies on the influence of these effects on the thermodynamic parameters being calculated.[34, 35] Potential energy calculations using analytical potentials are much more computationally efficient than rigorous quantum mechanical methods. By and large, the FF equations describe the shape of the potential function $U(\mathbf{r})$ of molecular ensembles near equilibrium positions with reasonable accuracy.

Numerical calculations of thermodynamic functions are carried out for statistical ensembles of states of a system, which contain complete information on microscopic states at each point of the system. Ensembles of states are generated using molecular dynamics [1, 12 – 14, 18, 19, 21, 36, 37] or Monte Carlo [1, 13, 18, 19] simulations. The interaction energies are described by the FF potentials. The dynamics of atomic motion are usually determined by numerically solving the Newton equations in Cartesian coordinates. Monte Carlo simulations only generate ensembles of spatial configurations using estimated probabilities of their appearance obtained from corresponding Boltzmann factors; the motion of atoms is not studied in explicit form. Computational modelling of the dynamics makes it possible to obtain larger sets from the phase space of biomolecules compared to construction of trajectories using MC simulations.[12] Average values at specified temperatures are calculated using statistical sets of system configurations upon thermodynamic equilibration. It should be noted that even very long trajectories are composed of a finite number of points from an infinitely large phase space of the system. In this case, estimates of thermodynamic parameters are stochastic because they depend on the choice of trajectory, so we can only say about their distribution near certain average values.

In practice, direct calculations of statistical integral and free energy using expressions (1) and (2) are possible only for simple systems. Since mechanical energy is an additive quantity, the expression for Q is separated into two factors corresponding to kinetic energy (ideal part) and potential energy (the excess part). The main difficulty is to obtain exact estimates of the excess contribution because direct calculation of Q would require uniform and sufficiently full sampling of system configurations. In MD and MC calculations the phase trajectories of atoms mainly pass through low-energy configurations because they are highly populated in the Boltzmann statistics. The contribution of high-energy states is included only partially and thus introduces uncertainty in the calculations of absolute free energy using expressions (1) and (2).[1, 13, 34] However, the mechanisms of chemical processes can be analyzed using relative rather than absolute free energies of states in the system.

Methods for free-energy calculations for various classes of chemical problems are available at the moment.[1 – 4, 12 – 25] More than 3500 articles concerning free-energy calculations by molecular modelling methods [38] have been published worldwide between 2000 and 2010 and the number of new publications continues to rapidly increase. It is of interest to compare the free energies of specified states and to determine the free-energy surfaces. Usually, calculations of the properties of biochemical models are carried out using the Gibbs free energy $G = F + PV$ for the (N,P,T) ensemble because it best corresponds to experimental conditions. The quantities ΔG and ΔF differ by $P\Delta V$; this contribution is small provided minor variations of the volume V and pressure $P = 1$ atm and is therefore usually ignored. Often, modern large-scale calculations allow ΔG values to be determined with an error of at most 1 kcal mol^{-1}; this is comparable with experimental accuracy. Main methods of determination of relative free energies are as follows:

— extrapolation of ΔG value using the results of molecular simulation of only one of the state being compared;

— summation of ΔG values along non-physical path between the states being compared;

— determination of ΔG and free energies of activation along the reaction coordinate between the states being compared; and

— calculations of ΔG using the results of simulation of the end states of the system.

These approaches are schematically shown in Fig. 1. Below we will consider methods for free-energy calculations of equilibrium systems and corresponding computing techniques.

II.1. Thermodynamic perturbation theory

Thermodynamic perturbation theory (TPT) is widely used in calculations of relative free energies of equilibrium states. In this method the free energy difference $\Delta F_{BA} \equiv F_B - F_A = \Delta F(A \rightarrow B)$ between the initial and perturbed states (A and B, respectively) with different Hamiltonians is treated as function of the perturbation potential $\Delta U(\mathbf{r}) = U_B(\mathbf{r}) - U_A(\mathbf{r})$ (case *1* in Fig. 1). The development of TPT began with a power series expansion of ΔF_{BA} in ΔU. The corresponding expression including the second-order terms has the form

$$\Delta F_{BA}(A \rightarrow B) = \langle \Delta U \rangle_A - \frac{\beta}{2} \langle (\Delta U - \langle \Delta U \rangle)^2 \rangle_A \qquad (5)$$

The angle brackets $\langle ... \rangle_A$ denote the averaging over the phase space of the state A.

Formula (5) appeared for the first time in a monograph [26] (1st Ed., see §). Various methods for derivation of this can be found in the literature.[1, 27, 40]

In addition to the series expansion, the free-energy difference ΔF_{BA} can be expressed as true function of perturbation potential ΔU

§ Mention was made of R.Peierls' contribution to the development of thermodynamic perturbation theory.[26, 39]

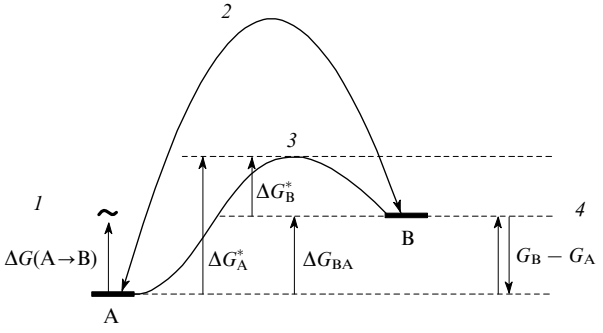


Figure 1. Free-energy levels of states A and B: a schematic. Free-energy (ΔG) calculations for the perturbation A→B (*1*); non-physical path between states A and B (*2*); reaction coordinate (ΔG^*_A and ΔG^*_B are the free energies of activation for the motion from states A and B, respectively) (*3*); and free-energy (ΔG) calculations from the free energies of the end states (*4*).

$$\Delta F_{BA}(A \rightarrow B) = -\frac{1}{\beta} \ln \left(\frac{Z_B}{Z_A} \right) = -\frac{1}{\beta} \ln \langle e^{-\beta \Delta U(\mathbf{r})} \rangle_A \qquad (6)$$

where Z is the configurational integral of the system, which is, up to a factor, the excess part of the statistical integral

$$Z = \int e^{-\beta U(\mathbf{r})} d\mathbf{r} \qquad (7)$$

Equation (6) reported by R Zwanzig [40] in 1954 gives the limiting value for the power series expansion of ΔF_{BA} in ΔU and represents the essence of the free energy perturbation method. The method was considered in detail and is widely used in applied chemical and biological studies (see, *e.g.*, Refs 2, 16, 17, 20, 38, 39, 41 – 71).

The exponential average (6) is numerically calculated from MD or MC trajectories ignoring the contribution of kinetic energy because it is cancelled in the expression for ΔF_{BA} owing to separation of the integration over momenta and coordinates, uniform kinetic energy distribution over degrees of freedom and(or) through the use of thermodynamic cycles (see below).[51, 59]

When calculating the free-energy difference using formula (6), the states being compared, A and B, are generally, identical from the standpoint of their treatment as the end and perturbed one. For this reason, to make numerical estimates more realistic, both states are often modelled independently and the free-energy difference ΔF is calculated in the forward (A → B) and reverse (B → A) directions. The total free-energy difference, ΔF_{BA}, is determined as the arithmetic mean of the exponential averages for two final states

$$\Delta F_{BA} = -\frac{1}{2\beta} \ln \frac{\langle e^{-\beta \Delta U(\mathbf{r})} \rangle_A}{\langle e^{\beta \Delta U(\mathbf{r})} \rangle_B} \qquad (8)$$

Expression (8) is termed the double-end FEP method.[39, 42]

A common condition applicability of TPT is smallness of ΔU compared to kT.[13, 20, 26] In applied studies, the states being compared are usually characterized by much larger potential energy differences. The problem is solved by augmenting the computational scheme with intermediate states whose potentials are parameterically coupled to the potentials at the end points [41, 42, 72]

$$U(\mathbf{r}, \lambda_m) = (1 - \lambda_m) U_A(\mathbf{r}) + \lambda_m U_B(\mathbf{r}) \qquad (9)$$

$$\lambda_m = 0 \rightarrow 1; \; m = 1, ..., n$$

The coupling parameter λ changes from 0 (for state A) to 1 (for state B). At small increment of the parameter λ the ΔU values for the intermediate states become sufficiently small, which improves the accuracy of calculations. The total free-energy difference is obtained by summation of intermediate contributions

$$\beta \Delta F_{BA}(\lambda = 0 \rightarrow 1) = -\sum_{m=1}^{n-1} \ln \langle e^{-\beta \Delta U(\mathbf{r}, \lambda_m \rightarrow \lambda_{m+1})} \rangle_m \qquad (10)$$

Transitions between states upon variation of λ correspond to the motion along curve *2* in Fig. 1. The distribution pattern of λ values should be chosen by researchers

depending on the problem to be solved,[39, 59, 64] however, it seems also appropriate to apply the condition $\Delta F(\lambda_m \rightarrow \lambda_{m+1}) \leqslant 2kT$ (see Ref. 15). The state of the system is simulated separately for each λ value and data for calculations of exponential averages at different points of the phase trajectory of a given state are collected after equilibration. The contributions of the states being simulated can be combined in different manner.[39, 43, 61, 62] Practical aspects of free-energy calculations are well documented.[13, 16, 18, 73]

In principle, the Zwanzig equation gives an exact free-energy difference between states,[15] but numerical calculations yield an estimate only. Similarly to evaluation of absolute free energies, difficulties in attainment of sufficiently high accuracy in FEP calculations are related to nonuniform sampling over the configuration space for finite phase trajectories of the state being modelled. There are stochastic and systematic errors in the ΔF calculations. The former depends on the completeness of the statistical set with respect to thermally accessible configurations. The latter is determined by low-populated regions of the configuration space and, in particular, depends on the overlap between the coordinate spaces of the populated configurations of the ground and perturbed states.[15, 20, 61] The systematic error in double-end FEP calculations using formula (8) can be evaluated from the ΔF hysteresis in the forward and reverse directions, defined by the difference between absolute free energies. In applied studies the accuracy of calculations of ΔF_{BA} values using expressions (6) – (10) is at most 1 kcal mol^{-1}, being much smaller than the error in direct calculations of ΔU_{BA} and ΔS_{BA}. Relatively low error in calculations of ΔF_{BA} is due to calculations of exponential averages of ΔU over the statistical ensemble, whereas, *e.g.*, ΔU_{BA} is calculated as the difference between the total average potential energies (large numbers).[15, 51]

An intuitively expected advantage of the modelling of both states being compared for the determination of free energy differences is successively implemented in the Bennett acceptance ratio method.[74] In his derivation Bennett used the ratio of the configuration integrals Z of two states of the canonical ensemble

$$\frac{Z_A}{Z_B} = \frac{\left\langle W(\mathbf{r}) e^{-\beta U_A(\mathbf{r})} \right\rangle_B}{\left\langle W(\mathbf{r}) e^{-\beta U_B(\mathbf{r})} \right\rangle_A} \tag{11}$$

where $W(\mathbf{r})$ is an arbitrary finite weighting function. The optimum weighting function obtained from the stationary condition for the dispersion of ΔF_{BA} with respect to a variation of $W(\mathbf{r})$ has the form

$$W(\mathbf{r}) = \frac{\text{const}}{(Z_A e^{-\beta U_B}/n_A + Z_B e^{-\beta U_A}/n_B)} \tag{12}$$

where n_A and n_B are the sizes of statistical sets for the two states. The ΔF_{BA} value for this type of weighting function can be found from the self-consistent numerical solution of equations

$$\beta \Delta F_{BA} = \ln \frac{\left\langle 1/\left[1 + e^{\beta\left(C - \Delta U(\mathbf{r})\right)}\right] \right\rangle_B}{\left\langle 1/\left[1 + e^{\beta\left(\Delta U(\mathbf{r}) - C\right)}\right] \right\rangle_A} + \beta C \tag{13}$$

$$\beta C = \ln\left(\frac{Z_A n_B}{Z_B n_A}\right)$$

The probability density distributions $\rho(\Delta U)$ for the states A and B in the energy coordinate U are interrelated [75]

$$\rho(\Delta U)_B e^{\beta \Delta U} = \rho(\Delta U)_A e^{\beta \Delta F} \tag{14}$$

According to Eqn. (14), the intersection point of $\rho(\Delta U)_A$ and $\rho(\Delta U)_B$ corresponds to the equality $\Delta U = \Delta F$ and to maximum value of the weighting function $W(\Delta U)$.

The potential and applications of the BAR method are well documented (see, *e.g.*, Refs 54, 61, 62, 66, 68, 76 – 83). For instance, the expression for statistical errors was derived using the maximum likehood method and it was shown that BAR has a smaller systematic error than the FEP method.[78] At large ΔU values, the free energy obtained from double-end FEP calculations is an asymptotic solution for BAR.[66] In this case, formal convergence of the self-consistency procedure in the BAR method is achieved only if the initial C value equals the ΔF value obtained from Eqn (8). The initial C value has little effect on the convergence at high overlap of occupied regions and can be taken equal to ΔU at the intersection point of $r(\Delta U)_A$ and $r(\Delta U)_B$ (see Ref. 61) or set equal to zero.[74] Expressions for free-energy calculations of multiple states by the BAR method were derived;[79] they can be used to analyze statistical sets or experimental data.

II.2. Thermodynamic integration method

The thermodynamic integration method proposed by J.Kirkwood in 1935 [84] is based on the expression of statistical integral as function of the coupling parameter λ

$$\Delta F_{BA} = F_B(\lambda = 1) - F_A(\lambda = 0) = \int_0^1 \left(\frac{\partial F(\lambda)}{\partial \lambda}\right) d\lambda \tag{15}$$

Having transformed expression (15) taking account of the equality

$$\frac{\partial F(\lambda)}{\partial \lambda} = -\frac{1}{\beta}\frac{\partial \ln Q(\lambda)}{\partial \lambda}, \tag{16}$$

one gets

$$\Delta F_{BA} = \int_0^1 \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \approx \sum_i \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda_i} \Delta\lambda_i \tag{17}$$

Integration over the coupling parameter λ corresponds to non-physical transition of the system (see curve *2* in Fig. 1). The free-energy difference $\Delta F(A \rightarrow B)$ is calculated not only by the TPT method, but also by the TI approach. In particular, TI calculations are carried out by the gradual growth [15, 85, 86] and finite differences [15, 43, 60] methods. In the former case, the coupling parameter λ continuously increases at each step of an MD trajectory and the derivative of the Hamiltonian with respect to λ is calculated analytically. In the latter case, one deals with a finite number of discrete λ values for which the dependence of $U(\mathbf{r})$ on λ is specified by expression (10). If $U(\mathbf{r},\lambda)$ values are calculated using linear combinations (9), there exists a direct analogy between the finite difference TI and FEP methods.[59]

II.3. Alchemical transformations and molecular topologies

Modelling of transitions between states by the TPT and TI methods involves non-physical *in silico* 'transmutations' of chemical compounds. As a consequence, such calculations are termed 'computational alchemy'.[17] Protocols for calculating alchemical mutations are designed using a kind of mathematical description called molecular topology. A topology is based on the chemical formula of the compound and contains complete information on the atoms, chemical bonds, bond angles and torsion angles. Alchemical mutations of molecules correspond to changes in their topological parameters. A FEP/MC study of alchemical mutation of ethane to methanol is one of the first examples of such calculations (Fig. 2).[42]

If mutating molecules are structurally similar, a common topology for the initial and final states is used. The force fields of intermediate states are specified using the coupling parameter λ. Parameters of the common topology can take both real values corresponding to one model molecule and fictitious or zero values for those structural fragments that are absent in the other molecule. If the ligand structures differ significantly, dual-topology schemes are used. They include separate description of the initial and final molecules and their superimposition in the common volume. Applications of both types of topologies have been considered in detail.[20, 51, 55, 64]

In applied calculations the free energies often include only changes in the nonbonding interactions of molecules, *i.e.*, the contributions of Coulomb and van der Waals potentials, which are determined by the atomic charges and Lennard-Jones parameters. In the course of alchemical mutations corresponding topological parameters can be varied separately; this allows one to evaluate the contributions of different nature to the total free energy.

Electrostatic interactions between molecules play a key role in most biochemical processes. Long-range electrostatic forces depend on the charge distribution and influence, in particular, the mutual arrangement of molecules in solution. A fundamental characteristic of electrostatic interactions is linear response of system polarization to perturbation of the electric field.[87 – 89] For instance, linear response underlies the description of dielectrics as continuous media.[88, 89] In microscopic models linear dependence occurs if polarization is taken into account explicitly in the case of induced dipoles and if the perturbing field is weak in the case of permanent dipoles.[88, 89] The free energy depends quadratically on the polarization coordinate of the system[44, 45, 53, 89] as, *e.g.*, in the Marcus theory of electron transfer.[87] For systems with Gaussian distribution functions the linear response relations are exact.[45, 53] The linear response principle is valid for polar solutions and, presumably, for proteins.[16, 44, 45, 53, 87 – 90]

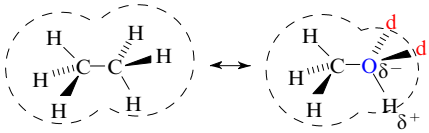


Figure 2. Alchemical mutation of ethane to methanol.[42] Corresponding changes in common topology involve the type of atoms and electron charges. The dashed curves outline the space occupied by the molecules in solution. One can see that for methanol, the fictitious atoms d are among solvent molecules.

The Lennard-Jones potentials in Eqn (4) define short-range atom – atom interactions in the condensed medium. The parameters (A_i,B_i) of nonpolar potentials influence the local packing of atoms and molecules, being usually unique for each type of atoms. Changes in these parameters represent an important type of alchemical mutations of atoms and molecules. The limiting cases of such transformations include annihilation of the atom X, *i.e.*, its conversion to a fictitious atom d (ideal gas particle) X→d, and the reverse process d→X (generation of a new atom from the fictitious atom). Fictitious atoms have zero van der Waals parameters and charges; therefore they are not involved in nonbonding interactions with the remaining part of the system. In the case of alchemical annihilation of particles the initial and final states of molecular ensembles have significantly different spatial configurations. For instance, position of a fictitious atom may overlap with positions of other atoms in the system (see Fig. 2). Conversion of a fictitious atom to a real atom causes those configurations where real particles occupy the space of the atom d to become energy-forbidden. Calculations by the TPT method with inclusion of this type of configurations can be accompanied by computational instabilities related to singularities in the potential energy of van der Waals interactions at nearly-zero interatomic distances. In this case, calculations using expression (6) give fifferent ΔG values for the forward and reverse alchemical annihilation.

Calculations of chemical potentials face the problem of overlapping solute and solvent molecules.[61, 75 – 77] It can be solved using so-called soft FF potentials that admit the overlap of atoms.[91, 92] The systematic error in ΔG calculations for the transformations X ⇄ d decreases when using the multistep schemes (9), (10), the BAR method being more preferable than FEP.[61, 68, 76, 77] In particular, a study on ligand incorporation into the cavity with high density of surrounding atoms taking endofullerene Ne@C_{60} as an example revealed a smaller systematic error for the BAR method.[68]

In some cases, the description of alchemical transformations of ligands shoud take account of changes in the molecular geometry that is specified by the valence bond lengths, bond angles and torsion angles.[55] This requires a correct inclusion of those spatial configurations which influence nonbonding interactions between atoms of the mutating chemical group, just like in the case of simulation of atomic annihilation. In particular, it was proposed to calculate the free energies corresponding to changes in the parameters of chemical bonds using an additional contribution from the perturbation of the reaction coordinate $\xi(\lambda_m \rightarrow \lambda_{m+1})$ in each step of summation of ΔG values in expression (10).[46, 50]

II.4. Free-energy thermodynamic cycles

The FEP and TI free-energy calculations make use of closed thermodynamic cycles.[39, 41, 51, 59, 67] In particular, the binding free energies of BAC to target proteins (ΔG_{bind}) are calculated by considering thermodynamic cycles that involve free ligands and protein – ligand complexes in solution (Fig. 3).[41]

The thermodynamic cycle shown in Fig. 3 describes transition of ligands from water to the complex with

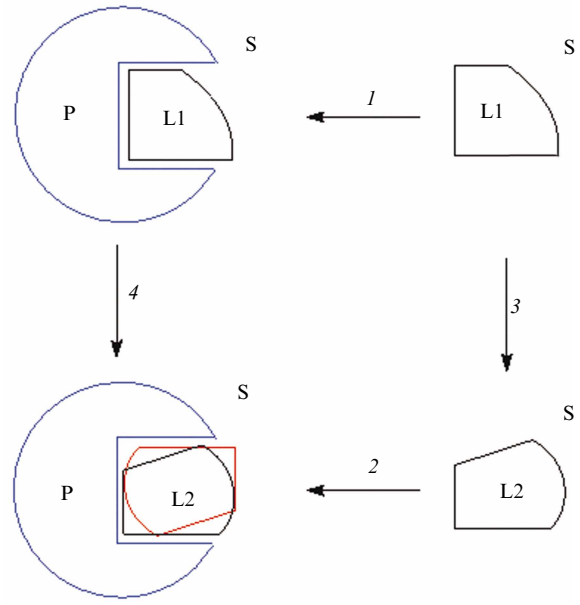


Figure 3. Thermodynamic cycle describing transition of ligands L1 and L2 from water S to complex with the protein P.[41, 51]
Shown are two possible orientations of the ligand L2 in the binding site.
For *1* – *4*, see text.

theprotein. Here, the free-energy changes along lines *1* (ΔG_1) and *2* (ΔG_2) correspond to the binding free energies of ligands L1 ($\Delta G^{\mathrm{L1}}_{\mathrm{bind}}$) and L2 ($\Delta G^{\mathrm{L2}}_{\mathrm{bind}}$). The free energies $\Delta G^{\mathrm{wat}}_{\mathrm{mut}}$ and $\Delta G^{\mathrm{prot}}_{\mathrm{mut}}$ in lines *3* and *4* correspond to the free energy of mutation L1→L2 in water and in the complex with protein, respectively. The total change in the free energy, ΔG_{sum}, for steps *1* – *4*, expressed through the configuration integrals of states Z and changes in the volume is zero.

$$\Delta G_{\mathrm{sum}} = -\frac{1}{\beta}\left(\ln\frac{Z_{\mathrm{PL1S}}}{Z_{\mathrm{L1S...PS}}} + \ln\frac{Z_{\mathrm{PL2S}}}{Z_{\mathrm{PL1S}}} + \ln\frac{Z_{\mathrm{L2S...PS}}}{Z_{\mathrm{PL2S}}} + \right. \quad (18)$$

$$\left. + \ln\frac{Z_{\mathrm{L1S...PS}}}{Z_{\mathrm{L2S...PS}}}\right) + P(\Delta V_1 + \Delta V_2 + \Delta V_3 + \Delta V_4) =$$

$$= -\frac{1}{\beta}\ln\frac{Z_{\mathrm{PL1S}}\, Z_{\mathrm{PL2S}}\, Z_{\mathrm{L2S...PS}}\, Z_{\mathrm{L1S...PS}}}{Z_{\mathrm{L1S...PS}}\, Z_{\mathrm{PL1S}} Z_{\mathrm{PL2S}} Z_{\mathrm{L2S...PS}}} = 0$$

where the subscripts PLS and LS...PS describe the complex of protein P with ligand L in water S and the molecules L and P at long separation in water.

Expression (18) confirms a known rule of zero total change in the thermodynamic potential along the closed cycles. Direct calculations of changes in the free energy in steps *1* and *2* are difficult-to-perform and difficult-to-simulate, whereas the free energies of alchemical mutations in the protein (step *3*) and in solution (step *4*) can be calculated by the TPT or TI methods with relative ease. Since the change in the thermodynamic potential along the overall cycle is zero, the difference between the binding free energies ($\Delta\Delta G^{21}_{\mathrm{bind}}$) of the compounds being compared is[41]

$$\Delta\Delta G^{21}_{\mathrm{bind}} = \Delta G^{\mathrm{L2}}_{\mathrm{bind}} - \Delta G^{\mathrm{L1}}_{\mathrm{bind}} = \Delta G^{\mathrm{prot}}_{\mathrm{mut}} - \Delta G^{\mathrm{wat}}_{\mathrm{mut}} \quad (19)$$

Expression (19) includes calculations only for the states of ligand L in the complex with the protein and in solution; the unbound protein is ignored. If one ligand is specified as an ideal gas molecule (no interaction with the environment), Eqn (19) makes it possible to calculate the binding free energy of the second ligand. This computational technique was named double annihilation of ligand.[93] In addition, if the experimental binding free energy $\Delta G^{\mathrm{L1}}_{\mathrm{bind}}$ of one ligand is know, formula (19) gives the binding free energy ΔG_{bind} of the second compound, $\Delta G^{\mathrm{L2}}_{\mathrm{bind}} = \Delta G^{\mathrm{L1}}_{\mathrm{bind}} + \Delta\Delta G^{21}_{\mathrm{bind}}$. Free-energy calculations using the thermodynamic cycle approach reduce possible systematic errors in MS, originating from approximate character of the force fields and the overall description of the model systems.

Calculations of free-energy differences require thorough analysis of possible issues and errors caused by incomplete examination of the allowed configurations of the system. Generally, the ligands being compared have different binding modes that can be separated by high kinetic barriers. In addition, the binding modes can be different in ligand position and orientation at the binding site, in the conformations of functional groups of the ligands and amino acid side groups at the binding site. Modelling of alchemical mutations using TPT methods demonstrates a trend towards retention of the conformation and position of the initial state and is in some cases insufficient to determine the optimum ligand arrangement. The drawback can be eliminated by inclusion of multiple binding modes in the mutation scheme. This is accompanied by a considerable increase in the computational cost and requires the use of additional methods, *e.g.*, automated docking (see below) for preselection of low-energy binding modes. When a number of ligands are compared, complex thermodynamic cycles combining multiple mutations are used. Examples of solution to this group of problems are provided by FEP/MD calculations of the free energies of protein – ligand complexes[57] and transition states of molecular transformations in the active site.[56] Similar approaches are employed in the calculations of free energies of solvation and partition coefficients between different solvents.[39, 42, 47 – 49, 51, 52]

II.5. Enhanced sampling of configurational space

A key problem in applied free-energy calculations is the need to generate large representative sets using finite-size phase trajectories. The use of the Boltzmann distribution imposes significant restriction on the accessible configurations of molecular ensembles; in particular, efficient modelling of transitions between spatial regions separated by high barriers is not easily performed. Besides, the characteristic times of conformational transitions of biomolecular systems are many orders of magnitude longer than time steps of MD simulations that are limited by the short times of chemical bond vibrations. The limited sampling also poses problems in the calculations of alchemical mutations of complex molecules containing mutating groups with significantly different chemical structures and(or) if the mutating groups are conformationally rich. As a consequence, the estimates of thermodynamic parameters of complex systems (*e.g.*, protein – ligand complexes) obtained using conventional MD and MC protocols can contain indeterminate errors. In this connection, recently considerable attention has been

paid to the development and application of mathematical methods for efficient extension of the accessible configurational space in the course of molecular modelling (see, *e.g.*, reviews [20, 21, 38, 94 – 96]).

In MD simulations, the sampling of phase space can be enhanced using mathematical techniques based on replica exchange of system configurations under different conditions. A considerable extension of the configuration space of the solvent in the course of modelling of alchemical mutations is achieved upon replica exchange at different values of the coupling parameter λ.[60] Besides, efficient replica exchange at different temperatures and(or) scaled interactions in the Hamiltonian of the system was reported;[97 – 99] a 2D version of RE for the parameter λ and Hamiltonian has also been reported.[65] Selection of possible configurations in the RE method is based on calculations of the Boltzmann factors responsible for the probability of emergence of the states being compared. This procedure is an example of incorporation of MC simulations into the MD computational scheme. Replica exchange enables a huge enhancement of the sampled space compared to conventional MD calculations. The RE method was integrated into the TPT and TI schemes and convergence of the results of free-energy calculations for complex biochemical systems was studied.[60, 65, 100 – 104]

In some cases, it is appropriate to directly consider rotational conformation, which can be calculated using additional software. Assuming that the degrees of freedom for internal rotation are not coupled with other rotational and vibrational states and that conformers are separated by high barriers, the partition function for n rotational conformers is given by [105]

$$Z_{\text{conf}} = \sum_{i=1}^{n} \mathrm{e}^{-\beta\varepsilon_i} \tag{20}$$

where ε_i is the energy of the ith conformer. Taking account of conformational lability of ligands can be of great importance in calculations of the free energy of binding ΔG_{bind}. Usually, ligands in complexes of proteins have fixed conformations which may be different from the equilibrium conformation of the ligand in the free state in solution. Changes in the free energy due to the loss of rotational degrees of freedom of the ligand, which occurs on going from solution to the complex with the protein, are governed by the logarithm of the ratio of the corresponding partition functions for rotational conformers. The results of modelling of protein – ligand complexes using known drugs showed that corrections to ΔG_{bind} due to the inclusion of Z_{conf} can be as large as a few kcal per mole.[106] TPT calculations of ΔG_{bind} with taking account of multiple ligand conformations were reported.[67, 107]

Umbrella sampling technique [108, 109] is widely used to change and control the accessible spatial regions in the course of simulations. The procedure involves multiplication of the probability distribution on the potential surface of the state A with energy $U(\mathbf{r})$ by a weighting function $w(\mathbf{r})$. The function $w(\mathbf{r})$ is chosen in such a manner that the probability for the system to arrive at a particular region (*e.g*, the region of high positive potential energy values) be increased. Ideally, the use of the modified potential gives a uniform set over all points of the configuration space on the new surface A′ with the potential energy $w(\mathbf{r})U(\mathbf{r})$. The desired average value of the function $\Phi(\mathbf{r})$ on the initial surface A is determined using the expression

$$\langle\Phi\rangle_{\text{A}} = \frac{\langle\Phi(\mathbf{r})/w(\mathbf{r})\rangle_{\text{A}'}}{\langle 1/w(\mathbf{r})\rangle_{\text{A}'}} \tag{21}$$

The modifying potentials are used in the TPT protocols, but the widest application field of umbrella sampling technique includes calculations of free-energy changes along the reaction coordinate (see below).

II.6. The potential of mean force

A key problem in modelling of biomolecular systems is to evaluate the free energy along the reaction coordinate. Unlike alchemical mutations, in this case, the Hamiltonian of the system has the same form for the states being compared. The motion of particles along the reaction coordinate $\xi(\mathbf{r}) = \xi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is illustrated by curve *3* in Fig. 1. The function $\xi(\mathbf{r})$ can, in particular, describe the interparticle distance or be a parameter of the molecular conformation.

The profile $\xi(\mathbf{r})$ gives the position of transition state and allows one to evaluate the free energy and kinetic parameters of the process (see Fig. 1). The free energy of the system $W(\xi)$ along the reaction coordinate $\xi(\mathbf{r})$ is given by [12, 16]

$$W(\xi) = W_0 - \frac{1}{\beta}\ln\frac{\int \delta[\xi - \xi(\mathbf{r})]\mathrm{e}^{-\beta U(\mathbf{r})}\mathrm{d}\mathbf{r}}{\int \mathrm{e}^{-\beta U(\mathbf{r})}\mathrm{d}\mathbf{r}} \tag{22}$$

Here, W_0 is a constant; the function under the logarithm is the distribution function $\rho(\xi)$ of the equilibrium system, which is equal to the average value of the δ-function $\delta[\xi - \xi(\mathbf{r})]$ over all coordinates $\mathbf{r}$ of the system for a specified ξ value. The function $W(\xi)$ is the potential of mean force [84] because its negative gradient defines the averaged force $\mathbf{f}(\xi)$ acting on the reacting particles. The profile $W(\xi)$ can be obtained from calculations of $\rho(\xi)$, $\mathbf{f}(\xi)$ or immediately pointwise from the free energy values determined by the TI or FEP methods (see, *e.g.*, Refs 12, 16, 28, 93, 110 – 125). On going from Cartesian to generalized coordinates one should use corresponding transformation matrices to describe ξ.[122] Owing to nonuniform population of different domains in the configuration space, $W(\xi)$ is evaluated using various techniques,[122, 124] in particular, umbrella samplings [108, 109] based on modifying potentials. In this case, a weak point of direct use of Eqn (21) is the impossibility to determine the optimum form of the weighting function prior to calculations. To eliminate the drawback, an adaptive umbrella sampling procedure was developed, including that for the multiple potentials whose shape is optimized in the course of successive iterations.[113, 119, 120]

The weighted histogram analysis method is widely used in calculations of free energies of biochemical processes.[114, 115] The procedure involves umbrella sampling for multiple overlaps of the perturbing parabolic potential along the reaction coordinate. Equilibrium-state trajectories are calculated separately at each point of the coordinate. The overall PMF profile is obtained by summing the probabilities of discrete $\xi(\mathbf{r})$ values over all trajectories and removing the contributions of restraining potentials. The expression for the weighting coefficients at particular contributions corresponds to the minimum statistical error of the total distribution function. The results of the weighted

histogram calculations are independent of the size of reacting particles as, *e.g.*, calculations of particle insertion by the Widom method.[121] However, unlike TPT methods, the weighted histogram analysis gives only relative free energies and position of the PMF base line should be determined separately. The weighted histogram analysis method was extended to PMF calculations along several reaction coordinates[116] and algorithms for joint use of this method and replica exchange from MD trajectories in large-scale parallel computing of ligand – protein binding.[125]

Metadynamics calculations of the system in the space of collective variables ξ are yet another widely use method for constructing free-energy surfaces.[94–96, 126–130] A collective variable is a function of atomic coordinates of the system.[95] It can describe geometric parameters of molecules, as well as the dipole moment, the coordination number, *etc*.[130] To overcome the potential barriers along similar slow degrees of freedom, the potential energy of the system is augmented with time-dependent modifying potentials $V(\xi,t)$, where t is time. In the course of the dynamics, the height of the potentials $V(\xi,t)$ increases with time in such a manner that the potential energy wells be compensated.[130] Calculations allow successive motion between local energy minima. As convergence is achieved, the free-energy surface becomes flat and then the motion of the system is described as random walk along the collective variable. In this case, the free energy

$$W(\xi) = C - V(\xi, t \to \infty) \tag{23}$$

is estimated up to the constant C. The advantage of the method consists in that no *a priori* information on the shape of the free-energy surface is needed. Metadynamics calculations also allow one to monitor and scan multiple reaction paths.[130]

II.7. Protein – ligand binding constants

Noncovalent ligand binding to a target protein in solution can be described by the equations

$$\mathrm{P} + \mathrm{L} \rightleftarrows \mathrm{PL} \tag{24}$$

$$K_\mathrm{d} = \frac{C_\mathrm{P}\, C_\mathrm{L}}{C_\mathrm{PL}} \tag{25}$$

where K_d is the dissociation constant having the dimensionality of concentration and C_PL, C_P, C_L are the concentrations of complex PL, free protein P and unbound ligand L, respectively. The environment is not included in explicit form.

Drugs bound to a protein molecule act as inhibitors. The efficiency of inhibitor L can be evaluated in terms of the concentration IC_{50} corresponding to 50% inhibition of activity P. For competitive binding of substrate and inhibitor molecules to a protein one has

$$\mathrm{IC}_{50} = K_\mathrm{d}\left(1 + \frac{C_\mathrm{S}}{K_\mathrm{M}}\right) \tag{26}$$

where C_S is the substrate concentration and K_M is the Michaelis constant.[131] If $C_\mathrm{S} \ll K_\mathrm{M}$, then $\mathrm{IC}_{50} \approx K_\mathrm{d}$. Then, in accordance with Eqns (24) and (25), the proportion, y, of bound protein is

$$y = \frac{C_\mathrm{PL}}{C_\mathrm{P} + C_\mathrm{PL}} = \frac{C_\mathrm{L}}{K_\mathrm{d}}\left(1 + \frac{C_\mathrm{L}}{K_\mathrm{d}}\right)^{-1} \tag{27}$$

According to formula (27), K_d equals the concentration C_L at which 50% of protein molecules are in the free state.[132] At $C_\mathrm{L} < K_\mathrm{d}$, most protein molecules are in the free state. The equilibrium constants of complexes can be determined in different manner from free-energy calculations using microscopic models.

One approach treats complexation as the appearance of the particle PL and disappearance of two molecules, P and L, in solution. The standard free energy of binding, $\Delta G^\circ_\mathrm{bind}$, is related to the thermodynamic dissociation constant K_d (Refs 133 and 134):

$$\Delta G^\circ_\mathrm{bind} = kT \ln \frac{K_\mathrm{d}}{C^\circ} \tag{28}$$

where C° is the concentration of the standard state (C° = 1 mol litre^{-1} or 1/1660 Å^{-3} per molecule[30, 133])

In practice, the $\Delta G^\circ_\mathrm{bind}$ values are determined from calculations of the relative free energies of states. In particular, $\Delta G^\circ_\mathrm{bind}$ values are calculated using the double annihilation of ligand L→d in the thermodynamic cycle (see above), which involves replacement of all atoms of the ligand by fictitious atoms in the complex and in solution and switching off nonbonding interactions between the ligand and its environment.[93, 133–143] From expression (19) it follows that $\Delta G^\circ_\mathrm{bind}$ equals the free-energy difference between these two processes.

A scheme was reported for calculations of ΔG for conversion of a ligand to an ideal gas particle fixed at specified position in the binding site and subsequent release of the particle to a volume corresponding to the standard state.[133, 137] Rigorous calculation of $\Delta G^\circ_\mathrm{bind}$ within the framework of the DA procedure involves MS of reversible switching off and switching on of interactions between the ligand and its environment. To attain full reversibility of results, the ligand is fixed at the binding site using restraining potentials $u(\mathbf{r})$ for (i) position of the centre of mass of the molecule, (ii) orientation relative to the axes of rotation of the entire molecule, (iii) conformational mobility, and then corrections for transition to standard conditions are applied.[133, 137, 141] The contribution of $u(\mathbf{r})$ to ΔG can also be evaluated by the FEP method by modelling a gradual decrease in the constant of the harmonic potential $u(\mathbf{r})$.[139] This approach was also used for calculating the activation entropy of enzymatic reactions.[144]

In the second approach, the complexation in solution is treated as a transition of protein (P) and ligand (L) molecules between the free and bound states. The association constant K_a ($K_\mathrm{a} = 1/K_\mathrm{d}$) of the protein and ligand is found from the ratio of the configuration integrals for the bound and free molecules. In this case the expression for K_a has the form of integral of the Boltzmann factor at the function $W(\xi)$ specified by Eqn (22).[145–147] For a spherically symmetrical case, the dependence of K_a on the distance r between P and L is given by

$$K_\mathrm{a} = 4\pi \int_0^{R_\mathrm{bound}} r^2\, \mathrm{e}^{-\beta W(\mathbf{r})} \mathrm{d}r \tag{29}$$

where integration along the coordinate r is carried out over the space of the bound state.[133, 146] The integral (29) takes

account of all positions and rotational orientations of L in a spherical layer around P at $r < R_{bound}$. The space of the bound state can be outlined using different criteria.[112, 133, 148] The major contribution to the integral (29) comes from the states with low values of $W(\xi)$. The binding free energy ΔG°_{bind} differs from the minimum value W_{min} for the complex PL; it can be roughly estimated as the product of W_{min} by the logarithm of the volume of the bound state.[147]

In principle, PMF calculations give a detailed picture of the molecular mechanism of ligand binding. However, a common problem in applied PMF calculations is that constructing a continuous spatial reaction coordinate for transition of ligand from the buried binding cavity to the free state in solution is difficult or even impossible. The advantage of the function $W(\xi)$ is that is it independent of the size of the protein and ligand molecules. In particular, PMF calculations are used to determine the absolute values of ΔG°_{bind} in protein – protein complexes.

Calculations of K_a and ΔG°_{bind} using the function $W(\xi)$ were reported for a wide range of systems.[112, 117, 143, 147, 149 – 156] Most examples illustrating PMF calculations carried out to determine ΔG°_{bind} values are reduced to spherically symmetrical [112, 117, 146] or linear problems (motion of atoms and molecules along the ion channel pore).[149, 150, 153, 154] Generally, ligand binding is a multistep process following a complex spatial trajectory, thus being difficult-to-describe using the 1D reaction coordinate. Note also that for many complexes the binding free energy can be obtained from both TPT and PMF calculations that reproduce experimental data with comparable errors and computational cost.[143]

Recently, the attention of researchers has been paid to MD calculations of the rate constants for dissociation (k_d) and association (k_a) of protein – ligand complexes and subsequent evaluation of K_d from the k_d/k_a ratio.[157, 158] In this case, expression

$$K_d = \frac{k_d}{k_a} \tag{30}$$

gives the third method for evaluating the equilibrium constants of the complex PL. Examples are provided by the results of studies on the trypsin – benzamidine test system (see below).

II.8. Molecular modelling of the end states

Generally, evaluation of the relative free energies of two states (ΔG_{BA}) by the FEP and TI methods requires computationally expensive modelling of both end and numerous intermediate states. Often, as, *e.g.*, in comparing the free energies of binding or solvation for large sets of molecules, it is of interest to obtain approximate ΔG_{BA} values at moderate computational cost. The problem can be solved by modelling only the end states of a system under study.

Approximate estimates of the free energy differences between two states, ΔG_{BA}, can be obtained using expression (5) from MS of the initial state only.[59, 159 – 161] Expansion of ΔG_{BA} in a series in terms of ΔU gives reasonably good results for electrostatic energies of systems with linear response. A clear example is provided by modelling the 'charging' of a monatomic ion. From FEP/MD calculations [44, 53] and Born formula [89] it follows that the free energy of a monatomic ion in water, ΔG_s, depends quadratically on the charge q as it changes from 0 to 1. In aqueous solution, the free energy ΔG_s of an ion equals nearly half the potential energy of the interaction between the ion and its environment, ΔU_s, which is due to energy expenditure for polarization of the solvent.[89] Numerical calculations of the free energy of charging of two ions carried out taking account of the contributions of a series expansion of ΔG up to the second-order terms exhibit good convergence,[45] whereas calculations for a similar system with inclusion of the third- to fifth-order terms required longer MD trajectories to achieve convergence.[160]

The difference between electrostatic free energies of end states can be evaluated using the linear response principle for system's polarization. In the Markus electron transfer theory the free energy functionals of electrostatic interactions have the shape of parabolas of equal curvature (Fig. 4), which was confirmed by the results of numerical simulations of the charging of ions.[44, 53] An analysis of the contributions of the free energies shown in Fig. 4 gives the following approximate expression for the dependence of ΔG_{BA} on ΔU [139, 162, 163]

$$\Delta G_{BA} = \frac{\langle \Delta U(\mathbf{r}) \rangle_A + \langle \Delta U(\mathbf{r}) \rangle_B}{2} \tag{31}$$

The free energies of protein – ligand complexes can be evaluated from simulations of only the end states using the linear interaction energy (LIE) method.[164] In this case, the binding free energy is calculated using the free energies corresponding to portions of the thermodynamic cycle including binding of neutral ligand and transition of the ligand in solution and in protein from the zero-charge state to the state with the desired charge distribution. The electrostatic contribution is calculated assuming that the binding free energy ΔG is a linear function of ΔU [see Eqn (31)]. The polar and nonpolar contributions to the free energy of binding were scaled separately:

$$\Delta G_{bind} = \alpha\left(\langle U^{vdw}_{L-S}\rangle_{PL} - \langle U^{vdw}_{L-S}\rangle_{LS}\right) + \beta\left(\langle U^{el}_{L-S}\rangle_{PL} - \langle U^{el}_{L-S}\rangle_{LS}\right) + \gamma \tag{32}$$

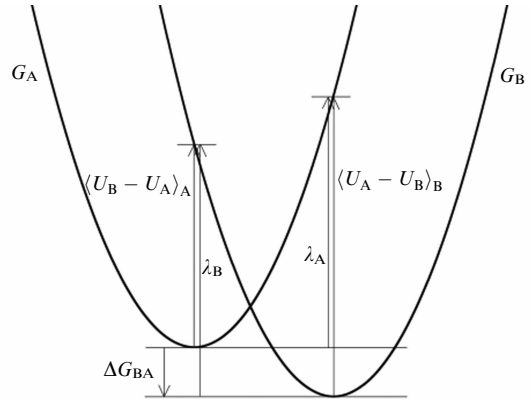


Figure 4. Electrostatic free energies of states A and B in the linear response approximation.[44, 139, 164]
The G_A and G_B parabolas have the same curvature. The corresponding reorganization energies, $\lambda_A = \langle U_B - U_A \rangle_A - \Delta G_{BA}$ and $\lambda_B = \langle U_A - U_B \rangle_B + \Delta G_{BA}$, are equal to each other.

The average energies of electrostatic and van der Waals interactions of ligand L with its environment S ($\langle U^{el}_{L-S}\rangle$ and $\langle U^{vdw}_{L-S}\rangle$, respectively) are determined from the MS of the protein – ligand complex PL and the free state in solution LS. The optimum value of the coefficient α for the nonpolar contribution was chosen empirically (0.14 – 0.18); the coefficient γ depends on the type of the system under study and is often equal to zero.[165] The theoretical value of the coefficient β at the polar contribution was obtained assuming linear response of system's polarization and is equal to 0.5 (*cf.* a range from 0.33 to 0.52 for optimal β values for different chemical groups).[165, 166]

In addition, polar contributions were calculated using a method based on linear response for the Langevin dipole model.[139, 167] Equation (32) can be augmented with internal energy difference between ligand conformations in solution and in the complex with biotarget.[168] The enthalpy and entropy contributions to ΔG_{bind} are not explicitly specified in the LIE method. In principle, the method imposes no restrictions on the structure of ligands; however, it is practically convenient to study homologous series of compounds.[165, 166] In addition to investigations of protein – ligand complexes, this approach was used to calculate the free energies of solvation of molecules.[165, 169]

The binding free energy ΔG_{bind} can also be calculated using the sum of the potential energy and entropy contributions of the end states [170, 171]

$$\Delta G_{bind} = \langle G_{protein-ligand}\rangle - \langle G_{protein}\rangle - \langle G_{ligand}\rangle \quad (33)$$

$$\langle G\rangle = \langle U_{MM}\rangle + \langle G_{s,el}\rangle + \langle G_{s,np}\rangle - T\langle S_{MM}\rangle$$

The average potential energy $\langle U_{MM}\rangle$ of protein, ligand and protein – ligand complex is obtained from MM calculations using the FF potentials. Electrostatic contribution to the free energy of solvation $\langle G_{s,el}\rangle$ is calculated within the framework of a continuum solvent model by the Poisson – Boltzmann method [172] or using generalized Born formula.[173] Nonpolar contribution to the free energy of solvation $\langle G_{s,np}\rangle$ is evaluated empirically from the linear dependence on the ligand surface area. Correspondingly, the methods described by expressions (33) are denoted as MM/PBSA and MM/GBSA. The averaged entropy of the ligand $\langle S_{MM}\rangle$ is determined by analyzing harmonic molecular vibrations. The average values appearing in Eqns (33) are calculated using molecular configurations obtained from separate computations of MD trajectories of the microscopic model of the system under study. One can either use separate ligand, protein and protein – ligand complex trajectories or compose the necessary configurations using only the trajectories of the complex. The latter method ignores structural differences between ligands in solution and in the complex with the protein; however, this approach is computationally more efficient and provides a better convergence. The authors of a critical publication [167] reported that the statistical error range of the MM/PBSA method for ΔG_{bind} calculations is a few kcal per mole larger than the error ranges of various versions of the LIE method.[167] Nevertheless, hundreds recent studies of protein – ligand complexes have been carried out using the MM/PBSA approach (see, *e.g.*, Refs 171, 174 – 176).

III. Construction and analysis of microscopic models

III.1. Microscopic models for biomolecules in solution and the boundary conditions

Until now our consideration of methods for free-energy calculations was based on the general laws of statistical thermodynamics regardless structural parameters of systems under study. Derivation of equations for thermodynamic potentials implies that the potential surfaces (energy levels) of a model system reasonably correlate with the properties of the target macroscopic system and that the model *per se* introduces minor distortions due to its finite size. However, this is not always valid. For instance, gas-phase calculations of protein molecules are insufficient for assessment of their properties in water (in aqueous solution). The reason is that the surface of solvated protein molecules bears a large number of ionized amino acid side groups shielded by water molecules. Correspondingly, strong electrostatic interactions between charged groups of isolated protein molecules can lead to large structural distortions in the course of MD calculations.

Yet another example is as follows. It was reported that isolated large molecular clusters are inappropriate to describe interactions between molecules in an infinitely large volume.[1, 13] Consider a cubic box with an edge of 100 Å (10 nm). Usually, this is sufficient to arrange a solvated protein molecule. Taking into account the fact that the action range of wall effect is about three layers of water molecules (*i.e.*, 10 Å), nearly half the total number of water molecules are affected even in this relatively large system. Transition from the description of finite systems to that of infinitely large systems is achieved by specifying additional boundary conditions. In practice, models with periodic and spherical boundary conditions are used.[1, 13, 14, 16, 19]

Models with periodic boundary conditions treat the system under consideration as an infinite 3D lattice built of identical boxes. When moving along the MD trajectories, molecules can freely cross boundaries between boxes. As a molecule escapes from a box, a symmetrically arranged molecule from another box simultaneously enters the box. Most calculations are carried out using boxes having the shape of a cube or rectangular parallelepiped. To reduce the number of surrounding water molecules, spherical proteins are also modelled using truncated octahedral boxes. A convenient model for prolate cylindrical entities (DNA, membrane proteins) is a hexagonal prism. Periodic boundary conditions allow one to simulate the properties of an infinitely large volume based on the treatment of a finite-size system. A key step in the simulation is the calculation of the potential energy $U(\mathbf{r})$ of the periodic lattice. This can be done using the minimum-image convention; in this case, each molecule in the box interacts only with the nearest images of other molecules from that box.[13, 19] Rigorous consideration of electrostatic interactions in the periodic lattice involves summation according to Ewald.[19]

The advantage of PBC models consists in that no restrictions are imposed on the arrangement of particles in the system. However, periodic boundary conditions have internal assumptions that should be taken into account when using this type of models.[1, 13, 14] In particular, the periodicity condition reproduces the structure of the primary box in an infinitely large volume and therefore

provides an approximate description of isotropic media. This class of models is inappropriate for simulating structural fluctuations with a characteristic size exceeding the box size. Also, models with periodic boundary generally cannot describe the native topology of biomolecules that usually form irregular structures in living cells. For instance, biomembranes are represented in PBC by alternating flat infinite layers of aqueous and nonpolar phases.

Many biomolecular systems can be described using models with spherical boundary conditions.[1,14,16] Such a model includes a protein being simulated (entire molecule or its fragment containing a reaction centre), small molecules and ions; the remaining space is filled with water molecules. The equations of motion are solved taking into account the interactions and motion of only the atoms confined within the sphere, all atoms lying outside the sphere being treated fixed. Considering models based on X-ray diffraction data, it is assumed that fixed position of protein periphery has little effect on the motion of atoms near the equilibrium positions corresponding to the experimental structure. Water molecules, ions and ligands can freely move in the system provided no additional restraining potentials or geometric conditions, the only limitation being no escape from the sphere. Water molecules within the boundary layer of the sphere obey the stochasticity condition simulating random molecular orientations in an infinitely large volume.[1,16] The solvent outside the sphere can be described using continuum models. A particular example of simulation of neat solvent using the SBC model is provided by the description of spherical volume as the surface of a four-dimensional hypersphere;[13] however, we left this model out of consideration.

The advantage of the SBC model is that, if the linear size of the protein–water system, is specified, the sphere of the corresponding diameter contains a minimum number of water molecules compared to other 3D bodies, *e.g.*, a cube as in the case of periodic lattice. Also, spherical symmetry is by default convenient for calculations of electrostatic interactions in the reaction centre. A drawback of this model consists in the need to use restraining potentials for atomic positions in the system. This restricts atomic motions in the course of MD simulations and thus makes it difficult, *e.g.*, to study ligand diffusion, to obtain a completely equilibrated protein structure and to examine large conformational rearrangements of the peptide backbone.

III.2. Molecular modelling of ligand binding in test systems

Wide use of computer chemistry in the studies of biomolecules dictated the need for the design of test systems for validation and comparison of computational methods. In particular, a test set including X-ray diffraction and physicochemical data for 195 complexes and 51 receptors,[177] a CCDC/Astex test set including 305 structures of protein–ligand complexes,[178] and a test set including structural data and binding constants for 200 complexes[179] were proposed to analyze the results of molecular modelling of protein–ligand complexes. There exists a Statistical Assessment of the Modelling of Proteins and Ligands (SAMPL) open Internet forum for discussing and validating computational predictions of the properties of small molecules and their complexes with proteins.[8,180,181] The test sets mentioned above were first of all composed for validation of VS methods. Since both MD and free-energy calculations of biochemical systems are much more computationally expensive and labour-consuming than VS methods, until now practical application of methods for free-energy calculations was limited to studies of the test sets of protein–ligand complexes including to 10^2 ligands.

III.2.a. Bacteriophage T4 lysozyme

Bacteriophage T4 lysozyme (from this point on, T4 lysozyme), a catalyst of the glycosidic bond hydrolysis, is a popular and well-studied test system used in the investigations of ligand binding by computer chemistry methods. Indeed, more than 200 references are retrieved from Web citation databases using search queries including the key words 'molecular simulation' and 'T4 lysozyme'. The polypeptide chain of T4 lysozyme contains 164 amino acids and folds into two domains. Mutations of T4 lysozyme allow one to engineer cavities in the core of the molecule.[182–185] For instance, the L99A mutant (Fig. 5) has a buried hydrophobic cavity of volume ~150 Å^3 (Ref. 182), which can

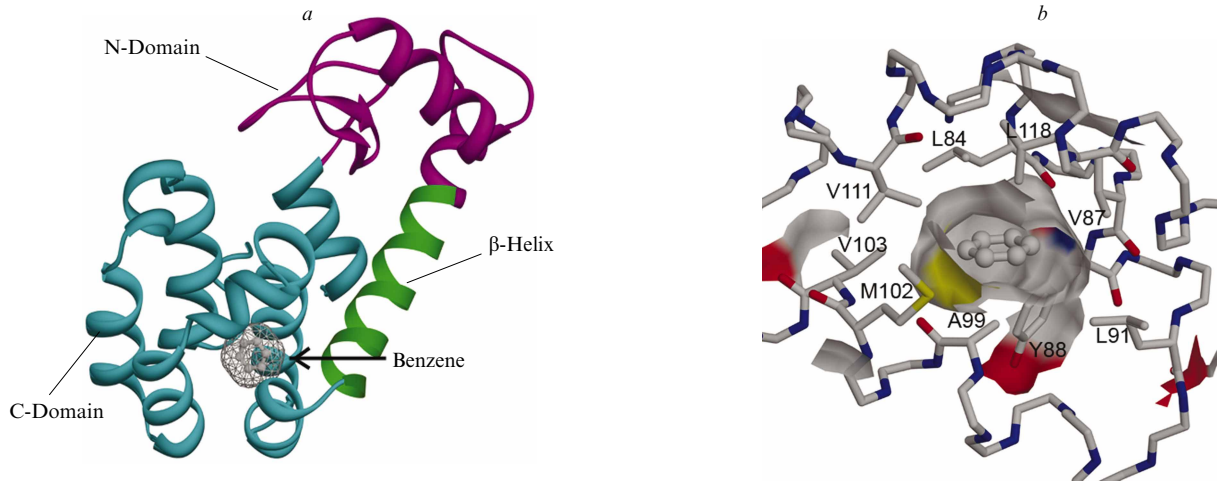


Figure 5. Complex of the L99A mutant of T4 lysozyme with benzene.[183]
(*a*) Common topology of the complex (PDB record 181L), (*b*) the surface of internal cavity and the arrangement of cavity-forming amino acid residues.

expand upon binding.[183] Benzene and its nonpolar derivatives bind to this cavity with $K_d \approx 0.1$ mmol litre^{-1}. The double mutant of T4 lysozyme, L99A/M102Q, serves a model for polar binding site. X-ray diffraction data with less than 2-Å resolution are available for complexes of T4 lysozyme and its mutants with different ligands. The structure of the cavity of T4 lysozyme mutants is stable even in long MD trajectories in spite of mutual motions of protein domains.[137]

Ligand binding to the buried cavity in T4 lysozyme mutants is modelled by the methods based on analysis of end states or on double alchemical annihilation. Table 1 presents the key results of free-energy calculations of this system. The first rigorous calculations of ΔG°_{bind} in the active site of the L99A mutant of T4 lysozyme were carried out taking a well-studied complex of the protein with benzene as an example.[137] The ΔG°_{bind} value was defined as the sum of the free energy ΔG°_1 of reversible alchemical annihilation of benzene molecule in the protein (obtained from TI calculations) and the free energy ΔG°_2 of transition of benzene molecule from the gas phase to aqueous solution. The contribution ΔG°_2 included analytically determined free energy of transition of benzene molecule from the bound to the standard state in the gas phase ΔG°_{21} and the experimentally determined standard free energy of hydration of benzene ΔG°_{22}. In the course of modelling the motion of the centre of mass and rotation of the benzene molecule in the active site were restrained by harmonic potentials. The ΔG°_{21} value calculated using the force constants equal to 100 kcal (mol Å^2)$^{-1}$ and 100 kcal (mol rad^2)$^{-1}$ was 11.06 kcal mol^{-1}; the corresponding ΔG°_1 values obtained for different conditions of calculations of the potential energy $U(\mathbf{r})$ were in the range from -15.37 to -16.50 kcal mol^{-1}. The final ΔG°_{bind} values were -4.01 to -5.14 kcal mol^{-1}, which is in good agreement with the experimental value equal to -5.19 kcal mol^{-1}.

Detailed calculations of the energy and entropy contributions (ΔU_{bind} and ΔS_{bind}, respectively) to ΔG°_{bind} were carried out for the complex of the L99A mutant of T4 lysozyme with benzene.[137, 187, 192] The average ΔU_{bind} value equal to -15.2 kcal mol^{-1} was obtained for rigid experimental protein structure from short-trajectory MD calculations; the corresponding entropy contribution for benzene transition from the standard state in the gas phase was 6.7 kcal mol^{-1}.[137] Estimation of the cratic entropy contribution gave a close value of 8.4 kcal mol^{-1}. The configurational entropy ΔS_{bind} was also directly calculated using alternative methods and the SBC model.[187] The $-T\Delta S_{bind}$ values determined from analysis of the root-mean-square deviations using Schlitter's formula and from quasi-harmonic analysis were in the range from 7.2 to 11.1 kcal mol^{-1} (Ref. 187). The binding entropy ΔS_{bind} calculated using thermodynamic cycles included the configurational entropy, as well as the polarization and hydrophobic contributions;[192] the total $-T\Delta S_{bind}$ value thus obtained was equal to 4.4 kcal mol^{-1}, being the closest to the experimental value of 2.8 kcal mol^{-1}.[185]

Calculations of the free energies of binding ΔG°_{bind} for complexes of mutants of T4 lysozyme with benzene derivatives are more complicated and serve good illustrations of the potential and bottlenecks of the available computational approaches (see Table 1). The free energies of binding in the complexes of the L99A mutant of T4 lysozyme with eleven nonpolar ligands were obtained from FEP calculations using the SBC model.[141] Higher rigidity of the protein core for the SBC model compared to the periodic lattice was mentioned. By and large, the SBC model correctly describes local fluctuations of atomic positions and the effect of restraining potentials on ΔG°_{bind} values. The absolute deviations of the calculated ΔG°_{bind} values from experimental ones were in the range from 0.16 (for benzofuran) to 4.39 kcal mol^{-1} (for *p*-xylene).

Table 1. Molecular simulation of ligand binding to the cavity of T4 lysozyme mutants.

Year	Model	Method	Results	Ref.
1995	L99A	Calorimetry	ΔG_{bind} of 91 ligands; calculations and evaluation of contributions to ΔG_{bind}	183
1997	L99A	MD, PBC, TI, DA	ΔG_{bind} of benzene taking into account standard conditions	137
2000	L99A	MD, PBC, TI, DA, MD H_2O in the active site	ΔG_{bind} of inert gases	186
2005	L99A	MD, SBC	ΔS_{bind} of benzene	187
2006	L99A/M102Q	MD, PBC, BAR, DA	ΔG_{bind} of phenol and catechol	188
2006	L99A	MD, SBC, FEP, DA	ΔG_{bind} of 11 ligands	141
2007	L99A	Calorimetry, docking, PBC, BAR, DA, multiple binding poses, rotation of amino acids	Testing and predictions of ΔG_{bind} for 18 ligands	189
2008	L99A/M102Q	Calorimetry, docking, MD, MM/GBSA	ΔG_{bind} of 33 ligands	190
2009	L99A/M102Q	Calorimetry, docking, MD, PBC, BAR, DA, multiple binding poses, rotation of amino acids	ΔG_{bind} of 13 ligands	191
2010	L99A	MD, SBC, FEP, DA, RE, rotation of amino acids	ΔG_{bind} of *p*-xylene	65
2010	L99A	MD, SBC	ΔS_{bind} of benzene	192
2011	L99A	Docking, MD, FEP, DA	ΔG_{bind} 38 ligands	193
2011	L99A	MD, DA, TI, RE, rotation of amino acids	ΔG_{bind} 12 ligands	102
2012	L99A	MD, FEP, DA, RE, rotation of amino acids	ΔG_{bind} of benzene and *p*-xylene	194
2014	L99A/M102Q	MD, density functional theory, MM/PBSA	ΔG_{bind} 8 ligands	195

The accuracy of calculations of ΔG°_{bind} values for complexes containing small ligands in the cavity of T4 lysozyme L99A/M102Q double mutant was improved by taking account of rotational conformers of the side group V111 (Fig. 5 *b*) in the active site and by analyzing multiple starting ligand orientations obtained by docking.[188] The need for forced expansion of the conformational space of amino acid side groups is due to the fact that conventional MD procedure usually deals with sluggish spontaneous rotation of these groups. Blind test of the results of calculations for five ligands gave correct, experimentally confirmed data on distinction between four bound molecules and one nonbinding molecule, arrangement of these molecules in the active site and their free energies of binding.[188] In the subsequent studies, simultaneous conformational changes in the side groups of a number of amino acid molecules in the binding site were taken into account using different versions of the RE method.[65, 102, 194] For instance, advanced accounting for the configuration space reduced the deviation of the calculated ΔG°_{bind} values from experimental data for *p*-xylene to -0.26 kcal mol^{-1}.[65, 141] Reliability of free-energy calculations was also improved using the AM1-CM2,[141] AM1-BCC,[188] and OPLS2.0 [194] quantum mechanical methods for calculating the atomic charge distributions in ligands.

In addition to conformational changes in the amino acid side groups, ligand binding to the active site of proteins can also influence the packing of the polypeptide chain of the protein. Recent X-ray diffraction studies of complexes containing benzene derivatives in the cavity of the L99A mutant of T4 lysozyme showed that minor changes in the ligand structure can cause large jumpwise conformational changes in the protein core.[196] Due to protein structure relaxation the experimentally measured changes in ΔG°_{bind} appeared to be smaller than those that could be predicted based on the changes in the physicochemical properties of ligands. It is expected that structural changes in the protein core in the complex with the ligand can be monitored in sufficiently long MD trajectories.[21, 196]

III.2.b. β-Trypsin

Yet another widely used test system for ligand binding studies is serine protease β-trypsin (Fig. 6). Molecular modelling of β-trypsin complexes with various ligands has been a subject of several tens of publications (for key results, see Table 2). The polypeptide chain of trypsin contains 230 amino acids. The binding site of inhibitor benzamidine having a total electron charge of +1 is located near the catalytic site (catalytic triad of trypsin includes the H57, D102 and S195 amino acid residues, see Fig. 6 *a*) and is accessible to the solvated ligand through an open channel.[197] Free trypsin is structurally similar to trypsin bound to benzamidine. Experimental K_d values for the trypsin complex with benzamidine vary between 7 and 15 μmol litre^{-1} depending on the measuring conditions.[181] Trypsin complexes with 500 ligands from the Maybridge fragment library were tested within the framework of research projects at the SAMPL open Internet forum.[180, 181]

The free energies of trypsin complexes with benzamidine and other ligands were calculated by the FEP,[200, 201, 207] TI,[202] LIE,[165, 203, 211] and MM/PBSA [206, 211] methods (Table 2); the results obtained are in reasonable agreement with experimental data. The entropy – enthalpy compensation effects were studied and it was shown that close ΔG°_{bind} values can correspond to strongly different enthalpy and entropy contributions.[209]

The open binding cavity of benzamidine (S4 domain in Fig. 6 *b*) makes it possible to clearly determine the binding path of the ligand in the active site. Earlier PMF calculations gave a qualitative description of dissociation of the trypsin – benzamidine complex.[156, 199] Subsequent extended studies disclosed intermediate complexes S1 – S3 that precede the bound state S4.[208] Recently, benzamidine binding and dissociation have been studied in detail and the corresponding rate constants, equilibrium constants and ΔG°_{bind} values were calculated.[208, 212 – 214] The rate constant for dissociation of a complex of 12 independent trajectories evaluated by the metadynamics method was 9.1 s^{-1} (*cf.* a value of 600 s^{-1} determined experimentally).[213] The rate constant for association calculated using a known K_d value [212] equals 1.18×10^7 litre mol^{-1} s^{-1} (*cf.* experimental

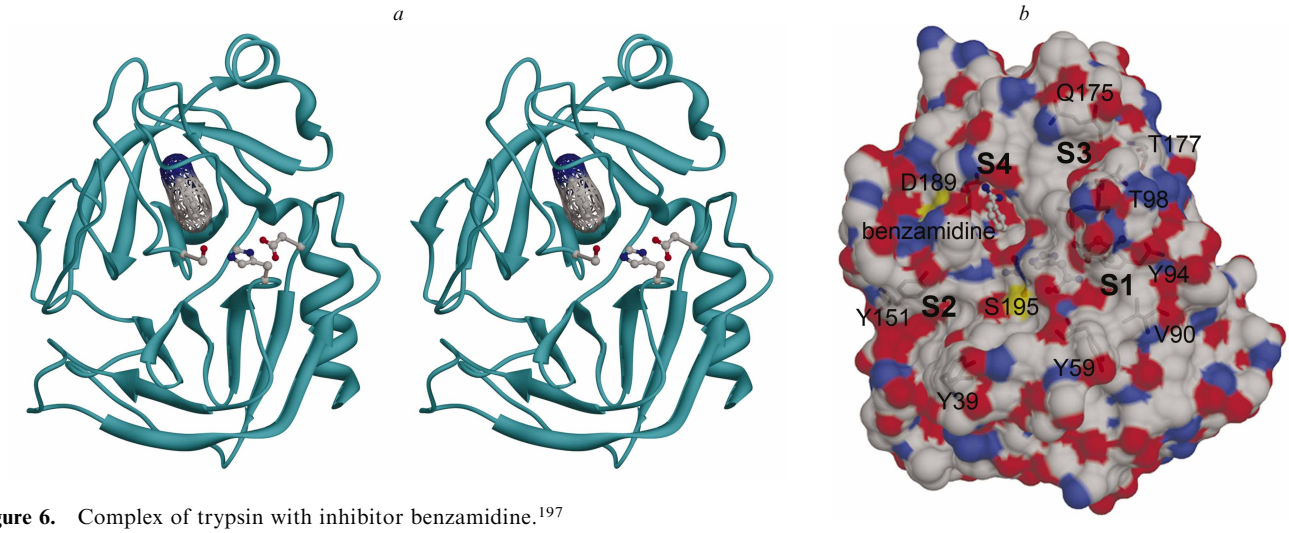


Figure 6. Complex of trypsin with inhibitor benzamidine.[197]
(*a*) Stereo image of β-trypsin – benzamidine complex (PDB record 3PTB); (*b*) the surface of trypsin – benzamidine complex (S1, S2, S3 are intermediate binding sites and S4 is the equilibrium binding site).

Table 2. Molecular simulations of ligand binding to β-trypsin.

Year	Method	Results	Ref.
1995	Grid-computed energies, docking, MD	G_{bind} of benzamidine	198
1997	Microscopic and continuum models for the medium, MD	PMF for benzamidine dissociation	199
1997	SBC, FEP, MC	$\Delta\Delta G_{bind}$ of benzamidine derivatives	200
1998	SBC, MD, LIE	ΔG_{bind} of benzamidine derivatives	165
1998	SBC, MD, FEP, graphical representation of ΔG of groups	ΔG_{bind} of benzamidine	201
1999	SBC, MD, TI, expanded set	$\Delta\Delta G_{bind}$ of benzamidine and benzylamine	202
2001	SBC, MD, LIE	ΔG_{bind} of inhibitors of different trypsins	203
2002	Docking, approximation of contributions to ΔG_{bind}	ΔG_{bind} of benzamidine derivatives	204
2005	Docking, metadynamics	ΔG_{bind} of benzamidine derivatives	205
2005	Docking, quantum mechanics, MM/PBSA	ΔG_{bind} of 47 benzamidine derivatives	206
2008	Polarizable FF, PBC, MD, DA	ΔG_{bind} of benzamidine and diazamidine	207
2009	MD, PBC, calculations of ΔG_{bind} from PMF	PMF of benzamidine dissociation	156
2011	MD, PBC, determination of metastable states S1 – S4 along the PMF	ΔG_{bind} and rate constants for benzamidine binding	208
2011	MD, PBC, calorimetry, binding to S3/S4 sites, entropy – enthalpy compensation	ΔG_{bind} of 8 benzamidine derivatives	209
2012	MD	2D PMF of benzamidine binding, collective coordinates	210
2012	MD, PBC, comparison of MM/PBSA and LIE	ΔG_{bind} of 34 benzamidine derivatives	211
2013	Metadynamics	ΔG_{bind} and PMF for benzamidine binding funnel	212
2015	Kinetics of binding and dissociation, metadynamics, Markovian processes	ΔG_{bind} of benzamidine	213
2015	Metastable states, Markovian processes, microsecond-scale MD	ΔG_{bind} and rate constants for benzamidine binding	214

value of 2.9×10^7 litre mol^{-1} s^{-1}). Yet another method for calculating the rate constants, namely, consideration of model Markovian states for long MD trajectories ($\sim$150 μs), makes it possible to study the kinetics of complexation in both directions.[214] The rate constants for dissociation and association thus obtained were 131×10^2 s^{-1} and 6.4×10^7 litre mol^{-1} s^{-1}, respectively. Modelling and detailed analysis of the behaviour of the system showed [214] that the binding kinetics is governed by the interaction between the ligand and various conformational states of protein molecules along the kinetically separated channels.

Summing up, studies on T4 lysozyme and β-trypsin carried out in the last two decades clearly illustrate the general trends in the development of methods for molecular modelling of protein – ligand complexes. The emphasis was placed on various techniques for calculations of ΔG°_{bind} and projections of the free energy along the reaction coordinates of the complexes. Methods for modelling the thermodynamics of ligand binding to the open and buried binding sites are available. Factors responsible for systematic errors in computations were established. For instance, taking account of conformational mobility of the protein and ligand is of great importance for the determination of exact values of free energies. Large-scale modelling of the dynamics of biomolecules on the time scale from 10^{-5} s to 10^{-3} s enables direct calculations of both association and dissociation constants of protein – ligand complexes.

IV. Search for new drugs and molecular modelling

Modern search for and structural design of new drugs are based on a wide range of experimental and computational methods. Often, computer chemists, including the author of this review, are asked 'Is there really a case where a drug that's on the market was designed by a computer?' [2] A short answer is yes. However, the question formulation is incorrect because the design of a new modern drug is a complex multistep process and requires the activity of many specialists. Practical implementation of a drug design project begins with the chemical synthesis and selection of compounds with respect to their biological activity followed by detailed experimental studies of candidate compounds. Based on the activity data, new compounds with higher expected activity and improved bioavailability, physicochemical and pharmacokinetic properties or ADME are synthesized. The efficiency of a drug is also characterized by the therapeutic index equal to the ratio of the average doses corresponding to 50% responses for toxicity and therapeutic effect.[131] Attainment of high therapeutic index is preferable, but this condition is not always met. Often, the chemical synthesis and activity tests are carried out in cyclic manner. The use of computational methods in these steps enables rational design and selection of compounds and gives considerable advantages to those project teams whose members are proficient with these tools. Computational methods are important part of structure-based drug design,[6 – 8] which is based on rational selection of chemical compounds taking account of their structural features and properties using detailed experimental data on the 3D atomic structure of the target protein.

A widely used method of search for new ligands acting on the biotarget is to carry out experimental screening of test sets of known chemical compounds. Screening reveals potential hits. Selection criteria for lead compounds include binding to a specified biotarget in micromole amounts and the possibility to improve the activity through chemical modification. Total activity tests of large sets of chemical compounds are most often unjustified and too expensive. However, efficient pre-selection of potentially active struc-

tures among a wide variety of compounds can be done using computational methods (see, *e.g.*, Refs 1 – 4, 215 – 222).

The choice of particular VS technique depends on the character of available information on the target and active ligands. One group of VS techniques takes account of the ligand properties only, that is, descriptors of the 2D and 3D structures and physicochemical properties. These methods are computationally inexpensive and enable high-throughput screening of virtual libraries containing millions of compounds. In particular, model pharmacophores and 2D templates of active molecules are widely used.[1, 216, 218, 220] Pre-selection of candidate compounds is also achieved by imposing restrictions on the physicochemical properties of the new compounds.[2, 218, 220] The next group of VS techniques includes structure-based methods using data on the 3D atomic structure of biotargets. An example of a structure-based approach is provided by automated docking into the active site; the procedure involves calculations of multiple admissible structures of protein – ligand complexes and their energy stabilities.[1 – 3, 189, 218 – 220, 223 – 225] This group also includes *de novo* design of chemical compounds given the structure of the biotarget active site.[2, 215, 216, 219, 221] When discussing the available methods, it is useful to pay attention to systematization of modern terminology in computational drug design.[226]

Virtual screening methods are widely used along with experimental methods in the search for new BAC. Large-scale VS techniques ignore conformational plasticity of biotargets, which imposes significant limitations on the quality of results obtained.[219] Owing to approximate character, VS can give either false-negative or false-positive results. Combining a number of VS techniques provides a better filter of candidate compounds for experimental testing. Successful application of VS methods in the design of new drugs is well documented (see, *e.g.*, Refs 2 – 4, 215 – 232).

Among computational approaches, comprehensive information on physicochemical properties, reactivity of compounds and binding to biotargets can be obtained by molecular simulations of microscopic systems. Owing to high computational cost it is appropriate to carry out free-energy calculations of microscopic models in the optimization of lead compounds. First of all, MS can serve to monitor how changes in the chemical structure of a molecule influence the binding affinity. Molecular modelling can also be used to evaluate the free energies of binding to specified antitargets and to assess the ADME properties. In the last decade, the number of such studies has increased in line with the development of computer hardware. The use of VS and molecular modelling in drug design [215, 216] is illustrated in Fig. 7.

An example of joint use of VS and molecular modelling (left side of Fig. 7) is provided by investigations of inhibitors of Dengue virus NS5 *O*-methyltransferase.[227, 228] The authors carried out step-by-step selection of structures from a virtual library containing two million of compounds. A pharmacophore and a 2D similarity search followed by docking into the binding site of *S*-adenosyl-L-methionine revealed a few dozens of potentially active compounds. Subsequent bioassay testing rejected false-positive results and demonstrated inhibitory effects ($IC_{50} \approx 60$ μmol litre^{-1}) for compounds containing a new structural motif, namely, urea derivatives of adamantane.[227] Different binding positions of adamantane in the binding site of the protein were studied using docking and LIE methods. A comparative study of binding parameters of antiviral adamantane-containing drugs amantadine and rimantadine and urea derivatives of adamantane was also reported.[228]

Design of non-nucleoside inhibitors of HIV reverse transcriptase[215, 229 – 231] is a good example of application of molecular simulations in the search for and optimization of

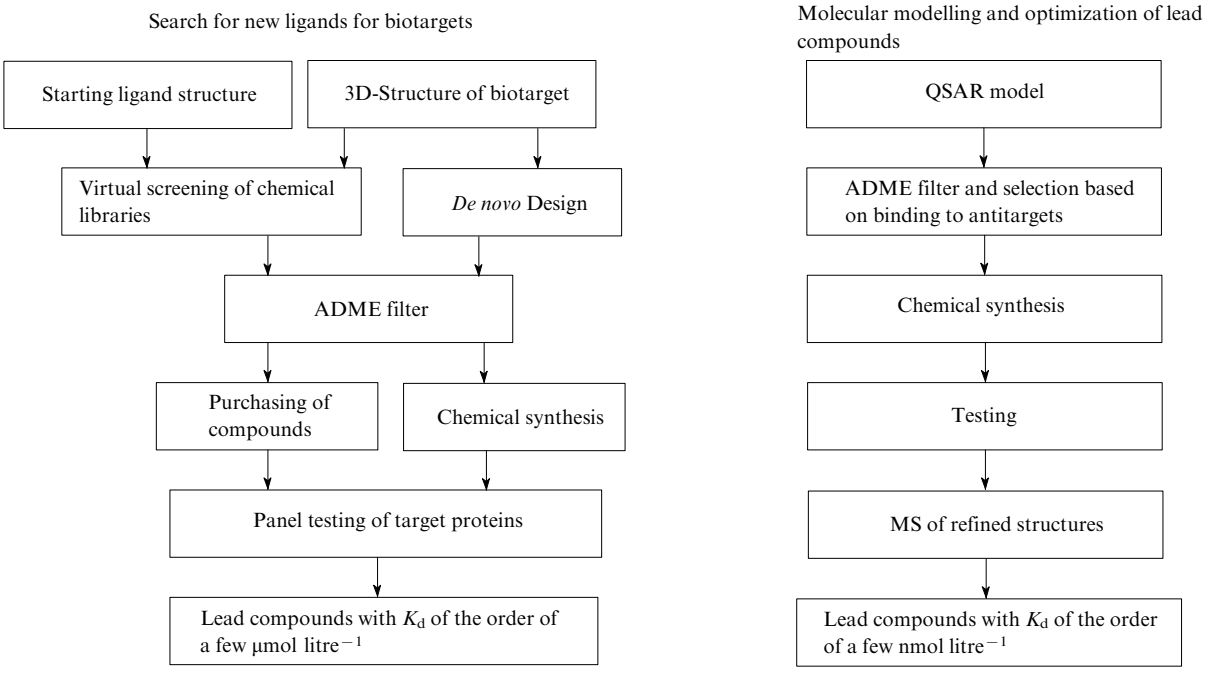


Figure 7. Pathways of computer-aided drug design.[215, 216]

lead compounds (right side of Fig. 7). Large-scale MS/FEP calculations of ΔG_{bind} for new series of compounds were used in combination with *de novo* design and bioactivity tests. Initial research efforts were aimed at optimizing two motifs selected from 100 lead compounds, namely, U – Het – NH – Ph and Het – NH – Ph – U (U is unsaturated hydrophobic group, Het is aromatic heterocycle and Ph is phenyl group). As a result, by replacing thiazole by pyrimidine as Het and appropriately choosing substituents, EC_{50} was reduced from 10 μmol litre^{-1} to 2 nmol litre^{-1} (Refs 2, 215, 229, 230). Further studies revealed the possibility to transfer the aromatic group to the distal region of the binding site of HIV reverse transcriptase outlined by the V106, E224, P225, P226, F227 and P236 amino acid residues. Molecular design, binding free-energy calculations and biotesting of new structures made it possible to synthesize oxadiazole and oxazole derivatives with high antiviral activity, therapeutic index and bioavailability.[215, 231]

V. Conclusion

Molecular modelling of microscopic systems is a powerful tool for studying the properties of biomacromolecules in condensed media. Methods of molecular modelling are widely used in basic research on the mechanisms of biochemical processes. Calculations of phase trajectories of atom-molecular ensembles provide detailed information on mechanistic aspects of chemical transformations and, which is equally important, underlie the possibility to obtain rigorous estimates of macroscopic properties including thermodynamic potentials, equilibrium constants and reaction rate constants. The results of modelling are widely used for independent interpretation of experimental data. Computer simulations of model systems underlie the understanding of molecular mechanisms of action of many biomolecules; in this respect, research into binding of biomolecules to small ligands is of great importance. Increased attention to free-energy calculations of biomolecular systems is associated with better theoretical description and capability enhancement of computational facilities.

Free-energy calculations of molecular ensembles by statistical mechanics methods is a well-developed field in modern theoretical chemistry and computer chemistry. Comprehensive inclusion of the configuration space of the system is of paramount importance for obtaining reliable data from molecular simulations. As applied to protein – ligand complexes, this requires extended analysis of different arrangement patterns of amino acid side groups and the polypeptide backbone, taking account of ligand conformations and possible arrangement of surrounding water molecules. Unlike quantum mechanical calculations of isolated molecules, free-energy calculations are far from being highly automated although a variety of programs are available. Construction of molecular models, preliminary analysis of the properties of a system under study, the choice of computational methods, carrying out MS and analysis of results obtained imply human activity. Computer simulations of biomolecular systems requires high-skilled staff. At present, there is reasonable understanding of methods for free-energy calculations in different situations.[38, 64, 73, 143] Modeling of alchemical mutations and determination of the potential of mean force along the reaction coordinate are two main methods for calculating the free energies of protein – ligand complexes. Further improvement of the accuracy and predictive power of molecular modelling methods is first of all associated with better description of the potential energy of microscopic systems using new force fields[33] and quantum mechanical techniques.[4, 233, 234]

Methods of molecular modelling are increasingly widely used to solve applied problems and become of great importance for drug design.[71, 157, 158, 215, 235] Molecular simulation of protein – ligand complexes gives a detailed description of their structure and, in principle, makes it possible to study the effect of binding on the functional dynamics of protein molecules and the influence of allosteric effects on the target activity.[236, 237] Systematic errors in calculations of the free energies of solvation and binding usually lie within 1 kcal mol^{-1} (and higher). Among known computational approaches, molecular simulations of microscopic systems give the most realistic estimates of thermodynamic parameters of interactions between ligands and targets and can therefore be used for rational optimization of lead compounds.

This review has been written with the financial support from the basic Research Programme of the Presidium of the Russian Academy of Sciences No. 22 (Investigation Line 2.3). The author expresses his gratitude to Doctor of Chemical Sciences A.F.Shestakov for valuable comments and to the staff of the Library at the Scientific Research Centre of the Russian Academy of Sciences in Chernogolovka for help in preparation of the manuscript.

References

1. A Leach *Molecular Modelling. Principles and Applications* (2nd Ed.) (London: Prentice Hall, 2001)
2. W L Jorgensen *Science* **303** 1813 (2004)
3. K M Merz, Jr, D Ringe, C H Reynolds *Drug Design. Structure- and Ligand-Based Approaches* (New York: Cambridge University Press, 2010)
4. B Honarparvar, T Govender, G E M Maguire, M E S Soliman, H G Kruger *Chem. Rev.* **114** 493 (2014)
5. C R Beddell, P J Goodford, F E Norrington, S Wilkinson, R Wootton *Br. J. Pharmacol.* **57** 201 (1976)
6. M Congreve, C W Murray, T L Blundell *Drug Discov. Today* **10** 895 (2005)
7. J Weigelt, L D B McBroom-Cerajewski, M Schapira, Y Zhao, C H Arrowmsmith *Curr. Opin. Chem. Biol.* **12** 32 (2008)
8. J Blaney *J. Comput.-Aid. Mol. Des.* **26** 13 (2012)
9. J P Overington, B Al-Lazikani, A L Hopkins *Nat. Rev. Drug Discov.* **5** 993 (2006)
10. D J Huggins, W Sherman, B Tidor *J. Med. Chem.* **55** 1424 (2012)
11. R Nussinov, C-J Tsai *Curr. Pharm. Des.,* **18** 1331 (2012)
12. J A McCammon, S C Harvey *Dynamics of Proteins and Nucleic Acids* (Cambridge: Cambridge University Press, 1987)
13. M P Allen, D J Tildesley *Computer Simulations of Liquids* (Oxford: Oxford University Press, 1987)
14. C Brooks III, M Karplus, B M Pettitt *Proteins: a Theoretical Perspective of Dynamics, Structure, and Thermodynamics. Advances in Chemical Physics* Vol. 71 (New York: Wiley, 1988)
15. D L Beveridge, F M DiCapua *Annu. Rev. Biophys. Biophys. Chem.* **18** 431 (1989)
16. A Warshel *Computer Modeling of Chemical Reactions in Enzymes and in Solutions* (New York: Wiley, 1991)
17. T P Straatsma, J A McCammon *Annu. Rev. Phys. Chem.* **43** 407 (1992)

18. D Frenkel, B Smit *Understanding Molecular Simulation. From Algorithms to Applications* (2nd Ed.) (San Diego: Academic Press, 2002)
19. T Schlick *Molecular Modeling and Simulation. An Interdisciplinary Guide* (New York: Springer, 2002)
20. Ch Chipot, A Pohorille *Free Energy Calculations. Theory and Applications in Chemistry and Biology* (Berlin: Springer, 2007)
21. R O Dror, R M Dirks, J P Grossman, H Xu, D E Shaw *Annu. Rev. Biophys.* **41** 429 (2012)
22. D E Makarov *Single Molecule Science: Physical Principles and Models* (Boca Raton, CA: CRC Press, 2015)
23. M Karplus *Angew. Chem. Int. Ed.* **53** 9992 (2014)
24. M Levitt *Angew. Chem. Int. Ed.* **53** 10006 (2014)
25. A Warshel *Angew. Chem. Int. Ed.* **53** 10020 (2014)
26. L D Landau, E M Lifshits *Teoreticheskaya Fizika. Statisticheskaya Fizika* (Theoretical Physics. Statistical Physics, Vol. 5) (2nd Ed.) (Moscow: Nauka, 1964)
27. Ya P Terletskii *Statisticheskaya Fzika* (Statistical Physics) (3rd Ed.) (Moscow: Vysshaya Shkola, 1994)
28. D A McQuarrie *Statistical Mechanics* (Sausalito: University Science Books, 2000)
29. I A Kvasnikov *Statisticheskaya Fizika. Teoriya Ravnovesnykh Sistem* (Statistical Physics. Theory of Equilibrium Systems) Vol. 2 (Moscow: URSS, 2002)
30. N M Bazhin, V A Ivanchenko, V N Parmon *Termodinamika Dlya Khimikov* (Thermodynamics for Chemists) (2nd Ed.) (Moscow: Khimiya, 2004)
31. V D Yagodovskii *Statisticheskaya Termodinamika v Fizicheskoi Khimii* (Statistical Thermodynamics in Physical Chemistry) (2nd Ed.) (Moscow: Binom. Laboratoriya Znanii, 2009)
32. A J Stone *Science* **321** 787 (2008)
33. O Demerdash, E-H Yap, T Head-Gordon *Annu. Rev. Phys. Chem.* **65** 149 (2014)
34. J C Owicki, H A Scheraga *J. Am. Chem. Soc.* **99** 7403 (1977)
35. G Czaku, A L Kaledin, J M Bowman *J. Chem. Phys.* **132** 164103 (2010)
36. D C Rapoport *The Art of Molecular Dynamics Simulation* (2nd Ed.) (Cambridge: Cambridge University Press, 2002)
37. R Elber *J. Chem. Phys.* **144** 060901 (2016)
38. J D Chodera, D L Mobley, M R Shirts, R W Dixon, K Branson, V S Pande *Curr. Opin. Struct. Biol.* **21** 150 (2011)
39. W L Jorgensen, L L Thomas *J. Chem. Theor. Comput.* **4** 869 (2008)
40. R W Zwanzig *J. Chem. Phys.* **22** 1420 (1954)
41. B L Tembe, J A McCammon *Comput. Chem.* **8** 281 (1984)
42. W L Jorgensen, C Ravimohan *J. Chem. Phys.* **83** 3050 (1985)
43. M Mezei *J. Chem. Phys.* **86** 7084 (1987)
44. G King, A Warshel *J. Chem. Phys.* **93** 8682 (1990)
45. R M Levy, M Belhadj, D B Kitchen *J. Chem. Phys.* **95** 3627 (1991)
46. D A Pearlman, P A Kollman *J. Chem. Phys.* **94** 4532 (1991)
47. B G Rao, U C Singh *J. Am. Chem. Soc.* **113** 4381 (1991)
48. J W Essex, C A Reynolds, W G Richards *J. Am. Chem. Soc.* **114** 3634 (1992)
49. V Luzhkov, A Warshel *J. Comput. Chem.* **13** 199 (1992)
50. C A Gough, D A Pearlman, P Kollman *J. Chem. Phys.* **99** 9103 (1993)
51. P Kollman *Chem. Rev.* **93** 2395 (1993)
52. F S Lee, Z T Chu, A Warshel *J. Comput. Chem.* **14** 161 (1993)
53. J Åqvist, T Hansson *J. Phys. Chem.* **100** 9512 (1996)
54. R J Radmer, P A Kollman *J. Comput. Chem.* **18** 902 (1997)
55. S Boresch, M Karplus *J. Phys. Chem. A* **103** 103 (1999)
56. V Luzhkov, J Åqvist *Chem. Phys. Lett.* **302** 267 (1999)
57. V Luzhkov, F Österberg, J Åqvist *FEBS Lett.* **554** 159 (2003)
58. C F Wong, J A McCammon *Adv. Protein. Chem.* **66** 87 (2003)
59. B O Brandsdal, F Österberg, M Almlöf, I Feierberg, V B Luzhkov, J Åqvist *Adv. Protein. Chem.* **66** 123 (2003)
60. C J Woods, J W Essex, M A King *J. Phys. Chem. B* **107** 13703 (2003)
61. N Lu, J K Singh, D A Kofke *J. Chem. Phys.* **118** 2977 (2003)
62. N Lu, T B Woolf *Mol. Phys.* **102** 173 (2004)
63. N Lu, D A Kofke, T B Woolf *J. Comput. Chem.* **25** 28 (2004)
64. J Michel, J W Essex *J. Comput.-Aid. Mol. Des.* **24** 639 (2010)
65. W Jiang, B Roux *J. Chem. Theory Comput.* **6** 2559 (2010)
66. V B Luzhkov *J. Chem. Phys.* **132** 194104 (2010)
67. D L Mobley, P V Klimovich *J. Chem. Phys.* **137** 230901 (2012)
68. V B Luzhkov *Mol. Simul.* **38** 326 604 (2012)
69. N E Dubinin, N A Vatolin, V V Filippov *Russ. Chem. Rev.* **83** 987 (2014)
70. L Boukharta, H Gutiérrez-de-Terán, J Åqvist *PLOS Comput. Biol.* **57** e1003585 (2014)
71. L Wang, Y Wu, Y Deng, B Kim, L Pierce, G Krilov, D Lupyan, S Robinson, M K Dahlgren, J Greenwood, D L Romero, C Masse, J L Knight, T Steinbrecher, T Beuming, W Damm, E Harder, W Sherman, M Brewer, R Wester, M Murcko, L Frye, R Farid, T Lin, D L Mobley, W L Jorgensen, B J Berne, R A Friesner, R Abel *J. Am. Chem. Soc.* **137** 2695 (2015)
72. J P Valleau, D N Card *J. Chem. Phys.* **57** 5457 (1972)
73. N Hansen, W F van Gunsteren *J. Comput. Theory Comput.* **10** 2632 (2014)
74. C H Bennett *J. Comput. Phys.* **22** 245 (1976)
75. K S Shing, K E Gubbins *Mol. Phys.* **46**1109 (1982)
76. G L Deitrick, L E Scriven, H T Davis *J. Chem. Phys.* **90** 2370 (1989)
77. D A Kofke, P T Cummings *Mol. Phys.* **92** 973 (1997)
78. M R Shirts, E Bair, G Hooker, V S Pande *Phys. Rev. Lett.* **91** 140601 (2003)
79. M R Shirts, J D Chodera *J. Chem. Phys.* **129** 124105 (2008)
80. G König, S Bruckner, S Boresch *J. Comput. Chem.* **30** 1712 (2009)
81. M Goette, H Grubmüller *J. Comput. Chem.* **30** 447 (2009)
82. A M Hahn, H Then *Phys. Rev. E* **80** 031111 (2009)
83. G König, S Boresch *J. Comput. Chem.* **32** 1082 (2010)
84. J G Kirkwood *J. Chem. Phys.* **3** 300 (1935)
85. H J C Berendsen, J P M Postma, W F van Gunsteren, in *Molecular Dynamics and Protein Structure* (Western Springs: Polycrystal Book Service, 1985)
86. T P Straatsma, H J C Berendsen *J. Chem. Phys.* **89** 5876 (1988)
87. R A Marcus *Annu. Rev. Phys. Chem.* **15** 155 (1964)
88. I E Tamm *Osnovy Teorii Elektrichestva* (Foundations of Theory of Electricity) (Moscow: Nauka, 1966)
89. A Warshel, S T Russel *Quart. Rev. Biophys.* **17** 283 (1984)
90. T Simonson *Proc. Natl. Acad. Sci., USA* **99** 6544 (2002)
91. M Zacharias, T P Straatsma, J A McCammon *J. Chem. Phys.* **100** 9025 (1994)
92. J W Pitera, W F van Gunsteren *Mol. Simul.* **28** 45 (2002)
93. W L Jorgensen, J K Buckner, S Boudon, J Tirado-Rives *J. Chem. Phys.* **89** 3742 (1988)
94. X Biarnés, S Bongarzone, A V Vargiu, P Carloni, P Ruggerone *J. Comput.-Aid. Mol. Des.* **25** 395 (2011)
95. C Abrams, G Bussi *Entropy* **16** 163 (2014)
96. R C Bernardi, M C R Melo, K Schulten *Biochim. Biophys. Acta* **1850** 872 (2015)
97. R H Swendsen, J-S Wang *Phys. Rev. Lett.* **57** 2607 (1986)
98. Y Sugita, A Kitao, Y Okamoto *J. Chem. Phys.* **113** 6042 (2000)
99. H Fukunishi, O Watanabe, S Takada *J. Chem. Phys.* **116** 9058 (2002)
100. J Hritz, C Oostenbrink *J. Chem. Phys.* **128** 144121 (2008)
101. M Fajer, D Hamelberg, J A McCammon *J. Chem. Theory Comput.*, **4** 1565 (2008)
102. I V Khavrutskii, A Wallqvist *J. Chem. Theory Comput.* **7** 3001 (2011)

103. L Wang, R A Friesner, B J Berne *J. Phys. Chem. B* **115** 9431 (2011)
104. K Wang, J D Chodera, Y Yang, M R Shirts *J. Comput.-Aid. Mol. Des.* **27** 989 (2013)
105. P J Flory *Statistical Mechanics of Chain Molecules* (New York: Interscience, 1969)
106. J Tirado-Rives, W L Jorgensen *J. Med. Chem.* **49** 5880 (2006)
107. A A Zeifman, V V Stroylov, F N Novikov, O V Stroganov, V Kulkov, G G Chilov *J. Chem. Theory Comput.* **9** 1093 (2013)
108. G M Torrie, J P Valleau *Chem. Phys. Lett.* **28** 578 (1974)
109. G M Torrie, J P Valleau *J. Comput. Phys.* **23** 187 (1977)
110. C Pangali, M Rao, B J Berne *J. Chem. Phys.* **71** 2975 (1979)
111. S H Northrup, M R Pear, C-Y Lee, J A McCammon, M Karplus *Proc. Natl. Acad. Sci.* **79** 4035 (1982)
112. W L Jorgensen *J. Am. Chem. Soc.* **111** 3770 (1989)
113. R W W Hooft, B P van Eijck, J Kroon *J. Chem. Phys.* **97** 6690 (1992)
114. S Kumar, D Bouzida, R H Swendsen, P A Kollman, J M Rosenberg *J. Comput. Chem.* **13** 1011 (1992)
115. S Kumar, J M Rosenberg, D Bouzida, R H Swendsen, P A Kollman *J. Comput. Chem.* **16** 1339 (1995)
116. B Roux *Comput. Phys. Commun.* **91** 275 (1995)
117. C Chipot, B Maigret, D A Pearlman, P A Kollman *J. Am. Chem. Soc.* **118** 2998 (1996)
118. S Izrailev, S Stepaniants, M Balsera, Y Oono, K Schulten *Biophys. J.* **72** 1568 (1997)
119. C Bartels, M Karplus *J. Comput. Chem.* **18** 1450 (1997)
120. C Bartels, M Schaefer, M Karplus *J. Chem. Phys.* **111** 8048 (1999)
121. C Czaplewski, S Kalinowski, A Liwo, H A Scheraga *Mol. Phys.* **103** 3153 (2005)
122. D Trzesniak, A-P E Kunz, W F van Gunsteren *Chem. Phys. Chem.* **8** 162 (2007)
123. V B Luzhkov *Chem. Phys. Lett.* **452** 72 (2008)
124. W Kästner *Wiley Interdiscipl. Rev.* **1** 932 (2011)
125. W Jiang, Y Luo, L Maragliano, B Roux *J. Chem. Theory Comput.* **8** 4672 (2012)
126. A Laio, M Parrinello *Proc. Natl. Acad. Sci.* **99** 12562 (2002)
127. G Bussi, A Laio, M Parrinello *Phys. Rev. Lett.* **96** 090601 (2006)
128. V Babin, C Roland *J. Chem. Phys.* **125** 204909 (2006)
129. A Barducci, G Bussi, M Parrinello *Phys. Rev. Lett.* **100** 020603 (2008)
130. A Barducci, M Bonomi, M Parrinello *Wiley Interdiscipl. Rev.* **1** 826 (2011)
131. D Voet, J G Voet *Biochemistry* (3rd Ed.) (New York: Wiley, 2004)
132. T E Creighton *Proteins. Structures and Molecular Properties* (2nd Ed.) (New York: W H Freeman & Co, 1993)
133. M K Gilson, J A Given, B L Bush, J A McCammon *Biophys. J.* **72** 1047 (1997)
134. Y Deng, B Roux *J. Phys. Chem. B* **113** 2234 (2009)
135. S Miyamoto, P A Kollman *Proteins* **16** 226 (1993)
136. T Z Mordasini Denti, W F van Gunsteren, F Diederich *J. Am. Chem. Soc.* **118** 6044 (1996)
137. J Hermans, L Wang *J. Am. Chem. Soc.* **119** 2707 (1997)
138. S B Singh, P A Kollman *J. Am. Chem. Soc.* **121** 3267 (1999)
139. Y Y Sham, Z T Chu, H Tao, A Warshel *Proteins* **39** 393 (2000)
140. S Boresch, F Tettinger, M Leitgeb, M Karplus *J. Phys. Chem. B* **107** 9535 (2003)
141. Y Deng, B Roux *J. Chem. Theory Comput.* **2** 1255 (2006)
142. I J General, R Dragomirova, H Meirovitch *J. Chem. Theory Comput.* **7** 4196 (2011)
143. J C Gumbart, B Roux, C Chipot *J. Chem. Theory Comput.* **9** 794 (2013)
144. M Strajbl, Y Y Sham, J Villà, Z-T Chu, A Warshel *J. Phys. Chem. B* **104** 4578 (2000)
145. J E Prue *J. Chem. Ed.* **46** 12 (1969)
146. D Shoup, A Szabo *Biophys. J.* **40** 33 (1982)
147. H-J Woo, B Roux *Proc. Natl. Acad. Sci.* **102** 6825 (2005)
148. M Mihailescu, M Gilson *Biophys. J.* **87** 23 (2004)
149. S Crouzy, S Bernèche, B Roux *J. Gen. Physiol.* **118** 207 (2001)
150. T W Allen, O S Anderson, B Roux *Biophys. Chem.* **124** 251 (2006)
151. M S Lee, M A Olson *Biophys. J.* **90** 864 (2006)
152. L Charlier, C Nespoulous, S Fiorucci, S Antonczak, J Golebiowski *Phys. Chem. Chem. Phys.* **9** 5761 (2007)
153. T Bastuğ, S Kuyucak *J. Chem. Phys.* **126** 105103 (2007)
154. T Bastuğ, P-O Chen, S M Patra, S Kuyucak *J. Chem. Phys.* **128** 155104 (2008)
155. M S Lee, M A Olson *J. Phys. Chem. B* **112** 13411 (2008)
156. S Doudou, N A Burton, R H Henchman *J. Chem. Theory Comput.* **5** 909 (2009)
157. A C Pan, D W Borhani, R O Dror, D E Shaw *Drug Discov. Today* **18** 667 (2013)
158. M DeVivo, M Masetti, G Bottegoni, A Cavalli *J. Med. Chem.* **59** 4035 (2016)
159. B Jayaram, D L Beveridge *J. Phys. Chem.* **94** 7288 (1990)
160. P E Smith, W F van Gunsteren *J. Chem. Phys.* **100** 577 (1994)
161. H Liu, A E Mark, W F van Gunsteren *J. Phys. Chem.* **100** 9485 (1996)
162. R Langen, G D Brayer, A M Berghuis, G McLendon, F Sherman, A Warshel *J. Mol. Biol.* **224** 589 (1992)
163. F S Lee, Z-T Chu, M B Bolger, A Warshel *Prot. Eng.* **5** 215 (1992)
164. J Åqvist, C Medina, J E Samuelsson *Protein. Eng.* **7** 385 (1994)
165. T Hansson, J Marelius, J Åqvist *J. Comput.-Aid. Mol. Des.* **12** 27 (1998)
166. M Almlöf, J Carlsson, J Åqvist *J. Chem. Theory Comput.* **3** 2162 (2007)
167. N Singh, A Warshel *Proteins* **78** 1705 (2010)
168. E Stjernschantz, J Marelius, C Medina, M Jacobsson, N P E Vermeulen, C Oostenbrink *J. Chem. Inf. Model.* **46** 1972 (2006)
169. H A Carlson, W L Jorgensen *J. Phys. Chem.* **99** 10667 (1995)
170. J Srinivasan, T E Cheatham III, P Cieplak, P A Kollman, D A Case *J. Am. Chem. Soc.* **120** 9401 (1998)
171. P A Kollman, I Massova, C Reyes, B Kuhn, S Huo, L Chong, M Lee, T Lee, Y Duan, W Wang, O Donini, P Cieplak, J Srinivasan, D A Case, T E Cheatham III *Acc. Chem. Res.* **33** 889 (2000)
172. K A Sharp, B Honig *Annu. Rev. Biophys. Biophys. Chem.* **19** 301 (1990)
173. W C Still, A Tempczyk, R C Hawley, T Hendrickson *J. Am. Chem. Soc.* **112** 6127 (1990)
174. H Gouda, I D Kuntz, D A Case, P A Kollman *Biopolymers* **68** 16 (2003)
175. T Hou, J Wang, Y Li, W Wang *J. Chem. Inf. Model.* **51** 69 (2011)
176. S Genheden, U Ryde *Exp. Opin. Drug Discov.* **10** 449 (2015)
177. O Roche, R Kiyama, C L Brooks III *J. Med. Chem.* **44** 3592 (2001)
178. J W M Nissink, C Murray, M Hartshorn, M L Verdonk, J C Cole, R Taylor *Proteins* **49** 457 (2002)
179. E Perola, W P Walters, P S Charifson *Proteins* **56** 235 (2004)
180. J Newman, V J Fazio, T T Caradoc-Davies, K Branson, T S Peat *J. Biomol. Screen,* **14** 1245 (2009)
181. J Newman, O Dolezal, V Fazio, T Caradoc-Davies, T S Peat *J. Comput.-Aid. Mol. Des.* **26** 497 (2012)
182. A E Eriksson, W A Baase, J A Wozniak, B W Matthews *Nature (London)* **355** 371 (1992)
183. A Morton, B W Matthews *Biochemistry* **34** 8576 (1995)
184. B Q Q Wei, W A Baase, L H Weaver, B W Matthews, B K Shoichet *J. Mol. Biol.* **322** 339 (2002)

185. L Liu, W A Baase, B W Matthews *J. Mol. Biol.* **385** 595 (2009)
186. G Mann, J Hermans *J. Mol. Biol.* **302** 302 (2000)
187. J Carlsson, J Åqvist *J. Phys. Chem. B* **109** 6448 (2005)
188. D L Mobley, J D Chodera, K A Dill *J. Chem. Phys.* **125** 084902 (2006)
189. D L Mobley, A P Graves, J D Chodera, A C McReynolds, B K Shoichet, K A Dill *J. Mol. Biol.* **371** 1118 (2007)
190. A P Graves, D M Shivakumar, S E Boyce, M P Jacobson, D A Case, B K Shoichet *J. Mol. Biol.* **377** 914 (2008)
191. S E Boyce, D L Mobley, G J Rocklin, A P Graves, K A Dill, B K Shoichet *J. Mol. Biol.* **394** 747 (2009)
192. N Singh, A Warshel *Proteins* **78** 1724 (2010)
193. R D Malmstrom, S J Watowich *J. Chem. Inf. Model.* **51** 1648 (2011)
194. L Wang, B J Berne, R A Friesner *Proc. Natl. Acad. Sci.* **109** 1937 (2012)
195. S J Fox, J Dziedzic, T Fox, C S Tautermann, C-K Skylaris *Proteins* **82** 3335 (2014)
196. M Merski, M Fischer, T E Balius, O Eidam, B K Shoichet *Proc. Natl. Acad. Sci.* **112** 5039 (2015)
197. M Marquart, J Walter, J Deisenhofer, W Bode, R Huber *Acta Crystallogr., Sect. B* **39** 480 (1983)
198. B A Luty, Z R Wasserman, P F W Stouten, C N Hodge, M Zacharias, J A McCammon *J. Comput. Chem.* **16** 454 (1995)
199. H Resat, T J Marrone, J A McCammon *Biophys. J.* **72** 522 (1997)
200. J W Essex, D L Severance, J Tirado-Rives, W L Jorgensen *J. Phys. Chem. B* **101** 9663 (1997)
201. R J Radmer, P A Kollman *J. Comput.-Aid. Mol. Des.* **12** 215 (1998)
202. N Ota, C Stroupe, J M S Ferreira-da-Silva, S A Shah, M Mares-Guia, A T Bruger *Proteins* **37** 641 (1999)
203. B O Brandsdal, A O Smalås, J Åqvist *FEBS Lett.* **499** 171 (2001)
204. S M Schwarzl, T B Tschopp, J C Smith, S Fischer *J. Comput. Chem.* **23** 1143 (2002)
205. F L Gervasio, A Laio, M Parrinello *J. Am. Chem. Soc.* **127** 2600 (2005)
206. F Gäter, S M Schwarzl, A Dejaegere, S Fischer, J C Smith *J. Phys. Chem. B* **109** 10474 (2005)
207. D Jiao, P A Golubkov, T A Darden, P Ren *Proc. Natl. Acad. Sci.* **105** 6290 (2008)
208. I Buch, T Giorgino, G De Fabritiis *Proc. Natl. Acad. Sci.* **108** 10184 (2011)
209. T Brandt, N Holzmann, L Muley, M Khayat, C Wegscheid-Gerlach, B Baum, A Heine, D Hangauer, G Klebe *J. Mol. Biol.* **405** 1170 (2011)
210. G A Tribello, M Ceriotti, M Parrinello *Proc. Natl. Acad. Sci.* **109** 5196 (2012)
211. P Mikulskis, S Genheden, P Rydberg, L Sandberg, L Olsen, U Ryde *J. Comput.-Aid. Mol. Des.* **26** 527 (2012)
212. V Limongelli, M Bonomi, M Parrinello *Proc. Natl. Acad. Sci.* **110** 6358 (2013)
213. P Tiwary, V Limongelli, M Salvalaglio, M Parrinello *Proc. Natl. Acad. Sci.* **112** E386 (2015)
214. N Plattner, F Nog *Nat. Commun.* **6** art8653 (2015)
215. W L Jorgensen *Acc. Chem. Res.* **42** 724 (2009)
216. A V Veselovsky, A S Ivanov *Curr. Drug Targ. Infect. Disord.* **3** 33 (2003)
217. R G Efremov, D E Nolde, L G Konchina, N P Syrtcev, A S Arseniev *Curr. Med. Chem.* **11** 2421 (2004)
218. G Klebe *Drug Discov. Today* **11** 580 (2006)
219. G Schneider *Nat. Rev. Drug Discov.* **9** 273 (2010)
220. F Svensson, A Karlgn, C Sklld *J. Chem. Inf. Model.* **52** 225 (2012)
221. O A Filz, V V Poroikov *Russ. Chem. Rev.* **81** 158 (2012)
222. S B Sosnin, E V Radchenko, V A Palyulin, N S Zefirov *Dokl. Russ. Acad. Nauk* **463** 293 (2015)
223. D B Kitchen, H Decornez, J R Furr, J Bajorath *Nat. Rev. Drug Discov.* **3** 1549 (2004)
224. C N Cavasotto, N Singh *Curr. Comput.-Aid. Drug Des.* **4** 221 (2008)
225. T Tuccinardini *Comb. Chem. High Throughput Scr.* **12** 303 (2009)
226. Y C Martin, R Abagyan, G G Ferenczy, V J Gillet, T I Oprea, J Ulander, D Winkler, N S Zefirov *Pure Appl. Chem.* **88** 239 (2016)
227. V B Luzhkov, B Selisko, A Nordqvist, F Peyrane, E Decroly, K Alvarez, A Karlen, B Canard, J Åqvist *Bioorg. Med. Chem.* **15** 7795 (2007)
228. V Luzhkov, E Decroly, B Canard, B Selisko, J Åqvist *Mol. Inf.* **32** 155 (2013)
229. W L Jorgensen, J Ruiz-Caro, J Tirado-Rives, A Basavapathruni, K S Anderson, A D Hamilton *Bioorg. Med. Chem. Lett.* **16** 663 (2006)
230. V V Thakur, J T Kim, A D Hamilton, C M Bailey, R A Domaoal, L Wang, K S Anderson, W L Jorgensen *Bioorg. Med. Chem. Lett.* **16** 5664 (2006)
231. C S Leung, J G Zeevaart, R A Domaoal, M Bollini, V V Thakur, K A Spasov, K S Anderson, W L Jorgensen *Bioorg. Med. Chem. Lett.* **20** 2485 (2010)
232. T V Rakitina, A A Zeifman, F N Novikov, O V Stroganov, V S Stroylov, I V Svitanko, A Frank-Kamenetskii, G G Chilov *Mendeleev Commun.* **25** 364 (2015)
233. A V Nemukhin, B L Grigorenko, S V Lushchekina, S D Varfolomeev *Russ. Chem. Rev.* **81** 1011 (2012)
234. R A Bryce, I H Hillier *Curr. Pharm. Res.* **20** 3293 (2014)
235. M J Harvey, G De Fabritiis *Drug Discov. Today* **17** 1059 (2012)
236. J-P Changeux, S Edelstein *F1000 Biol. Rep.* **3** 19 (2011)
237. A C Gibbs *J. Med. Chem.* **57** 7819 (2014)