Review

Molecular Modelling of Lipase Catalysed Reactions. Prediction of Enantioselectivities

Fredrik HÆFFNER and Torbjörn NORIN

Department of Chemistry, Organic Chemistry, Royal Institute of Technology, SE-100 44 Stockholm, Sweden. Received April 7, 1999

The use of *Candida antarctica* lipase B and lipases from *Candida rugosa* and *Rhizomucor miehei* for kinetic resolutions is discussed. Theories and methods for molecular modelling are presented briefly. We have studied the transition state geometry in acyl transfer reactions of serine hydrolases by *ab initio* calculations of proton transfer mechanisms in ester hydrolysis in order to be able to predict enantioselectivity by means of molecular modelling. A combination of molecular modelling and kinetic studies was used to construct models that explain and predict the enantioselectivity of these lipases. The calculations and the experimental results provide support for a proposal that enantiomers can bind in different orientations. Structural factors of importance for the enantioselectivity have been identified. Our results provide information on the scope and limitations of the use of these enzymes in asymmetric synthesis. Molecular modelling also provides information about possible genetic modifications of the enzymes in order obtain a useful catalyst for a specific enantioselective transformation or for the use of the lipase framework and part of the catalytic site for the catalysis of unnatural reactions.

Key words molecular modelling; *ab initio*; ester hydrolysis; lipase, catalysis; enantioselectivity; serine hydrolase

Lipases are hydrolytic enzymes and belong to the serine hydrolase class of enzymes. They have proven to be efficient for resolving racemic alcohols and esters. Many lipases are easily available, heat-stable, and rather robust enzymes. Furthermore they do not need co-factors and can often be used in organic solvents. Therefore, lipases have become useful tools in asymmetric organic syntheses.

Several lipase structures have been determined during the last few years.[1a—l)] Studies of structure–function relationship by means of molecular graphics and molecular modelling are thus possible. Molecular modelling can also be used for guidance on redesigning the enzyme by site-directed mutagenesis in order to improve the capability of the enzyme to accomplish a specific enantioselective transformation. Furthermore, molecular modelling is a useful tool for guidance when one wish to transform the enzyme in such a way that it can catalyze reactions in which an activated carbonyl group is operating in the transition.

Lipases usually have an α/β-hydrolase fold with mostly parallel β-sheets, flanked on both sides by α-helixes. They do not use co-factors for catalysing chemical reactions. Their natural substrates are glycerides. The active site consists of the catalytic triad (Asp/Glu,His,Ser), the oxyanion hole, and sites able to bind and orient the substrates. The nucleophilic serine residue is always positioned in a short loop between a strand and an α-helix. This residue is usually found in the sequence Gly-x-Ser-x-Gly.[2)] The active site pocket is hydrophobic which makes it suitable for accommodating triglycerides and other hydrophobic substrates.

Many lipases contain one or more α-helixes which form a lid (or flap). These lipases are usually inactive when this lid covers the active site. When the lipase approaches a hydrophobic surface, the lid uncovers the active site exposing it to the substrate, and the lipase becomes active. This phenomenon is called interfacial activation.[3,4)] Lipases are usually active not only at a water/hydrophob interface but also in organic solvents.[5,6)] This makes lipases useful not only for hydrolytic reactions but also for esterifications and transesterifications. Lipases have proven to be able to accept a broad number of substrates, a fact that makes them useful for resolving chiral organic substrates.

Molecular Modelling—Theories and Methods

Theoretical models are important for the understanding and prediction of physical and chemical observations. The models can be based on quantum chemistry or molecular mechanics calculations and combinations thereof. Reaction paths through transition states and intermediates, and electronic charge distributions can be studied with quantum chemistry methods. Molecular mechanics can be used to compute the geometry of large molecules in the ground state, heats of formation, and the relative energy of conformers.

Quantum Chemistry and Semiempirical Methods[7)] The quantum theory was born in 1900, when Max Planck presented the distribution law of black body radiation. Later on, quantum mechanics was developed and in 1926 Erwin Schrödinger formulated his famous partial differential equation, the Schrödinger equation. Quantum mechanics is still the best model known for describing systems at the atomic level, and in theory the Schrödinger equation holds the key to all molecular properties.

The time-independent non-relativistic Schrödinger equation describes the wave function of the electrons in a molecule by means of quantum mechanics (*ab initio*). If it were possible to solve the Schrödinger equation, that is, to calculate the wave function, all molecular properties could be determined. However, this can be done exactly only for two particles, the hydrogen atom. One is thus restricted to approximate solutions for larger systems. For this reason, various methods have been developed to derive approximate solutions with high accuracy.

It was not until the beginning of the 1970s that *ab initio* calculations on molecules of chemical interest became feasible. This was the time when computers fast enough to perform quantum chemical calculations first become available. Today quantum chemistry is an important tool in research on

* To whom correspondence should be addressed.

molecular systems.

Semiempirical methods are used to find approximate solutions of the Schrödinger equation. Only the valence electrons are treated with quantum mechanics. Several approximations are made in the treatment of these electrons to reduce the computational costs. The inner core is parameterised from experiments or *ab initio* calculations. The advantage of the semiempirical methods is that they are computationally cheap. The drawback is that they can give erroneous results if applied to systems for which they are not parameterised.

The Hartree–Fock method (HF),[7] which provides an approximate solution of the Schrödinger equation, has been very successful in predicting molecular geometries and energies. It is based on the approximation that the electrons move independently of each other, each one feeling only the mean field of the other plus the attraction from the nuclei.

The density functional theory (DFT) was formulated in 1964 by Hohenberg and Kohn and has become an important complement to the computer-intensive correlated *ab initio* methods.[7] DFT is based on the finding that all ground state electronic properties of a molecule are uniquely determined by the electron density. The advantage of DFT is that it includes electron correlation at approximately the same cost as that of HF calculations. This means that the DFT methods can be good complements to the computer-intensive correlated *ab initio* methods. Unlike the *ab initio* methods, a calculation using the DFT method can not be improved in a systematic way.

Force Field Methods, Molecular Mechanics, and Molecular Dynamics The force field methods and molecular mechanics are empirically based. These methods describe a molecule as a group of atoms connected by springs. In spite of their simplicity, the methods are often good and can be used for larger molecular systems. The force field is based on classical mechanics. It relates the geometry and the potential energy of a molecule by means of an analytical function. The basic idea is that molecules have natural bond lengths and bond angles, which they strive to maintain. The electrons and nuclei are not treated separately. The atoms of the molecule are instead parameterised to act as spheres connected by springs. The first serious attempt to use molecular mechanics was made by T. L. Hill in 1946.[8]

Molecular mechanics consists of an analytical function defined in terms of classical physics. Hooke's law is used to describe the chemical bonds as simple harmonic springs, whereas the atoms are described by van der Waals surfaces and partial charges localised in the centre of the atoms. Molecular mechanics relies upon a number of parameters (force fields), to which values must be assigned before any calculations can be made. Moreover, it does not treat the electrons in the molecule explicitly. Therefore, chemical reactions, in which chemical bonds are formed and broken, can not be studied.

Several force fields have been developed for the handling of various kinds of molecular systems. There are force fields such as MM2,[9] MM3[10—14] and CVFF[15—19] for small organic molecules, and force fields such as CHARMM,[20] AMBER,[21,22] OPLS[23] and GROMOS[24] for proteins, sugars, RNA and DNA. One can ask why there is not just one force field that is good enough to cover all kinds of molecules. One reason is that there were different needs, and consequently, several force fields were developed. For example, the cross-terms (present in *e.g.* MM3), which are of minor importance in protein modelling, would give rise to unnecessary computations if applied to proteins. This is one of the reasons why the force fields developed for protein modelling contain no cross-terms but only independent harmonic bonding terms.

In molecular dynamics (MD) the classical laws of motion

Fredrik Hæffner graduated from the School of Chemistry and Chemical Engineering, Royal Institute of Technology (KTH), Stockholm, Sweden, in 1994 and joined the biocatlysis group of the Chemistry Department of the same university to work for Professor Torbjörn Norin. He obtained his PhD in organic chemistry in 1998 and went for post-doctoral work in the U.S.A. for Professor K. N. Houk, University of California, Los Angeles, U.S.A., and just recently he moved to continue his work for Professor B. Tidor, MIT, Cambridge, U.S.A.

Torbjörn Norin

Torbjörn Norin graduated from the School of Chemistry and Chemical Engineering, Royal Institute of Technology (KTH), Stockholm, in 1957. He continued for a PhD and a DSc in organic chemistry under the supervision of Professor Holger Erdtman. He went for post-doctoral work for Professor Sir Ewart Jones and Professor T. G. Halsall at Dyson Perrins Laboratory, Oxford University, England. He returned to KTH to take up a position as Associate Professor. In 1966 he was appointed Director of Research and head of the chemistry department of the Swedish Forest Products Research Laboratory (STFI) and in 1969 he was appointed to his present position as Professor of Organic Chemistry of KTH. He is fellow of the Royal Swedish Academy of Sciences, the Royal Swedish Academty of Engineering Sciences, and the Academia Europaea. He has served at various boards and councils. He is the President of the Swedish Chemical Society, the President Elect of the Organic Chemistry Division of IUPAC, and has recently served as the chairman of the chemistry section of the Royal Swedish Academy of Sciences, where he now is a member of the board ("Akademinämnden"). His main research interests are within synthetic organic chemistry, asymmetric synthesis, and biocatalysis. The chemistry of natural products and ecological chemistry are also recent topics of his research. He is the author or co-author of more than 200 scientific publications and two textbooks.

are expressed numerically so as to describe the thermal fluctuations of molecules over time.[25,26] A number of phenomena and properties such as structural changes, diffusion coefficients *etc*. can be derived from MD-simulations. In our modelling work we have used MD to search the conformational space for low energy conformations.

Like MD, the Monte Carlo method[25] samples the conformational space. However, the way of doing this is different. By means of a Boltzmann weighted probability (Metropolis condition[25,27]), representative conformations can be sampled for specific ensembles. In contrast to MD, the method can not be used for deriving time-dependent properties.

Other methods that rely on force fields are GRID[28,29,30] and FEP (Free Energy Perturbation).[31] Homology Modelling is another important tool for studies of protein structures.[32] The GRID method has proved to be excellent for highlighting binding sites in proteins.[33] The method is based on the concept that a probe, which mimics a specific functional group, visits arbitrary lattice points in the space around the target protein and calculates the interactions between the probe and the points. It is possible to choose different types of probes. Parameters are established for a number of probes such as water, alcohol, aliphatic groups, phosphate groups, halogens, metal ions and carboxyl groups. The interaction between the probe and the protein is calculated by a force field.[28,29,30] The interaction energies at each site can be visualised by contour maps on a graphic display.

Reaction Mechanism and Transition State Approximation of a Serine Hydrolase Catalysed Reaction

Knowledge of the detailed mechanism of the enzyme-catalysed reaction is a prerequisite for a prediction of enantioselectivity by means of molecular modelling. Lipase-catalysed reactions proceed *via* the "bi-bi ping-pong" mechanism.[34] There is an acyl/enzyme intermediate and two transition states—one for the acylation of the active serine of the enzyme and one for the deacylation of the intermediate serine ester (Chart 1).

The enzyme catalysed reaction is both base and acid promoted. This is possible because the protein scaffold separates those amino acids residues, which function as the base and the acid in the active site. The enzymatic machinery that promotes the base catalysis consists of the catalytic triad (Asp/Glu,His,Ser). The nucleophilicity of the catalytic serine (and the water molecule that enters in the deacylation step) is enhanced by the histidine residue and the aspartate/glutamate residues by means of the so-called charge relay system.[35] The acidic catalysis is promoted by the so-called oxyanion hole, which is formed from several hydrogen bond donors (the amide protons of the enzyme backbone, or of its side-chains). These form several hydrogen bonds to the oxyanion of the transition state.

The serine hydrolases catalyse the ester hydrolysis with a rate which is several orders of magnitude higher than those of the acid or base catalysed reactions in the gas phase and in solution. Several theoretical studies have been published on the catalytic mechanism of the serine hydrolases.[36—38] The Kollmann group studied the mechanism of the serine proteases by means of a model system using semiempirical calculations.[39,40]

In order to obtain further information about the reaction mechanism of the serine hydrolase catalysed ester hydrolysis reaction, we studied in our laboratory a simple model system (Chart 2).[41] The ester studied was methyl formate. The oxyanion hole was modelled by two water molecules. The catalytic triad (Asp/Glu,His,Ser) was modelled by a formate anion, an imidazole ring, and methanol. The geometry optimisations were performed by means of HF and DFT calculations. The calculated potential energy of the acylation step is presented in Fig. 1.

ATS1, ATS2, and ATS3 are transition states, and the points ATd1 and ATd2 are intermediates, see Chart 2 and Fig. 1. The structures at the transition states (ATS1 and ATS3) and the intermediates (ATd1 and ATd2) along the reaction coordinate (Fig. 1) were geometry optimised and checked by means of frequency calculations. However, when

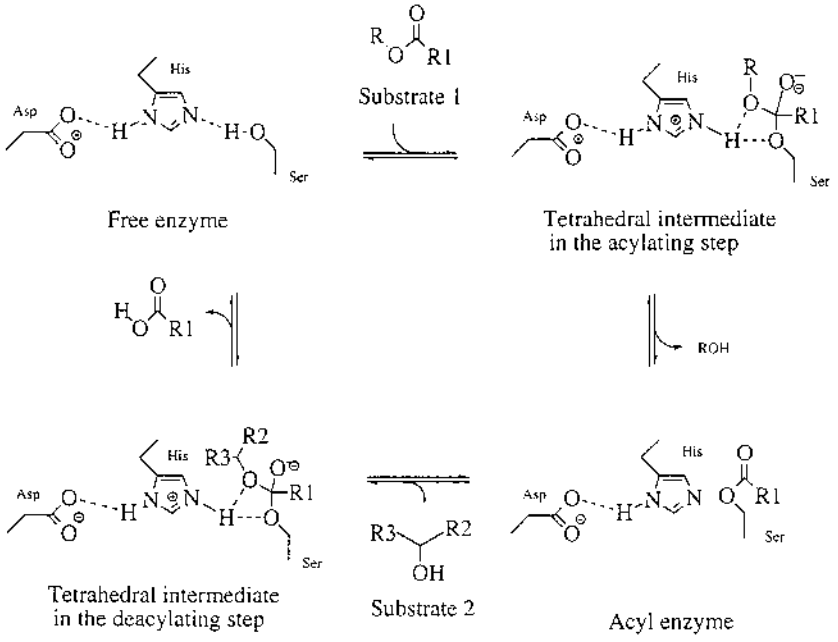

Chart 1. The "Bi-bi Ping-pong" Mechanism of a Serine Hydrolase Catalysed Ester Hydrolysis

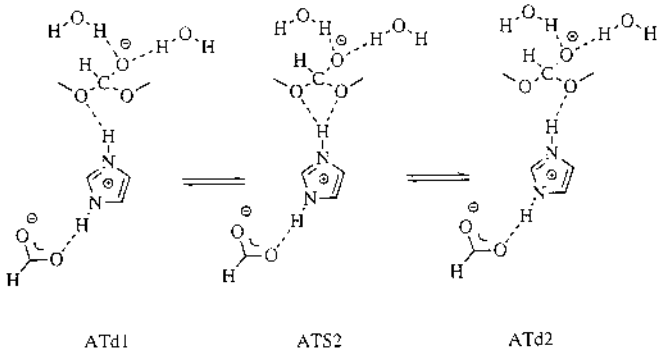

Chart 2. A Model System of a Serine Hydrolase Catalysed Ester Hydrolysis Reaction[41)]

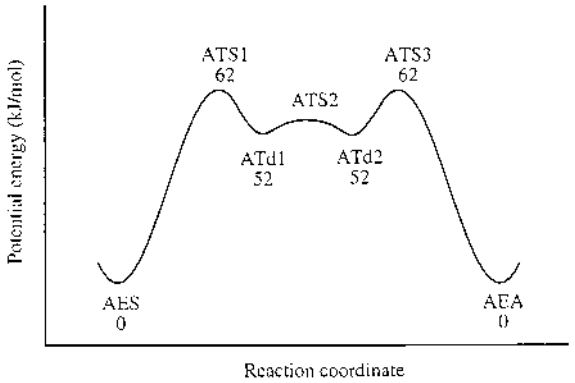

Fig. 1. The Potential Energy Profile of the Acylation Step in the Hydrolysis of Methyl Formate[41)]

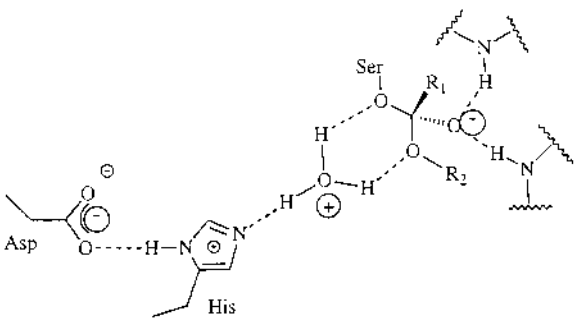

Fig. 2. A Hypothetical 6-Membered Ring Transition State of an Enzyme Catalyzed Ester Hydrolysis Reaction

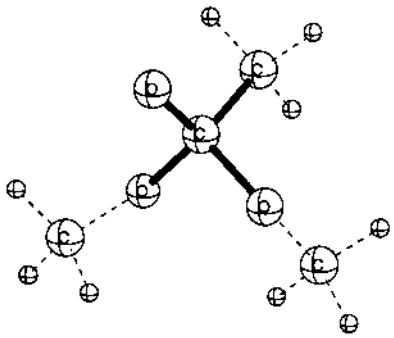

Fig. 3. A Tetrahedral Intermediate with Its Rigid Central Part (the Atoms Connected by Solid Lines) Was Used in Ref.[53)] to Simulate the Enzymatic Transition State

the thermal energy contributions were added to the potential energies, the resulting energy differences between the transition states and the tetrahedral intermediates disappeared.[41)]

The presence of a tetrahedral intermediate could thus not be confirmed in the acylating step. Similar results were obtained in calculations on water assisted hydrolysis of methyl acetate under neutral conditions.[42)]

A 4-membered cyclic transition state (as shown in the Charts 1 and 2), similar to the gas phase transition states, is generally postulated to be valid for the docking to the active site of the enzyme. Our calculations on the hydrolysis of methyl acetate[42)] show, however, that the 4-membered ring transition state is energetically less favoured than the water assisted 6-membered ring transition state. Therefore, it is tempting to postulate a 6-membered ring transition state even in an enzyme catalysed reaction. Perhaps there are enzymes that stabilise 6-membered ring transition states similar to the one presented in Fig. 2. This mechanism would agree well with recent views of proton transfer and high mobility of protons in water.[43,44)]

The importance of using the correct geometry of the transition states in the modelling of enantioselectivity is obvious. The stability of this activated complex determines the chemical reaction rate. Ideally, the model of the true transition state of a serine hydrolase catalysed hydrolysis (or of any enzymatic reaction) should include the entire enzyme, the docked-in substrate(s), and the surrounding solvent. The model should be constructed by means of quantum mechanical calculations. As this is not feasible, one must make several approximations.

Simple models of the tetrahedral intermediates in the acylating as well as the deacylating steps in lipase-catalysed acyl transfer reactions have been to simulate the true transition states in several theoretical studies.[45—50)] We used a simple model of a tetrahedral intermediate structure, in which bond lengths and bond angles were chosen from existing parameters in the AMBER force field.[50,51)] Point charges of the tetrahedral intermediate were calculated from the electrostatic potential, which was computed by means of the semiempirical method MNDO.[52)] We have also used a more sophisticated model of the tetrahedral intermediate.[53)] In this model the geometry presented in Fig. 3 was optimised by means of an HF calculation. This structure was used to simulate the enzymatic tetrahedral transition state. Point charges were derived from the electrostatic potential. The central part (see Fig. 3) of the tetrahedral intermediate was kept fixed during the molecular dynamics and the molecular mechanics simulations by means of large force constants.

In our modelling studies the tetrahedral transition state structure has been introduced into the active site with the oxyanion stabilised by hydrogen bonds to the oxyanion hole.[46,51,53)] Furthermore, one hydrogen bond is formed between the ring nitrogen (N) in the catalytic histidine residue and the side-chain oxygen (O) of the catalytic serine, and one further hydrogen bond is formed from the same nitrogen atom to the alcohol oxygen (cf. the tetrahedral intermediates in Chart 1). In the light of the *ab initio* studies performed by Brinck *et al.*,[41)] this model has been justified as a good representation of the enzymatic transition state.

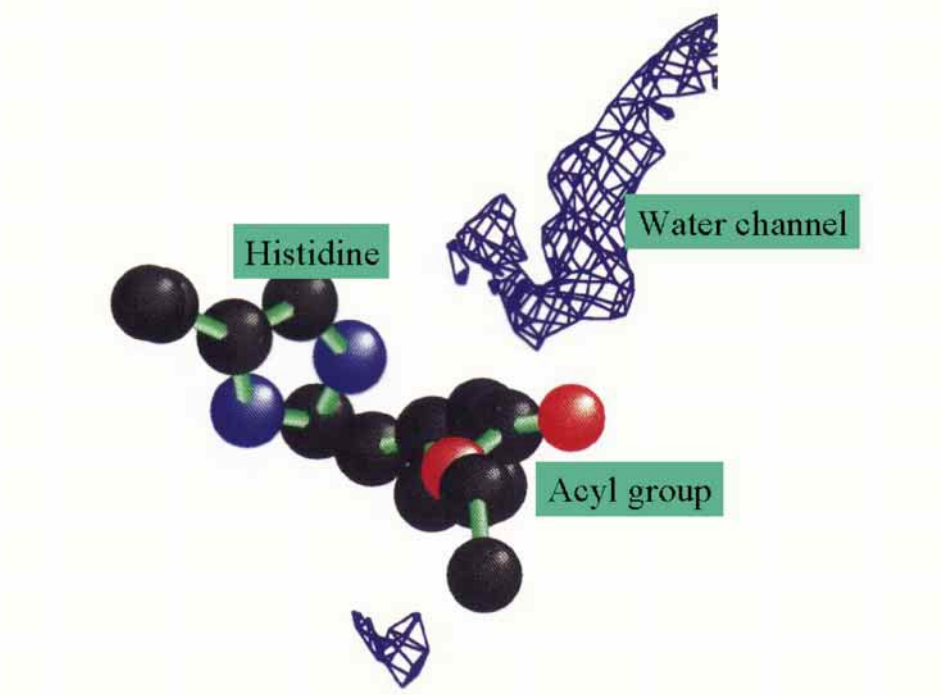

Fig. 4. Part of the Active Site of *Rhizomucor miehei* Lipase Showing the Water Channel Ending at the Carbonyl Carbon of the Acyl Enzyme Intermediate
The cannel is identified by means of the GRID method.[33]

The Treatment of Solvent in Molecular Modelling

It is important to take the solvent into account in the calculations. The modelling of solvent interactions can be treated in two fundamentally different ways. One is the explicit treatment of the solvent molecules and the other is the treatment of the solvent as a non-structured dielectric continuum. Explicit solvation, in which the solute molecule is surrounded by hundreds or thousands of solvent molecules, is used in molecular dynamics and Monte Carlo simulations. The solute is solvated in a periodic box filled with the solvent molecules.[54] To keep the number of solvent molecules constant in the box during the simulation, those solvent molecules leaving the box at one side will be put back in at the opposite side of the box. Explicit solvation is not feasible in quantum chemistry, because of the system size and the statistical sampling needed.

Explicit solvent representation has several advantages. One is, of course, the structural information obtained regarding the orientation of the solvent molecules around the solute molecule. This is especially important when the solvent molecules can form hydrogen bonds with each other and the solute molecule.

The disadvantages of explicit solvation are the costly simulations, the lack of polarizability (which is a general problem in the force field method), and the statistical sampling needed to derive ensembles.

The structural treatment of the solvent is most important in modelling the structure and function of a protein. The three-dimensional architecture of proteins is often dependent on several water molecules that are buried in the protein molecule or attached to its surface.[55]

The specific solvation was taken into account in our modelling of the basic hydrolysis of methyl acetate.[42] Using *ab initio* calculations, we found that the proton transfer was mediated by water. Unpublished results from similar *ab initio* studies on the specific solvation in the hydrolysis of methyl acetate under acidic and neutral conditions indicate a similar water mediated mechanism.

The GRID method has been used successfully in studies of the interaction between the enzyme and the solvent.[33] In particular, this method is useful for the localisation of hydrophobic and hydrophilic interfaces of the enzyme. One significant observation concerns the binding of water molecules near the active site of the *Rhizomucor miehei* lipase. In this enzyme there is a water channel in the hydrophobic interface and this channel ends at the active site. Calculations of water-binding sites in the model acyl enzyme intermediate showed that the active site may accommodate a water molecule in a position that is favourable for in-plane nucleophilic attack on the carbonyl carbon of the acyl group (Fig. 4). In this position the water molecule may donate one proton to the imidazole of the catalytic triade. This arrangement fits nicely with a 6-membered transition state, as depicted in Fig. 2.

Several non-structured continuum solvation methods have been developed. The simplest way to simulate the solvent in the force field method is to set the dielectric constant in the term describing the electrostatics at that of the solvent. This is an easy but not very accurate way of treating the solute–solvent interactions. However, because of its simplicity it can be used as a first approximation. In a MD-simulation study, we have found vacuum to be fairly good model of a hydrophobic solvent.[56] In this work a lipase from *Rhizomucor miehei* was studied by means of molecular dynamics. The hydrophobic solvent was modelled in two ways: as a vacuum and as a periodic box, filled with the lipase and methyl hexanoate molecules. The results from these simulations showed that the lipase behaved similarly in these two solvent representations.

A model, in which the solvent is represented by a dielectric continuum, was developed as early as 1934 by Onsager.[57] The solute is placed in a spherical cavity, with a radius that should be representative of the solute, in this continuum and interacts with the dielectric continuum *via* a dipole, centred in the cavity. The dipole polarises the dielectric continuum, which will in return polarise the solute. The Onsager model can be implemented in the Hartree–Fock equations as an additional term in the Hamiltonian operator. The main limitation of this model is that the electrostatics is approximated only by the dipole moment. Moreover, the spherical cavity is not a proper representation of non-spherical molecules. The advantages are that the model is simple and can easily be implemented in methods that include electron correlation.

The second non-structured dielectric continuum model to be discussed is the Polarised Continuum Model (PCM).[58,59] In the PCM method, the solvent is represented by an apparent charge distribution over the cavity surface. We have used this method in modelling of the hydrolysis of methyl acetate.[42]

Molecular Modelling for Prediction of Enantioselectivity

Empirical models of the active sites of enzymes were derived by screening the reactivity of many substrates. Jones *et al*. constructed a box-model of the active site of pig liver esterase.[60] Kazlauskas *et al*. derived an empirical rule that can predict the enantioselectivity of cholesterol esterase, *Pseudomonas cepacia* lipase, and *Candida rugosa* lipase towards secondary alcohols.[61] However, the details in which the enzymes actually discriminate between substrates and enantiomers can not be understood by using these empirical rules.

Several attempts have been made to calculate the enantioselectivity by means of force field methods resulting in good prediction of the fast reacting enantiomer on a qualitative level.[45,47,48,50,51,62—66] The calculated energy differences between the enantiomers are of proper magnitude and provide correct leads for a proper evaluation of the reaction. One should bear in mind that in a reaction, which provides an enantiomeric excess of approximately 95% (corresponding to an E-value of approximately 100), the free energy difference between the reactions of the two enantiomers is slightly less than 3 kcal/mol. This is on the limit of the accuracy of the calculations. However, since the calculations are made on very similar systems (enantiomeric substrates) and the ground state energies of the two systems are equal, the results of the calculations can provide correct and useful information.

Various strategies can be used for determining the binding of the two enantiomers in the active site during the reaction. One strategy is to use chiral inhibitors that mimic the activated complexes of the enantiomers in the active site of the enzyme.[60] Separate crystallisations of the enzyme with the two enantiomeric inhibitor complexes, followed by 3-D structure determination, will reveal how the enantiomers are oriented in the active site.[67,68] Another way, which has been used in our laboratory, is to combine enzyme kinetics with molecular modelling to elucidate the binding orientation of the enantiomers.[48,50,51,65]

Conformational Space for Calculations of Enantioselectivity

Manual docking of the tetrahedral intermediate is, of course, a very important part in searching the conformational space for the calculations of enantiomeric selectivity. In order to limit the computer time needed for the calculations, only atoms within a certain distance of any atom of the tetrahedral intermediate were included in the calculations.[50] No constrains were put on the system, but all atoms were allowed to relax in molecular mechanics minimisations. Such a method did not search the conformational space to any great extent but provided the possibility to manage calculations with modest computational resources.

We used a somewhat similar approach for calculating the enantiomeric ratios in acyl transfer reactions catalysed by α-chymotrypsin.[48] The active site region and the tetrahedral transition state structure were free to move in the simulations. The region surrounding the active site was kept fixed. The rest of the protein and the water molecules were not included in the calculations. A simulated annealing protocol was developed, where the system was alternatively heated and cooled by means of a molecular dynamics simulation. Low temperature structures were sampled and refined by means of molecular mechanics minimisations. This method searched the conformational space of the enzyme–tetrahedral intermediate complex more efficiently than the ones discussed earlier. The difference in energy between the diastereomeric enzyme–tetrahedral intermediate complexes was approximated by the difference in force field energy between the two complexes. A similar protocol was used in our modelling of reactions with *Candida antarctica* lipase B[51] and *Candida rugosa* lipase.[50]

More recently we developed a method using energy-based subsets.[51] These subsets were constructed from amino acid residues and water molecules in the enzyme structure, which interacted differently with the two diastereomeric enzyme–tetrahedral intermediates. The structures, from which these energy-based subsets were collected, had been subjected to molecular dynamics simulations (where all atoms were free to move) as well as to structure refinement by means of molecular mechanics calculations of some of the structures collected from the molecular dynamics simulation. Using this method we could predict the fast reacting enantiomers for both small and bulky substrates (*e.g*. those shown in Chart 3).

The method nicely presents how the interactions causing the enantioselectivity are distributed in the active site region. The amino acids in the enzyme, which are important for the

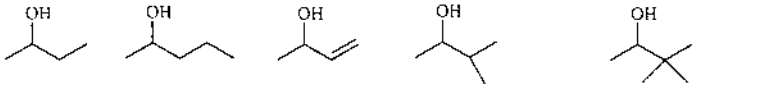

Chart 3. Substrates Used for the Studies of Conformational Space and Enantioselectivity[50,51]

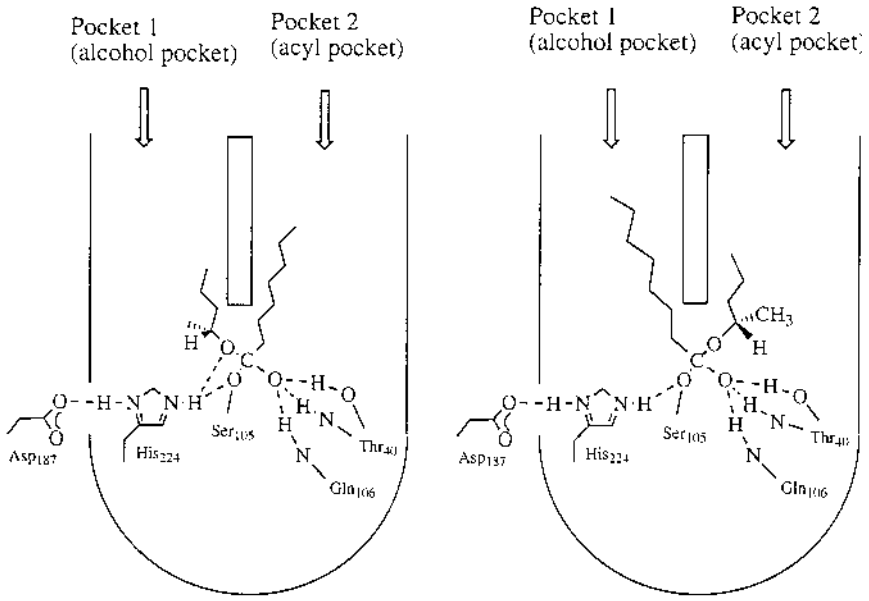

Fig. 5. The Enzyme Transition State Structure, Constituting the Alcohol and the Acyl Donor, Docked in Two Different Binding Modes in the Active Site of *Candida antarctica* Lipase B

The mode with the acyl group in Pocket 2 is the productive one.[2,50,51]

enantioselectivity, are extracted. This is of interest for the development of a genetically modified enzyme for a specific selective transformation. The modelling shows that different amino acid residues cause the chiral discrimination of 2-butanol and 3,3-dimethyl-2-butanol. Thus a static way of assigning the subsets (by means of geometrical constraints) is not appropriate for calculating the enantiomeric ratios, but the subsets should instead be constructed from residues selected from differences in interaction energy, as discussed above.

Candida antarctica Lipase B—Binding Modes of Transtion States

The active site in *Candida antarctica* lipase B can be considered as being built from two pockets separated by several hydrophobic amino acid residues, with the catalytic machinery positioned at the bottom of the pockets. In the X-ray structure of the lipase with a covalently bonded phosphonate inhibitor, the acyl part of the inhibitor was found to be located in Pocket 1, whereas the alcohol part was encountered in Pocket 2 (Fig. 5). However, it was found from molecular modelling, that with such an orientation, the proton transfer between the alcohol and the histidine residue was impeded.[2] For that reason the acyl part of the tetrahedral intermediate must be introduced into Pocket 2 (also labelled as the acyl pocket) and the alcohol must be placed in pocket 1 (labelled as the alcohol pocket).

Candida antarctica lipase B is most useful for the kinetic resolution of *sec*-alcohols, but it has no measurable activity on reactions with tertiary alcohols or with *sec*-alcohols with two large or bulky groups. The chiral alcohol part of the transition state must find the best orientation for productive catalysis when the acyl group is in the "acyl pocket". There is restricted space in the active site of the enzyme and a "stereospecificity pocket" in the alcohol binding site (Fig. 6). Considering the binding of the transition states, the enzyme allows two modes of orientation of the alcohol moiety for productive catalysis. Mode I places the large substituent straight out toward the active site entrance and the medium substituent in the stereospecificity pocket. In Mode II the medium substituent is positioned out towards the surface of the enzyme and the larger substituent is in the limited inner space of the active site.

Candida rugosa Lipase—Modelling of Enantioselectivity and Enantioselective Inhibition

X-ray crystallographic methods have proved that the *Candida rugosa* lipase contains a substrate-binding tunnel, which can accommodate unbranched acyl donors.[1f,68,69] A lid covers the active site when the lipase is in water solution. It is believed that the acyl part of the substrate is binding in the tunnel during the catalysis.[68] This lipase has been found to be more suitable than other lipases for the resolution of 2-methylalkanoic acids.[70—72]

We constructed the transition states of the enantiomers of heptyl α-methyldecanoate, covalently bonded to the side-chain oxygen of Ser209 (see Fig. 7).[50] We positioned the acyl donor parts in the tunnel. The oxyanions were then oriented in such a way that three hydrogen bonds were formed between the oxyanion and the oxyanion hole in both transition states. Furthermore, the oxygen of the side-chain of Ser209 and the oxygen of the alcohol (the leaving group) were placed so as to form two hydrogen bonds to the ring nitrogen of His449. We then used the molecular mechanics and molecular dynamics protocol[48] in order to find low energy enzyme–tetrahedral transition states.[50]

These simulations showed that the (*S*)-enantiomer was lower in energy than the (*R*)-enantiomer. In addition, the α-methyl group of the (*R*)-enantiomer had displaced His449 during the simulations in such a way that the two hydrogen bonds to this residue were broken thus giving rise to a non-productive conformation. The enantiopreference predicted was in qualitative agreement with the one found from experiments. However, the simulation of the (*R*)-enantiomer

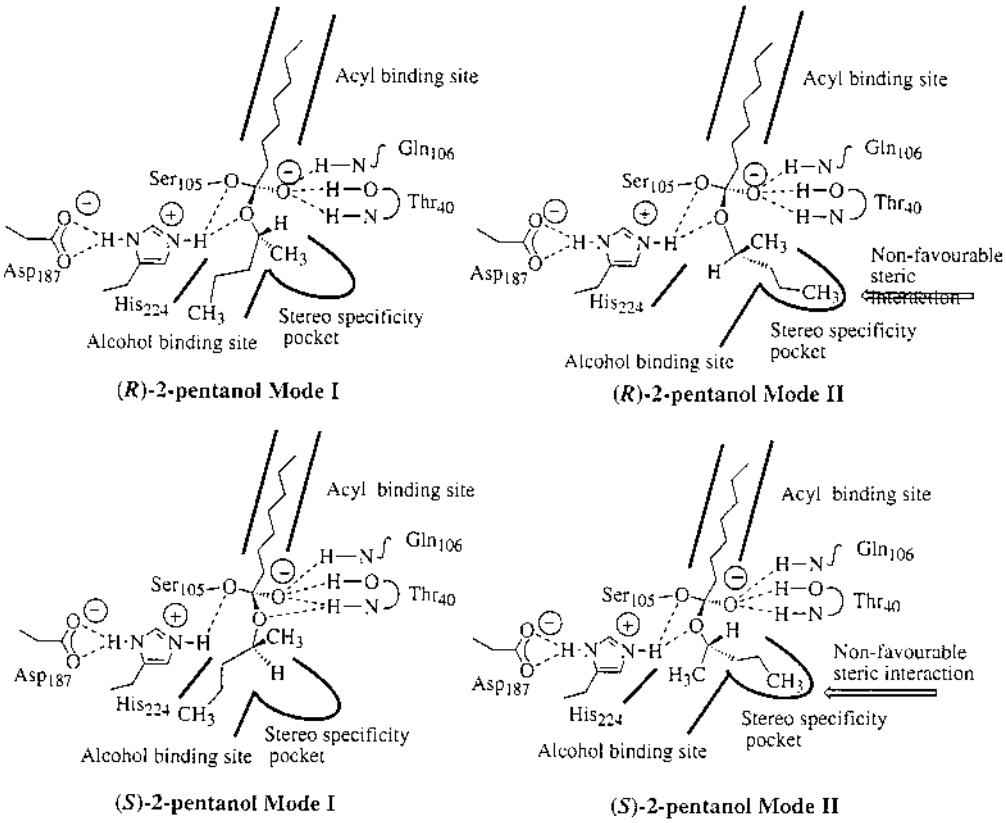

Fig. 6. Modelled Tetrahedral Transition States in the Kinetic Resolution of 2-Pentanol Using *Candida antarctica* Lipase B

The lowest potential energy is found for the (*R*)-enantiomer in the mode I docking. The (*S*)-enantiomer in mode I docking is high in energy and displays a ruptured hydrogen bonding pattern. In the mode II docking the (*S*)-enantiomer displays serious steric interactions and is high in energy.[51]

demonstrated that binding the acyl donor in the tunnel was a non-productive binding mode.

We performed two more simulations, in which the acyl moieties of both enantiomers were oriented so as to point out towards the solvent surface in a "hair-pin" mode (Fig. 7). The α-methyl group of the (*R*)-enantiomer pointed into the tunnel ("hair-pin" mode). With this mode of binding of the tetrahedral transition state structures, the resulting optimised structures showed catalytically necessary hydrogen bonding patterns. However, the (*S*)-enantiomer was higher in energy than the (*R*)-enantiomer. The conclusion that can be drawn from these simulations is that the (*S*)-enantiomer in the "hair-pin" mode will bind its acyl moiety in the tunnel, whereas the (*R*)-enantiomer will bind its acyl moiety towards the solvent surface ("hair-pin" mode).[50]

The enantioselectivity of the *Candida rugosa* lipase catalysed transesterification of α-methyldecanoic acid decreased with increasing heptanol concentration.[73] Molecular modelling has been used successfully to explain this behaviour.[50]

As discussed above, the (*S*)-enantiomer binds in the tunnel whereas the (*R*)-enantiomer does not. Instead it binds in a "hair-pin" binding mode leaving the tunnel empty (Fig. 7). The alcohol will act as a competitive inhibitor towards the (*S*)-enantiomer by binding in the tunnel as well. The (*R*)-enantiomer, on the other hand, with its acyl moiety pointing towards the solvent surface will not be affected by the heptanol (Fig. 8). These different modes of binding the enantiomers nicely explain why the enantioselectivity decreases when the concentration of the long non-branched aliphatic alcohol is increased.

This model is supported by the fact that the *Candida rugosa* lipase shows a reversed enantioselectivity towards substrates with branched (and bulky) acyl moieties.[74] Berglund *et al*. performed a study in order to test our hypothesis with the bulky substrate 2-methyl-6-(2-thienyl)hexanoic acid.[75] Their results strongly support our hypothesis of a different binding of the enantiomers.

Conclusions and Perspectives

Molecular modelling has become an important tool for studies of enzyme selectivities. In particular, the combination of experimental studies of enzyme kinetics and selectivities with computer modelling provides important information for the use of enzymes in selective organic synthesis. Recent protocols for modelling and calculations of molecular interactions provide possibilities for qualitative predictions with rather moderate computer facilities. Important substrate–enzyme interactions can be located and sites for site-directed mutagenesis or chemical transformations, leading to improved performance of the enzyme, can be identified. A tempting approach is to transform the enzyme in such a way that essential parts of its active site can be used for the catalysis of a reaction that is different from its natural function. Thus it should be possible to transform a lipase to catalyse reactions in which the carbonyl function of an ester, an aldehyde, or a ketone is activated for a condensation reaction. Many of the commercial enzymes are robust proteins and should be able in such a way to provide useful scaffolds for transformations to efficient and selective catalysts for organic synthesis.

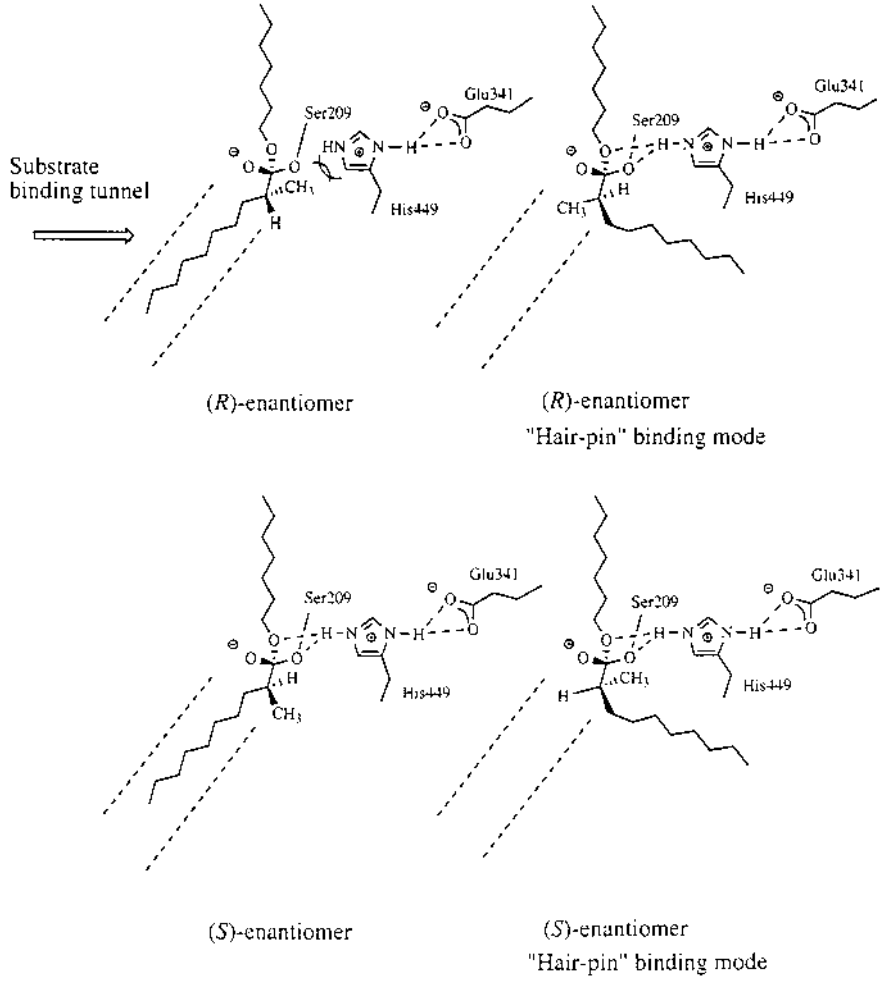

Fig. 7. Different Modes of Binding the Tetrahedral Transition States Constituting the Enantiomers of α-Methyldecanoic Acid in the Active Site of *Candida rugosa* Lipase[50]

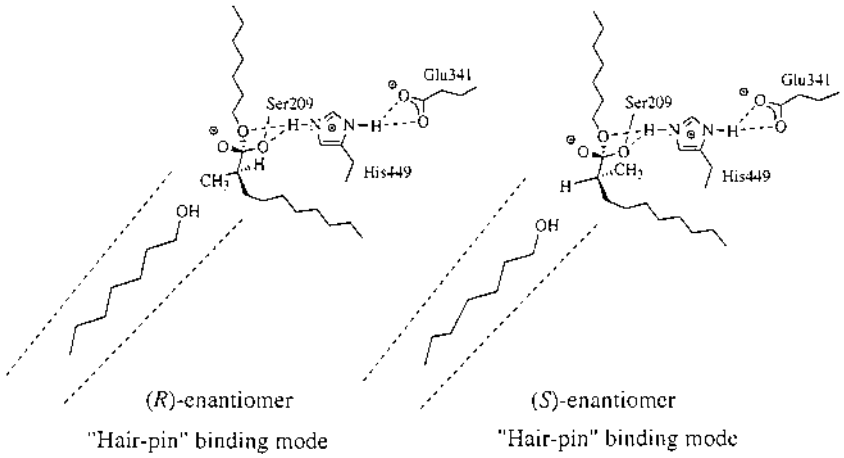

Fig. 8. Competitive Inhibition Between Heptanol and the Productive Binding Mode of the Tetrahedral Transition State of (*S*)-Enantiomer of Heptyl α-Methyldecanoate[50]

Acknowledgements We gratefully acknowledge the work of our colleagues in the biocatalysis group of KTH, on which this survey is founded. They are named as co-authors in the references. Dr. Gunhild Aulin-Erdtman is gratefully acknowledged for discussions. Financial support has been provided by the Swedish Natural Science Research Council (NFR), the Swedish Council for Forestry and Agricultural Research (SJFR), the Swedish Research Council for Engineering Sciences (TFR) and the Trygger Foundation.

References

1) *a*) Brady L., Brzozowski A. M., Derewenda Z. S., Dodson E., Dodson G. G., Tolley S., Turkenburg J. P., Christiansen L., Huge-Jensen B., Norskov L., Thim L., Menge U., *Nature*, **346**, 767—770 (1990); *b*) Winkler F. K., D'Arcy A., Hunziker W., *ibid*., **343**, 771—774 (1990); *c*) Schrag J. D., Li Y., Wu S., Cygler M., *ibid*., **351**, 761—764 (1991); *d*) Martinez C., De Geus P., Lauwereys M., Matthyssens G., Cambil-

lau C., *ibid*., **356**, 615—618 (1992); *e*) van Tilbeurgh H., Sarda L., Verger R., Cambillau C., *ibid*., **359**, 159—162 (1992); *f*) Grochulski P., Li Y., Schrag J. D., Bouthillier F., Smith P., Harrison D., Rubin B., Cygler M., *J. Biol. Chem*., **268**, 12843—12847 (1993); *g*) Martinez C., De Geus P., Stanssens P., Lauwereys M., Cambillau C., *Protein Eng*., **6**, 157—165 (1993); *h*) Schrag J. D., Cygler M., *J. Mol. Biol*., **230**, 575—591 (1993); *i*) van Tilbeurgh H., Egloff M. P., Martinez C., Rugani N., Verger R., Cambillau C., *Nature*, **362**, 814—820 (1993); *j*) Uppenberg J., Hansen M. T., Patkar S., Jones T. A., *Structure*, **2**, 293—308 (1994); *k*) Hermoso J., Pignol D., Kerfelec B., Crenon I., Chapus C., Fontecilla-Camps J. C., *J. Biol. Chem*., **271**, 18007—18016 (1996); *l*) Lang D., Hofmann B., Haalck L., Hecht H. J., Spener F., Schmid R. D., Schoburg D., *J. Mol. Biol*., **259**, 704—717 (1996).

2) Uppenberg J., Öhrner N., Norin M., Hult K., Kleywegt G. J., Patkar S., Waagen V., Anthonsen T., and Jones T. A., *Biochemistry*, **34**, 16838—16851 (1995).
3) Brzozowski A. M., Derewenda U., Derewenda Z. S., Dodson G. G., Lawson D. M., Turkenburg J. P., Björkling F., Huge-Jensen B., Patkar S. A., Thim L., *Nature*, **351**, 491—494 (1991).
4) Derewenda U., Brzozowski A. M., Lawson D. M., Derewenda Z. S., *Biochemistry*, **31**, 1532—1541 (1992).
5) Zaks A., Klibanow A. M., *Science*, **224**, 1249—1251 (1984).
6) Zaks A., Klibanow A. M., *J. Am. Chem. Soc*., **108**, 2767—2768 (1986).
7) For a presentation see: Szabo A., Ostlund N. S., "Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory," Dover Publications Inc., Mineola, New York, 1996.
8) Hill T. L., *J. Chem. Phys*., **14**, 465 (1946).
9) Burkert U., Allinger N. L., ACS Monograph **177**, Am. Chem. Soc., Washington, DC, (1982).
10) Allinger N. L., Chen K., Rahman M., Pathiaseril A., *J. Am. Chem. Soc*., **113**, 4505—4517 (1991).
11) Allinger N. L., Li F., Yan L., *J. Comput. Chem*., **11**, 848—867 (1990).
12) Allinger N. L., Rahman M., Lii J.-H., *J. Am. Chem. Soc*., **112**, 8293—8307 (1990).
13) Allinger N. L., Yuh Y. H., Lii J.-H., *J. Am. Chem. Soc*., **111**, 8551—8566 (1989).
14) Lii J.-H., Allinger N. L., *J. Comput. Chem*., **12**, 186—199 (1991).
15) Hagler A. T., Dauber P., Lifson S., *J. Am. Chem. Soc*., **101**, 5131—5141 (1979).
16) Hagler A. T., Ewig C. S., *Comp. Phys. Comm*., **84**, 131—155 (1994).
17) Hagler A. T., Lifson S., Dauber P., *J. Am. Chem. Soc*., **101**, 5122—5130 (1979).
18) Hagler A. T., Osguthorpe D. J., Dauber-Osguthorpe P., Hemple J. C., *Science*, **227**, 1309—1315 (1985).
19) Hagler A. T., Stern P. S., Sharon R., Becker J. M., Naider F., *J. Am. Chem. Soc*., **101**, 6842—6852 (1979).
20) Brooks C. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., Karplus M., *J. Comput. Chem*., **4**, 187—217 (1983).
21) Weiner S. J., Kollman P. A., Case D. A., Singh U. C., Ghio C., Alagona G. S., Profeta S. J., Weiner P., *J. Am. Chem. Soc*., **106**, 765 (1984).
22) Weiner S. J., Kollman P. A., Nguyen D. T., Case D. A., *J. Comput. Chem*., **7**, 230—252 (1986).
23) Pranata J., Wierschke S. G., Jorgensen W. L., *J. Am. Chem. Soc*., **113**, 2810—2819 (1991).
24) Hermans J., Berendsen H. J. C., van Gunsteren W. F., Postma J. P. M., *Biopolymers*, **23**, 1513 (1984).
25) Heermann D. W., "Computer Simulation Methods," Second Edition, Springer-Verlag: Berlin/Heidelberg, 1990.
26) Karplus M., Petsko G. A., *Nature*, **347**, 631—639 (1990).
27) Dean P. M., "Molecular Foundations of Drug-Receptor Interaction," First Edition, Cambridge University Press, Cambridge, 1987.
28) Goodford P. J. A., *J. Med. Chem*., **28**, 849—857 (1985).
29) Wade R. C., Clark K. J., Goodford P. J., *J. Med. Chem*., **36**, 140—147 (1993).
30) Wade R. C., Goodford P. J., *J. Med. Chem*., **36**, 148—156 (1993).
31) Rao S. N., Singh C., Bash P. A., XXXX K. P. A., *Nature*, **328**, 551—554 (1987).
32) Blundell T. L., Carney D., Gardner S., Hayes F., Howlin B., Hubbard T., *Eur. J. Biochem*., **172**, 513—520 (1988).
33) Norin M., Hæffner F., Achour A., Norin T., Hult K., *Protein Science*, **3**, 1493—1503 (1993).
34) Kraut J., *Ann. Rev. Biochem*., **46**, 331—358 (1977).
35) Blow D. M., Birktoft J. J., Hartley B. S., *Nature*, **221**, 337 (1969).
36) Kollman P. A., Hayes D. M., *J. Am. Chem. Soc*., **103**, 2955—2961 (1981).
37) Warshel A., Naray-Szabo G., Sussman F., Hwang J.-K., *Biochemistry*, **28**, 3629—3637 (1989).
38) Warshel A., Russell S. T., *J. Am. Chem. Soc*., **108**, 6569—6579 (1986).
39) Daggett V., Schröder S., Kollman P., *J. Am. Chem. Soc*., **113**, 8926 (1991).
40) Schröder S., Daggett V., Kollman P., *J. Am. Chem. Soc*., **113**, 8922 (1991).
41) Hu C.-H., Brinck T., Hult K., *Int. J. Quantum Chem*., **69**, 89—103 (1998).
42) Hæffner F., Han-Hu C., Brinck T., Norin T., *J. Mol. Struct*. (*THEOCHEM*), **459**, 85—93 (1999).
43) Marx D., Tuckerman M. E., Hutter J., Parinello M., *Nature*, **397**, 601—604 (1999).
44) Kirby A. J., *Acc. Chem. Res*., **30**, 290—296 (1997).
45) DeTar D. F., *Biochemistry*, **20**, 1730—1743 (1981).
46) Wipff G., Dearing A., Weiner P. K., Blaney J. M., Kollman P. A., *J. Am. Chem. Soc*., **105**, 997—1005 (1983).
47) Warshel A., Naray-Szabo G., Sussman F., Hwang J.-K., *Biochemistry*, **28**, 3629—3637 (1989).
48) Norin M., Hult K., Mattson A., Norin T., *Biocatalysis*, **7**, 131—147 (1993).
49) Norin M., Hæffner F., Achour A., Norin T., Hult K., *Protein Science*, **3**, 1493—1503 (1994).
50) Holmquist M., Hæffner F., Norin T., Hult K., *Protein Science*, **5**, 83—88 (1996).
51) Orrenius Ch., Hæffner F., Rotticci D., Öhrner N., Norin T., Hult K., *Biocat. Biotransform*., **16**, 1—15 (1998).
52) Dewar M. J. S. Thiel W., *J. Am. Chem. Soc*., **99**, 4899 (1977).
53) Hæffner F., Norin T., Hult K., *Biophys. J*., **74**, 1251—1262 (1998).
54) van Gunsteren W. F., *Eur. J. Biochem*., **204**, 947—961 (1992).
55) Denisov P. V., Peters J., Hörlein H. D., Halle B., *Nature Structural Biology*, **3**, 505—509 (1996).
56) Norin M., Hæffner F., Hult K., Edholm O., *Biophys. J*., **67**, 548—559 (1994).
57) Onsager L., *J. Am. Chem. Soc*., **58**, 1486 (1936).
58) Miertus S., Scrocco E., Tomasi J., *Chem. Phys*., **55**, 117—129 (1981).
59) Miertus S., Tomasi *J., Chem. Phys*, **65**, 239—245 (1982).
60) Lee T., Sakowicz R., Martichonok V., Hogan J. K., Gold M., Jones J. B., *Acta Chem. Scand.*, **50**, 697—706 (1996).
61) Kazlauskas R. J., Weissfloch A. N. E., Rappaport A. T., Cuccia L. A., *J. Org. Chem*., **56**, 2656—2665 (1991).
62) Bemis G.W., Carlson-Golab G., Katzenellenbogen J. A., *J. Am. Chem. Soc*., **114**, 570—578 (1992).
63) Faber K., Griengl H., Hönig H., Zuegg J., *Biocatalysis*, **9**, 227—239 (1994).
64) Norin M., Hæffner F., Hult K., *Biophys. J*., **67**, 548—559 (1998).
65) Rottichi D., Hæffner F., Orrenius Ch., Norin T., Hult K., *J. Mol. Cat. B*: *Enzymatic*, **5**, 267—272 (1998).
66) Orrenius Ch., van Heusden C., van Ruiten J., Overebeeke P. L. A., Kierkels H., Duine J. A., Jongejan J. A., *Protein Eng*. **11**, 1147—1153 (1998).
67) Ahmed S. N., Kazlauskas R. J., Morinville A. H., Grochulski P., Schrag J. D., Cygler M., *Biocatalysis*, **9**, 209—225 (1994).
68) Grochulski P., Bouthillier F., Kazlauskas R. J., Serreqi A. N., Schrag J. D., Ziomek E., Cygler M., *Biochemistry*, **33**, 3494—3500 (1994).
69) Grochulski P., Schrag J. D., Cygler M., *Protein Science*, **3**, 82—91 (1994).
70) Engel K. H., *Tetrahedron Asymmetry*, **2**, 165—168 (1991).
71) Holmberg E., Holmquist M., Hedenström E., Berglund P., Norin T., Högberg H.-E., Hult K., *Appl. Microbiol. Biotechnol*., **35**, 572—578 (1991).
72) Sonnet P. E., Baillargeon M. W., *Lipids*, **26**, 295—300 (1991).
73) Berglund P., Holmquist M., Hult,K., Högberg H.-E., *Biotechnol Lett*, **17**, 55—60 (1995).
74) Ahmed S. N., Kazlauskas R. J., Morinville A. H., Grochulski P., Schrag J. D., Cygler M., *Biocatalysis*, **9**, 209—225 (1994).
75) Berglund P., Holmquist M., Hult K., *J. Mol. Catalysis B*, *Enzymatic*, **5**, 283—287 (1998).