HUNGARIAN UNIVERSITY OF AGRICULTURAL AND LIFE SCIENCES

MODELLING THE INTERACTIONS OF EXTRACELLULAR ENZYMES AND BIOACTIVE SUBSTANCES

JÁNOS ECKER

GÖDÖLLŐ

2022

DOI: 10.54598/003090

The doctoral school

Name:	Doctoral School of Biological Sciences
Discipline:	Biological sciences
Head of school:	Dr. Zoltán Nagy Head of Department, DSc Hungarian University of Agriculture and Life Sciences, Institute of Agronomy, Department of Plant Physiology and Plant Ecology
Supervisor:	Dr. László Fülöp PhD, Habil.

....

Approval of the head of school Approval of the supervisor

BACKGROUND AND OBJECTIVES

Extracellular enzymes or exoenzymes perform their function in the extracellular space. Such enzymes are produced by bacteria and eukaryotic organisms and are involved in many biological processes, such as the degradation of macromolecules. Some enzymes can also convert xenobiotics, thus playing a role in bioremediation, and can even transform contaminants into biologically active secondary products that can be used as active ingredients in pesticides or antimicrobial drugs. Once secreted, exoenzymes are released from cellular control and, depending on the enzyme, must have catalytic activity in a variety of physico-chemical media. However, the study of enzymes is cumbersome, lengthy and often difficult to optimise the conditions. Thus, in addition to performing enzyme activity assays in different media, modelling them at the molecular level is a very important task. The advantage of the latter methods is that they are not very material-demanding, so that with a well-chosen software background and sufficiently accurate models in terms of the possibilities and the goals to be achieved, many molecular-level mechanisms can be visualised and many properties of the molecules can be analysed and quantified.

These approaches are still relatively new, but their methodology is constantly growing and the methods used are producing increasingly accurate results. Therefore, their scientific relevance is now unquestionable, as confirmed by the results of several national and international research groups published in reputable journals. Such research is therefore worthwhile, as the results are in demand internationally.

In practice, the basidiomycete fungus *Phanerochaete chrysosporium* is widely used for the degradation of organic

pollutants, as its exoenzymes are not substrate specific. The lignin peroxidase produced by the fungus is a key enzyme in lignin biodegradation. The fungus is capable of degrading various xenobiotics, but the specific role of the endo- and exoenzymes involved in its metabolism is not yet clear. Atrazine as a widely used pesticide and lignin peroxidase as an enzyme that can be used to degrade xenobiotics are good models to approach this topic.

A secondary metabolite produced by the fungus, veratryl alcohol (1,2-dimethoxybenzene), plays a mediating role in lignin peroxidase-catalysed reactions, and in its presence certain end products are produced in higher amounts. In the case of lignin peroxidase, 1,4-dimethoxybenzene is also known as a mediator. An important question is how the different mediators affect the structure of the enzyme and what substrates they act on.

The symmetrical triazine derivative atrazine (1-chloro-3-ethylamine-5-isopropylamine-2,4,6-triazine) is one of the most widely used herbicides in agriculture in the United States and Australia. It is mainly used for pre- and postemergence control of biennial weeds of maize and sugarcane. Metabolism of atrazine by extracellular enzymes is a relatively intensively researched area, but the results to date are rather controversial, and the mechanism of degradation by these enzymes is poorly understood. Many bacteria and fungi are capable of degrading atrazine to some extent, but the role of the specific enzymes and the reaction pathways are not fully understood.

Phanerochaete chrysosporium is also able to degrade the herbicide, but it was found that the purified form of lignin peroxidase does not react with atrazine. This would be a negative result at best and not worthy of attention, but lignin peroxidase is capable of degrading compounds that are structurally similar to atrazine. The

question is therefore what are the physical, chemical and structural parameters of the substrate (atrazine) and the enzyme (lignin peroxidase) that determine the degradability and accessibility of the compound.

It is known that the presence of lignin causes a significant reduction in the activity of cellulases. By modelling and random amino acid substitutions, more efficient cellulases could be produced and more could be learned about lignin-cellulase interactions.

In addition, molecular modelling methodologies could be applied to the production of biofuels (bioethanol, biodiesel), given the growing demand for these energy sources and the potential for their enzymatic production. By targeting enzymes and testing them under simulated conditions, time and financial resources could be saved on the way to producing enzymes with higher activity.

To get a more precise picture of the causes of the "incompatibility" between atrazine - lignin-peroxidase, an atomic-level description of the enzyme-substrate interactions is needed. To this end, we can use almost the entire methodology of molecular modelling and the following objectives are set for this work.

1. In order to use atrazine in molecular dynamics calculations, we will use quantum chemistry to calculate the parameters of the compound: non-bonding Lennard-Jones/van der Waals interactions, geometry, charges, bond lengths, bond angles and torsion angles.
2. Using the structure of lignin peroxidase and atrazine, we will investigate whether atrazine has an affinity for the active site of the enzyme by enzyme-ligand docking, and then further investigate the stability and behaviour of the resulting complex by molecular dynamics simulations. Quantitative data (binding energies, conformational changes of the amino

acids of the enzyme active site) will be collected for the complexes obtained during the simulations.

3. By evaluating the data obtained, we provide an atomic-level explanation for the fact that the purified form of lignin peroxidase cannot degrade atrazine (but compounds that are structurally similar to atrazine can).
4. We are using molecular modelling to try to map the function of biopolymerases. We will also try to find correlations between biopolymers and biopolymerases inhibitory mechanisms that may hinder the efficient processing of biomass. We also aim to review modelling-based improvements in biopolymerases that may play a role in more efficient biomass processing.
5. We will investigate the catalytic processes involved in the production of biofuels, with a view to how molecular modelling can be used to optimise them and make biofuel production more efficient.

MATERIALS AND METHODS

Introduction to the software used

This subsection briefly describes the applications that have produced the data and models needed for the different phases of the work. The different stages of the process, from the generation of quantum chemical (QM) data to the evaluation of molecular dynamics (MD) simulations, are represented by different software applications, which always use the data generated by the previous application as input, and are therefore discussed in the order of their use, since knowledge of the software applications makes the logic of the workflow transparent and clarifies the origin of the input data.

Avogadro. An easy-to-use modelling program capable of structural and energetic optimization of hand-drawn molecules. It was used to generate the 3D structure of the atrazine (ATZ) molecule, which was then transferred to Visual Molecular Dynamics.

Visual Molecular Dynamics (VMD). VMD is a molecular modelling and visualisation computer program developed primarily for viewing and analysing the results of MD simulations. It also includes tools for working with nucleic acid and protein sequence data and arbitrary graphical objects.

Force Field Toolkit (ffTK). An extension within the VMD that assists users in determining CHARMM-compatible force field parameters, including charges, bonds, angles and torsion angles. These tools can be accessed through the included graphical user interface, which greatly simplifies the setup and analysis of the underlying calculations. The set parameters are sent to QM software for quantum data determination and then back to ffTK, where a final

optimization is performed to make them compatible with CHARMM36 force fields.

Gaussian 09 (G09). Primarily used for QM modelling, this software determines charges, bonds, angles and torsion angles at the quantum level based on data entered with ffTK. Finally, ffTK makes the data compatible with the CHARMM36 force field. The data becomes suitable for use in MD simulations.

Nanoscale Molecular Dynamics (NAMD). A software for MD simulation using the Charm++ parallel programming model. It is known for its efficient parallelization and is often used to simulate systems with millions of atoms. It has no graphical interface and can be run from the command line. Its output files can be parsed with VMD.

Docking Server & Autodock Tools. Protein ligand docking software. The Docking Server is only available online, running on a central server, so it can perform computationally intensive docking faster and more accurately than software running on a traditional PC. The input structures (enzyme and ligand) can be derived from VMD, and its output files can also be displayed and analysed with VMD and prepared for MD simulation with NAMD.

Autodock is also a docking software, but we did not use it. Its graphical interface is Autodock Tools, which was used to display the data generated by the Docking Server.

Calculation of the atrazine parameters

The choice of force field is often based on what molecules the force field can handle in the first place. The CHARMM force field does not contain the data that can be applied additively to the ATZ molecule, so the force field needs to be extended to handle the new molecule. The optimisation procedure should be the same as that used to develop the CHARMM force field. This standard procedure has been used in this work.

The parameterization workflow includes the following steps and calculations: 1. Assign missing Lennard-Jones (LJ)/van der Waals (vdW) parameters, 2. Optimize geometry, 3. Calculate water interaction energy, 4. Optimize charges, 5. Calculate second derivative of potential energy, 6. Optimize bonds and angles, 7. Torsion test, 8. Optimize torsion angles. The parameterisation of the ATZ was performed using the Molefacture (v1.3) and ffTK (v1.1) extensions of the VMD (v1.9.3) software and NAMD (v2.12). The ffTK input QM data were calculated using G09 (Revision B.01). The 3D structure of the ATZ and the input file for the VMD were generated using Avogadro (v1.1.1). The charge representation of the ATZ was rendered using the Tachyon Parallel/Multiprocessor Ray Tracer extension of the VMD.

Assignment of missing Lennard-Jones/van der Waals parameters. Using Molefacture, the partial charges of non-polar hydrogens were fixed to +0.09 for compatibility with the CHARMM field. This is necessary because in the case of CHARMM, hydrogens are not always represented explicitly, but are rather treated as part of the non-hydrogen atom to which they are covalently bonded. For example, a methyl group can be treated as four independent atoms (one carbon and three hydrogens) or even as a single atom, and the

LJ/vdW parameters and charges can be modified to consider the omission of hydrogens. Although this approach can be applied to all hydrogens, it is typically used only for non-polar (aliphatic and aromatic) hydrogens; polar hydrogens, which are important for H-bond interactions, are present as separate atoms in the system.

After cross-checking the complete topology and parameter set of CHARMM36, 6 bonds, 11 angles, 11 torsion angles and 3 LJ/vdW parameters were missing, redundant data were not considered. LJ/vdW parameters were assigned by analogy, references were also obtained from CHARMM36 topology and parameter data.

Geometry optimisation. The G09 calculations were performed using the MP2/6-31G* theory-base set combination.

Calculation of water interaction energy. To determine the partial charges of the ATZ, it is necessary to characterise the water interaction sites of the molecule by two-dimensional optimisation.

All atoms with a partial positive charge can interact with the oxygen in the water and are therefore defined as donors, and all atoms with a partial negative charge can interact with the hydrogen in the water and are therefore defined as acceptors. The only exceptions are the sp^3 carbon atoms of the ethylamine and isopropyl groups, where the approach of water molecules is hindered by the hydrogens bonded to them, and these groups were therefore not included in the optimisation. The QM calculations were performed with the combination HF/6- 31G*.

Charge optimisation. The net charge of the ATZ was set to zero and the fixed charges of non-polar hydrogens were removed. The optimization routine was repeated until two iteration steps showed no difference, and the structure and topology file (PSF, TOP) were updated with the final charge values.

Calculating the second derivative of the potential energy. The calculation of the second derivative of the potential energy is an effective method for reconstructing the potential energy surface associated with distortions along bonds and angles.

The calculation has been applied to the geometry of the ATZ, which has already been optimised. The calculation of the frequencies was carried out at the MP2/6-31G* level, without using molecular symmetry in the calculation.

Optimisation of bonds and angles. Geometry Weight is set from 1.0 to 2.0, Angles-Eq. Deviation tolerance has been reduced from 10.0 to 5.0. The first modification represents a larger weighting related to how well the geometry optimized with conventional mechanics matches the QM-optimized geometry. The latter value determines the threshold below which the deviations no longer contribute to the objective function. A too tight Eq. Deviation thresholds tend to result in large force constants, so reasonable thresholds are critical for balancing the interaction between achieving sufficiently close optimized geometries and reproducing the surrounding potential energy surface. The iterations were performed with NAMD until the values converged to a minimum.

Torsion testing. Since ffTK excludes torsion angles ending in hydrogens when reading data directly from the parameter file in preparation, the 11 torsion angles found earlier were added manually. The scan intervals (+/-) were set to 180° and the step spacing to 10°, resulting in a total of 22 scans for QM calculations (MP2/6-31G*).

Optimisation of torsion angles. In the last phase of the parameterisation, the data provided by G09 is refined using the conventional NAMD mechanical method to fit the QM profile. For a complete description of the torsion angles, several expressions are

often needed that carry the appropriate periodicity (**n**) and phase shift (**σ**) for a force constant (**k**).

The input **n** and **σ** values are taken from the analogue torsion parameters of CHARMM36. The refinement algorithm was set to "downhill" and the tolerance value to 0.0001. The process was iterated until the mechanical profile was sufficiently close to the QM profile.

Interactions of atrazine and lignin peroxidase

The 3D model of the ATZ and the energy minimisation of the structure were carried out using the Avogadro program. The docking procedure was performed using the Docking Server web application and the results were analysed using Autodock Tools (v1.5.6). The enzyme structure used was obtained from the RCSB PDB database (PDB: 1B82) with a resolution of 1.8 Ångstroms (Å, 10^{-10} m). The enzyme was derived from the organism *Phanerodontia chrysosporium* and contained an R114A mutation that is not relevant for the present work. MD simulations of lignin peroxidase (LiP) were performed using NAMD software with the CHARMM36 force field. The CHARMM-compatible parameters of ATZ were calculated using the ffTK extension of the VMD software and the G09 software as described in the previous chapter. The images were rendered using the Tachyon Parallel/Multiprocessor Ray Tracer and Autodock Tools software.

Preparing for docking. The histidines were protonated (HSD form; histidine with hydrogen on delta nitrogen), the enzyme structure was solvated (with heme and two Ca^{2+} ions), the charges were neutralized and the NaCl concentration was adjusted to 0.15 mol dm^{-3} (physiological value). Energy minimization was run for 30 000 steps using the Conjugate Gradients method. This method is suitable for

solving linear (i.e. **Ax = b**) systems of equations; it is considered an iterative method due to rounding errors. The temperature was set to 288 K, 298 K and 308 K to produce three different conformers. MD simulations were then run on each minimized structure for 10 nanoseconds (ns, 10^{-9} s). We used the isotherm-isobaric NPT (constant particle number, pressure and temperature) sequence (which is well suited for hydrated complexes) with Periodic Boundary Conditions, which assume a practically infinite large system, since there is a vacuum around the macromolecule in solution, but the solution is considered as an elementary cell, thus minimizing surface effects, and Particle Mesh Ewald Electrostatics with a constant pressure of 1 bar. It is possible to perform simulations of biological molecules in the condensed phase such that virtually all non-bonding interactions can be retained. Indeed, in the past, to save resources, atom-atom non-bonding interactions beyond a certain distance were not considered in the calculations; the use of Particle Mesh Ewald makes this simplification unnecessary, i.e., there is no distance-based truncation of non-bonding interactions. From each of the 10 ns simulations, we selected 1-1 structure with the ligand access channel in the open state. These conformers were the input structures for the docking procedure.

Enzyme-ligand docking. The docking method is used to predict the affinity/orientation of a molecule (ligand) to another (biopolymer) molecule. In the case of a method based on geometric complementarity alone, the ligand and the protein are evaluated based on shape, surface area, solubility, and the conclusion whether it makes sense to dock the two structures together, whether they are compatible. Although the method is quite fast, it does not consider the dynamic changes of the components, the effects of conformational changes on the binding. In this simulation, the protein and ligand are

separated from each other, and once docking is initiated, the ligand must find the ideal conformational state within the target sequence. After each conformational change, the software calculates the energy of the current state. The method is much more computationally intensive, but this approach is less abstract than the geometric-based one.

For docking, only subunit A of LiP was used (the two subunits are identical) without solvent. The center of the Simulation Box was the amino acid His82, which is the entrance to the ligand access channel, and the size of the box was set to 20 20 20 Å. The partial atomic charges of LiP and ATZ were calculated using the conventional Gasteiger method, which considers the electronegativity of the atoms. For each LiP conformer, the total docking process consisted of 255 docking trajectories. During the process, the LiP structure remained rigid, only the ATZ was flexible, hence the need to select a conformer with an open state ligand channel earlier.

Molecular dynamics refinement and energy evaluation. The docking result with the highest frequency value was selected and the ATZ-LiP complex was solvated and minimized under the same conditions. The complex was refined by MD simulations with a length of 5 ns, using the same temperature and settings used in the generation of the conformer to study its stability. Binding free energies in the solvated phase were calculated using the NAMD Energy plugin. The bonding free energies are composed of vdW, electrostatic, polar and non-polar members. Finally, in addition to the 5 ns refinement, a 100 ns long MD simulation was performed with the canonical NVT set (constant particle number, volume and temperature) to gather slightly more information about the ATZ behavior on a larger time scale. The NVT was only used for a rough and fast simulation; it is not very common as it does not allow the system to relax properly and is not

representative of experimental conditions. Since the 5 ns simulation with the NPT sequence had by this time clarified how ATZ behaves in the ligand channel of the enzyme, this 100 ns NVT simulation was intended as a confirmation that could be run quickly and easily evaluated.

The basics of modelling, biopolymerases

The following search terms were used to search the databases: *polysaccharides*, *cellulose hemicellulose lignin biodegradation*, *modification*, *immobilization*, *industry*, *O-glucosyl hydrolase*, *cellulase enzyme modeling*, *molecular modeling*, *computation*, *cellulase modification*, *cellulose enzyme modification modeling*. The search was combined with OR and AND terms. Searches were performed up to 2020. Articles and their references were also reviewed. Articles found to be of interest for the problem highlighted above were categorised into: *Biodegradation of polysaccharides; Mechanical and physicochemical modification of substrates; Modification of cellulose, hemicellulose, lignin; Modification of enzymes; Immobilisation of enzymes; Molecular modelling of enzymes*. From these articles we have compiled our study to investigate the prevalence of model-based modification of extracellular enzymes and their rational design application over conventional methods.

Based on the data presented in this study, the following enzyme-ligand docking was performed in order to outline further research possibilities on extracellular enzymes: a previously studied *Cellulomonas* sp. CelB7 cellulase, a cellulose and a lignin fragment were docked separately, a xylanase (PDB: 1J01) was docked separately to its catalytic domain, a mannanase (PDB: 2X2Y) active

site, separately a mannan and a lignin part, and for LiP (PDB: 1B82), already used in the ATZ work, separately a lignin and a cellulase part. Both subunits of the enzyme were present during the docking. Docking was performed using the Patch Dock program.

Modelling in the production of biofuels

As an illustration, a model of a transesterification reaction obtained by energy minimization was constructed. There is perspective in the approach to produce biofuels in whole or in part with enzymes that have been previously optimised by molecular modelling and only then synthesised.

RESULTS AND DISCUSSIONS

The CHARMM36 compatible parameters of atrazine

The specific parameters read in by the CHARMM36 force field during MD calculations are contained in four files: PDB (atrazine.pdb), structure (atrazine.psf), topology (atrazine.top) and CHARMM36-compatible parameter file (par_charmm36_atrazine.par). These files are freely available.

Assignment of missing Lennard-Jones/van der Waals parameters. The cross-checking of the parameters resulted in 3 LJ/vdW parameters missing: a C2 (sp^2 hybrid carbon), C3 (sp^3 hybrid carbon) and a Npl (sp^2 hybrid nitrogen bonded to three other atoms, formally neutral) type. The assigned parameters were obtained from the topological and parameter data of the CHARMM36 force field.

Type C2 is designated as an aromatic carbon in a 6-membered ring between 2 or 3 nitrogens, one of which is double bonded; type C3 is designated as a carbon in the sp^3 hybrid state and type Npl as a neutral nitrogen, with ATZ as an aromatic amine.

Charge optimisation. The charge values of the ATZ atoms can be found in the PSF file. The carbon atoms and N2 of the 1,3,5-triazine ring have the highest absolute charge values.

Optimisation of torsion angles. The Root Mean Square Deviation (RMSD) after the first optimization was 0.446; after the final (5th) optimization it was 0.225. The RMSD is a commonly used measure of the differences between the model predicted values and the observed values; it is the square root of the second sample point of the differences between the predicted and observed values, or the root mean square of these differences. The result of the final refinement shows a good fit to the target QM data.

Dynamics of the atrazine-lignin-peroxidase complex

Evaluation of enzyme-ligand docking. For all conformers, the binding of ATZ to the LiP target site is thermodynamically favourable, but the frequency value is only 16% and 10% at 288 K and 308 K, respectively. For the LiP conformer generated at 298 K, the frequency was 53%, so this was the input structure for the 5 ns refinement process to study the stability of the complex and the interaction of the ATZ with the amino acids of the ligand channel.

The protein-ligand interactions between amino acids of the interacting side chains of ATZ and LiP can be classified into the following categories: H-bonding, polar, hydrophobic and other. The Cα carbonyl group of Asp183 interacts with the isopropylamine nitrogen of ATZ through H-bonds. There is also an unspecified link between the oxygen of the Cα carbonyl group and the number 1 nitrogen of the s-triazine ring. The distance between these two atoms is only 2.68 Å. Since there is no direct link between nitrogen and oxygen and both atoms have a negative partial atomic charge, this state is energetically unfavourable and cannot be maintained in a fully dynamic system.

Structural stability of the docking complex. The Cα RMSD values of the ligand channel amino acids and the minimization of ATZ showed that the initial structure, the [298 K] docking complex, was not in an optimal state. The 30 000 steps were sufficient for the complex to reach the relaxed state. The minimum RMSD value was 0.040 Å and the maximum value was 1.490 Å. The system reached the maximum value in 29.55 picoseconds (ps, 10^{-12} s) and after about 20 ps the system was in a relaxed state, but conclusions can only be drawn from the results obtained by MD.

Ligand channel behaviour in the presence of atrazine. Fluctuation and deformation of LiP ligand channel amino acids in aqueous solution is a known phenomenon. The channel shows remarkable distortions with respect to the crystal structure even during a 150 ps MD simulation. Such rapid conformational changes of the amino acids of the ligand channel are typical of LiP when no potential substrate or other small molecule is present in the vicinity of the channel. In the presence of certain molecules (e.g., veratryl alcohol, which is a natural substrate of the LiP enzyme), the observed phenomena may vary according to the different interactions between the ligand and the ligand channel amino acids.

In the presence of the ATZ, the channel fluctuation decreases over a short period of time, and the last 150 ps of the refinement (5 ns) MD simulation shows significantly lower RMSD values compared to data from different time scales. Since rapid conformational changes of the amino acids in the ligand channel were clearly detectable in the absence of ATZ, the first 150 ps of the MD simulation without ATZ was used as a reference. In the presence of ATZ, the first 150 ps of refinement showed little difference, and the values in the 151-300 ps interval were clearly lower than the reference over the entire range. After about 80 ps these changes are mostly below the values of the first 150 ps. Visualization of the end-state of the ATZ-LiP complex confirms that the ATZ molecule does not reach the active site containing heme through the channel because the ligand is surrounded by amino acids of the slightly fluctuating ligand channel. This causes stoichiometric obstruction of the active site from the ligand. Shrinkage of the ligand channel amino acids occurs in a short time (ps scale) in the presence of ATZ, and this mild fluctuation is stably maintained over 4-5 ns. The relative positions of ATZ and the heme skeleton show that only the hydrophobic isopropylamine group of

ATZ and some side chains of the heme are oriented towards each other.

Substrates can be oxidized by electron transfer mediated by the LiP surface amino acid Trp171, a hypothesis based on the observation that mutation of this amino acid led to loss of enzymatic activity against veratryl alcohol oxidation. However, the mutants were still able to oxidize two substrates with lower redox potentials compared to veratryl alcohol, as was ATZ. In the absence of Trp171, LiP can still catalyse the reactions, but direct contact between the heme rim and the ligand through the ligand channel is essential. The exact function of Trp171 is not fully understood, as it is a rigid region of the enzyme, and the W171A mutation does not cause relevant structural changes. Based on its rigid character, it is unlikely that the Trp171 domain could function as a competitive ligand channel.

The energy data suggest that the reduced fluctuation of amino acids in the ligand channel in the presence of ATZ is due to favourable interactions and stable contact between the amino acids and the ligand. The energy results for the last 150 ps of the 5 ns simulation show that the complex is in an equilibrium state, with negative energy values over the entire range.

Calculated binding free energies ($\mathbf{\Delta G_{bind}}$) confirm that ATZ can incorporate into the ligand channel environment, but a slowly increasing trend of binding free energies is observed in the last 150 ps of the MD. In addition, $\mathbf{\Delta G_{bind}}$ between ATZ and the heme skeleton shows a positive value at some time points; these results suggest an energetically unfavourable and transient relationship.

An additional MD simulation of 100 ns shows that after 5.6-5.7 ns, the connections between the ATZ and the ligand channel start to break down. The RMSD values clearly indicate that ligand integration is only a transient phenomenon. The partial atomic charges of ATZ

create a temporary ATZ-channel complex, but the kinetic energy of the enzyme, together with the unfavourable interactions from some heme side-chain atoms interacting with the isopropylamine group of ATZ, overcomes these effects in a short time, and therefore ATZ cannot reach the heme edge. The ATZ can be displaced from the channel and the enzyme returns to its original state by fluctuations in the amino acids of the ligand channel and periodic alternation of open and closed states. By 100 ns, the secondary structure of the enzyme has not changed significantly.

Three amino acids of the ligand channel have the most significant effect on ATZ unavailability: Phe148, Asp183 and Gln222. The phenyl group of Phe148 interacts with the delocalized electron system of the s-triazine ring. The carbon atom of the carboxyl group of Asp183 has a partial charge of 0.62 and interacts with the N1 atom of the s-triazine ring, which has a partial charge of -0.759. The hydrogens of the amino group of Gln222 have a partial charge of 0.32 and 0.30, respectively. These atoms interact with the N1, N2, N3 and Cl atoms of the s-triazine ring, which have partial charges of -0.759, -0.244, -0.19 and -0.26, respectively.

It is important to note that the previously calculated partial atomic charges of the carbon atoms in the delocalized s-triazine ring of ATZ and nitrogen 1 show significantly higher absolute values compared to the other atoms of the compound. The position of the incorporated ATZ suggests that these partial charges generated by the delocalized electron system are possible limiting factors and may be responsible for the inaccessibility of the ligand.

Biopolymerases and molecular modelling

Cellulases. Biopolymers in the plant cell wall can be divided into three main groups: cellulose, hemicellulose and lignin. It is the tight coupling of these that makes biomass utilisation so difficult.

The strain of bacteria we used (*Cellulomonas* sp.) was chosen because it contains a good proportion of the enzymes required for the degradation of biopolymers, which allowed us to follow the 3D structure and similarities of the different enzymes more easily. We first used the catalytic domain and cellulase binding domain (CBM) of a previously studied cellulase for modelling.

The two types of domain are fundamentally different. The catalytic domain is dominated by α-helix parts, whereas the CBM is dominated by the β-strain structure. This is generally true for the binding and catalytic domains of other O-glycosyl hydrolases. Many other enzymes besides cellulases are involved in the degradation of cellulose, including, but not limited to, various β-glucanases, β-glucosidases and cellobiohydrolases. Cellulases are very important in the degradation of biomass; the CBM binds the enzyme to the cellulose fibres, loosening them so that the active site of the catalytic domain can carry out the catalysis.

Due to the activity of cellulases, the resulting smaller oligomers often inhibit the enzyme's further function by binding to the enzyme, thus inhibiting its function.

Higher concentrations of lignin inhibit cellulase activity, a fact that has been demonstrated in several cases. We have demonstrated by modelling that lignin can bind to the catalytic domain of *Cellulomonas* sp. CelB7 cellulase in some cases. The docking results obtained showed this coupling to a lesser extent, so this may confirm that lignin has little inhibition of cellulase function or that a high

concentration is required for lignin to significantly inhibit the cellulase enzyme.

Xylanases. Another important group of enzymes involved in the degradation of biopolymers are the various xylanases. Xylanases degrade the xylan part of hemicellulose in different ways. The xyloglucan side chains in hemicellulose are a significant part of the hemicellulose, so xylanases are important enzymes in the degradation of biomass.

There are many similarities between cellulases and xylanases, and this is reflected in their properties. Our modelling of this is very similar to that of the cellulase complex. Perhaps this explains the fact that there are several cellulase enzymes that have xylanase activity. The possibility of cooperation between cellulolytic and hemicellulolytic enzymes has been investigated and mutual synergism has been demonstrated.

The catalytic domain of *Cellulomonas fimi* xylanase shows a strong affinity for lignin. This was also demonstrated by enzyme-ligand docking. The model shows a high similarity to the binding of cellulose to lignin, which is not surprising since many cellulases also have xylanase activity.

The link can be partly explained by the adsorption of enzymes (cellulase and xylanase) onto the lignin surface, but the detailed mechanisms of binding are still poorly understood. The mechanisms underlying lignin-enzyme interactions remain unclear. Non-specific adsorption of cellulases and xylanases to lignin prevents enzymatic conversion of biomass. Calculations show that negatively charged surface cellulase and xylanase can reduce inhibition by lignin. It is possible to produce highly active cellulases and xylanases that resist lignin-mediated inactivation, although further work is needed to understand this problem.

Mannanases. Mannans are also hemicelluloses. Mannanases are involved in the degradation of the mannopyranose oligomers attached to the rhamnogalacturonan moieties of hemicellulose as side chains. Mannanases can be found in both monomeric and dimeric forms. The mannanase-mannose docking complex can be of several types, depending on whether the mannan is attached to the monomeric enzyme or the dimeric enzyme. The mannanase enzyme contains both a catalytic and a substrate-binding moiety. The docking complex, like the other enzyme-substrate interactions discussed so far, is stable and may be capable of substrate degradation.

A synergistic effect between xylanase and mannanase has been described. Nevertheless, the two enzymes have little similarity, so this can hardly explain their cooperative action.

The possibilities are not exhausted with hemicelluloses and the enzymes used to break them down. In addition to the ones discussed so far, hemicellulose also contains, but not exhaustively, glucopyranoside, galactopyranoside, rhamnopyranoside and arabino-furanoside side chains. Gelatinous pectin is a heteropolysaccharide composed mainly of galacturonic acid - it is linked to hemicelluloses at several points. The additional enzymes involved in the degradation of these side chains are also very diverse. The diversity of hemicelluloses and hemicellulases and their modelling will be a future task that may be challenging.

There are almost no references in the literature on whether lignin-mannanase interactions exist and what their effects are. It has been possible to model this relationship. The lignin moiety is well aligned with the active site of the mannanase, although the nature of the interaction is not yet fully understood and further studies are needed.

It has been reported that water-soluble (with low molecular weight) lignin enhances the rate of enzymatic digestibility, but the

catalytic mechanism of lignin-enzyme interaction remains elusive. This may be aided by molecular modelling, which may help to explain the phenomenon. To mitigate the negative effects of lignin, extensive research has been carried out to elucidate the basic mechanisms of lignin enzyme interactions in order to develop technologies that can overcome the negative effects of lignin on enzymatic hydrolysis.

Lignin peroxidases. The third group of biopolymers is lignin. LiP is produced by fungi that cause white rot (white sooty mold). It catalyses the oxidation of veratryl alcohol and the degradation of lignin, and inhibits the depolymerisation of lignin. (Note: brown rot fungi degrade cellulose and hemicellulose in wood. They break down cellulose and hemicellulose with hydrogen peroxide (H_2O_2), so no saccharides are produced, no utilization, only oxidation).

Both monomers contain a heme cofactor, so these enzymes are called heme peroxidases, which require H_2O_2 as oxidant.

The ligand fits between the two subunits. The possibility arises that the presence of both subunits (with the same sequence) may be necessary for lignin recognition and degradation - further measurements are needed to determine this.

Under extremely acidic conditions, LiP does not catalyse properly. However, if the heme of LiP is replaced by the heme in the manganese peroxidase, the catalytic efficiency is significantly increased.

Once we had studied the effect of lignin on cellulase, we also tried to model the effect of cellulose on LiP. We were surprised to find that there is some relationship between cellulose and LiP. In most cases, cellulose bound to the non-catalytic site of the LiP homodimer, but in some cases, cellulose bound to the enzyme near the active site. Even if cellulose is not tightly bound to the active site, it may be able to block lignin binding partially or completely.

This phenomenon is completely new, no similar result has been described. It was made possible by molecular modelling. Of course, it must be verified in practice, many experiments have to be done to describe it accurately. In any case, modelling shows the new potential of this method.

One of our later research aims is to find out whether random amino acid substitutions cause different structures, and whether the different structures cause differences in enzyme activity and substrate binding capacity, and whether we can explain these at the molecular level. In addition, it would be interesting to investigate why the presence of lignin causes a significant reduction in the activity of cellulases (and other hydrolases). If we could succeed in generating more efficient cellulases by random amino acid substitutions, it would be possible to learn more about lignin-cellulase interactions.

Exploring new opportunities. The recycling of biomass is now essential not only because fossil energy sources are gradually being depleted, but also because of the need to reduce pollution caused by the growing use of energy. This article aims to review the results of currently used plant biomass processing methods. We also aimed to review currently unused but published methods. We have explored new methods for biomass recycling and the potential of newer methods using enzyme modelling. The results of this review are striking in almost all areas. Advances have been made in the pre-treatment of biomass and in the diversity and application of the enzymes used. However, very little progress has been made in advances based on molecular modelling. There has also been minimal progress in the modification of existing enzymes to adapt to changed functionality and environmental conditions during biomass processing. There are hardly any publications that use molecular modelling techniques to modify and improve enzyme function and

adapt enzymes to different environmental conditions. We believe that modern computational, biochemical and biotechnological methods could be used to design more efficient enzymes suitable for biomass processing.

Together with the methods that have been recently investigated and applied, we have explored the latest possibilities and results of the utilisation of plant structural biomass. Progress has been made in all areas, and this field continues to evolve. New physical and chemical methods have been developed for the pre-treatment of biomass, which may lead to more efficient extraction of biomass materials. A wide range of enzymes in microorganisms have been isolated and used in biomass processing. The mechanism of action of enzymes has been studied and new and more efficient methods have been developed. In our opinion, there are some areas where less progress has been made recently. These include the modification of existing enzyme properties by rational design based on structural modelling, MD simulations, enzyme-substrate interactions and virtual mutagenesis.

There is untapped and huge potential for producing modified enzymes that work more efficiently under the industrial conditions used for biomass processing. Rational modification of enzymes can significantly increase the efficiency of biomass conversion, enzymes or processes. Physical, chemical, biochemical, biotechnological and molecular biological methods are available for modifying the protein side chains of enzymes to make more efficient use of biomass, which are more appropriate to the techniques used.

With the widespread application of these methods, further progress in this field can be expected and can only be supported by in silico modelling, virtual modification of molecules, molecular simulations to study the function of enzymes under different environmental and/or

industrial conditions. This would help to improve the industrial biomass utilisation methods used so far without changing the method, but only modifying the enzymes to work more efficiently under the given conditions, thus increasing productivity and reducing costs.

There are many ways to improve the enzymes currently used to process biomass. Computer-aided design and modelling techniques can be used to modify enzymes to work more efficiently under specific physico-chemical conditions. By studying enzyme function and catalytic conditions, it is possible to intervene in chemical reactions by modifying the enzyme structure to better suit the actual operation. This requires knowledge of the enzymatic processes involved and the use of computer modelling and design to modify the enzymes' function for better use.

Enzymes that degrade polysaccharides are often linked to glycosylation, N- and O-linked glycans, whose role is only partially understood. Glycans can affect critical properties of enzymes: N-glycosylation improves thermal and proteolytic stability, O-glycosylation improves CBM-binding affinity and stability in addition to proteolytic stability, but their presence does not necessarily affect catalytic activity. Modelling glycosylated cellulases may improve our knowledge of the different functions of glycans.

A glycosylated cellulase was described in which mainly galactose disaccharides were found. This glycosylation dramatically affected the hydrolysis of insoluble substrates, proteolytic and thermal stability, and proved necessary for this enzyme to function in harsh environments, including industrial environments.

The modelled methanolysis

For lipase enzymes catalysing the transesterification reaction, the possibility of molecular modelling-based developments also arises. For example, triacylglycerol lipase is found in animals, plants, fungi and bacteria. It hydrolyses triglycerides to diglycerides and then to monoglycerides and free fatty acids. The enzyme is highly soluble in water and acts on the surface of oil droplets. Access to the active site is controlled by the opening of a so-called lid which, when closed, hides the hydrophobic surface surrounding the active site. The lid opens when the enzyme comes into contact with the oil-water interface (interfacial activation).

It would be possible, as with cellulases and ligninases, to substitute amino acids at the relevant sites of the enzyme and then test the resulting (mutant) models under varying environmental conditions (pH, temperature). If the results of the simulations show that the modifications result in one or more enzyme variants with higher activity, which can catalyse the reactions more rapidly under more easily producible conditions, we could then target these enzyme populations for synthesis.

CONCLUSIONS AND SUGGESTIONS

Quantum chemical parameters of atrazine. Most of the publications refer to parameters from the Param Chem web server, a resource that allows access to initial parameters based on analogy alone. The problem is that these results contain a so-called penalty score, since the loaded ligands are only compared to the CHARMM general force field (CgenFF), and therefore further calculations would be needed to refine the parameters in the new chemical context. In this work, the data set based on the analogy is limited to the three non-binding parameters, the ATZ has been cross-checked with all topologies and parameter data of the CHARMM36 force field (including CgenFF), and the other output figures are the result of precise optimization methods and quantum chemical calculations.

The result of the presented workflow is the complete set of parameters needed to simulate the ATZ with MD.

The incompatibility of lignin peroxidase and atrazine. The enzyme is capable of degrading various xenobiotics, but the enzyme-substrate relationships required for catalysis are not fully understood. Since LiP is not able to degrade ATZ, describing the interactions at the molecular level may provide a better understanding of the ligand channel structure-function relationship and the factors that determine or limit the availability of each ligand. The enzyme-ligand docking results suggest that ATZ may find energetically favourable positions in the vicinity of amino acids in the ligand channel, although the mathematical probability of forming these complexes is not significant (10%, 16% and 53% frequency values at three different temperatures). In a 5 ns MD simulation of the ATZ-LiP complex, ATZ integrated with amino acids with the best docking result (53% abundance at 298 K) and failed to cross the channel. This

phenomenon is somewhat contradictory to the fact that small substrates have free access to the heme when the channel is in the open state conformation. Partial atomic charges on the delocalized heterocyclic s-triazine ring of ATZ can form a stable enzyme-ligand complex and may account for the inaccessibility of ATZ.

Since ATZ can bind to amino acids that provide access to substrates towards the active site, an important question is whether this binding transiently affects enzyme activity. If ATZ is able to act as an inhibitor for each substrate, such experimental results would demonstrate that our in silico approach correctly describes the phenomenon and is therefore the area for further investigation. Our results suggest that the kinetic energy of the atoms may be sufficient to remove ATZ from the channel, a possible phenomenon that could be the subject of another investigation. Moreover, the full function of the ligand channel could be mapped and explained. Again, we note that LiP has two oxidation sites, but the relationship between them is not yet clear.

Targeted modification of enzymes to process biomass more efficiently. It can be concluded that very few papers address the design of extracellular enzymes by molecular modelling. Although the methods are available, this field is far from being well established. A priority for further research is to investigate whether it is possible to generate more efficient cellulases by random amino acid substitutions and, in this context, to describe as precisely as possible the specific mechanism of lignin as an inhibitor of cellulases that induces enzyme inhibition. Indeed, based on the docking work carried out so far, it is well established that lignin can inhibit cellulases and xylanases, and there also appears to be a lignin-mannanase interaction, and even a cellulose-lignin-peroxidase interaction - the

latter is also an interesting area for future research, as it may also influence enzyme activity.

Biofuels production. Although it is still at a very early stage, the potential to use the results of molecular modelling to produce bioethanol and biodiesel more efficiently is clear. These biofuels can also be synthesised by enzymes, but their use is still costly and slow. With well-chosen methods, it would be possible to design more efficient and faster enzymes in a targeted way, test them under simulated conditions and then express and produce only the ones best suited to our needs.

NEW SCIENTIFIC RESULTS

1. Quantum chemical calculations have allowed us to determine the partial charges, bond lengths and angles, and torsion angles of the photosynthetic herbicide atrazine. The obtained values were then translated back into the language of conventional mechanics and, for the first time, we have produced force-field compatible parameters for atrazine CHARMM36, which are much more accurate than the conventional parameters based on analogy. Atrazine can now be used in molecular dynamics simulations to study protein-ligand, enzyme-ligand and other interactions at the atomic level.

2. By enzyme-ligand docking, we found that the binding of atrazine to the ligand channel of lignin-peroxidase is thermodynamically favourable, but somewhat temperature-dependent and that only a minority of the possible conformations of atrazine in the open state of the channel represent the correct orientation. Using molecular dynamics simulations and calculations, we have shown for the first time that the presence of atrazine in the ligand channel of lignin peroxidase significantly reduces the amino acid fluctuations. These fluctuations are essential for the transport of small potential substrates to the active site. This explains the inaccessibility of atrazine to the active site of lignin peroxidase. By energetically evaluating the interactions of the atrazine-lignin-peroxidase complex, we have described the interactions between the enzyme and atrazine that explain the inaccessibility. We found that the delocalized electron arrangement of the s-triazine ring of atrazine causes a drastic decrease in the lignin peroxidase channel fluctuations. This explains the decrease in the channel functionality.

3. Using molecular modelling, we have demonstrated that lignin at higher concentrations can indeed inhibit cellulases and xylanases. By modelling, we were able to establish for the first time that a lignin-mannanase relationship exists, so that enzyme inhibition may also exist in this case. Using modelling, we have described for the first time that the cellulose oligomer may be able to inhibit lignin peroxidase under certain conditions.

4. We have published a comprehensive review of state-of-the-art biomass utilisation techniques. We have found that methodologies based on molecular modelling and enzyme modification are still scarce.

LIST OF PUBLICATIONS

Publications in scientific journals in the topic of the dissertation

1. FÜLÖP, L.; ECKER, J. (2020): An overview of biomass conversion: exploring new opportunities. In: *PeerJ*, 8: e9586. DOI: https://doi.org/10.7717%2Fpeerj.9586
2. ECKER, J.; FÜLÖP, L. (2018): Lignin peroxidase ligand access channel dysfunction in the presence of atrazine. In: *Scientific Reports*, 8: 5989. DOI: https://doi.org/10.1038%2Fs41598-018-24478-w
3. ECKER, J. (2016): Generating CHARMM-compatible force field parameters for atrazine. In: *Journal of Universal Science Online*, 3: (1) 1-8. p. DOI: https://doi.org/10.17202%2FJUSO.2016.3.1

Article under review in a scientific journal on the topic of the thesis

1. DANCS, G.; KAKUCSKA, G.; DOBRÁNYI, SZ.; ECKER, J.; FÜLÖP, L. (2022): Efficient method for the determination of the neutral lipid content of oil-producing microalgae strains required for biodiesel. In: *Fuel*, Under review.

Publication in scientific journals outside the topic of the dissertation

1. ECKER, J.; FÜLÖP, L. (2014): Molecular modeling of DDT's and its major metabolites' adsorption in the interlaminar space of montmorillonite. In: *Journal of Universal Science Online*, 1: (1) 12-19. p. DOI: https://doi.org/10.17202%2FJUSO.2014.1.12

Conference publications

1. ECKER, J., FÜLÖP, L. (2017): Dysfunction of the lignin peroxidase ligand channel in the presence of atrazine. In: REMÉNYI, A.; DORMÁN, GY. (Szerk.), *ECBS 2017 5th European Chemical Biology Symposium: Program and book of Abstracts*, Budapest, Magyarország: Magyar Kémikusok Egyesülete (MKE). 43. p. http://www.ecbs2017.eu/
2. ECKER, J.; FÜLÖP, L. (2017): Dysfunction of the lignin peroxidase ligand channel in the presence of atrazine. *5th European Chemical Biology Symposium - ECBS 2017*, 2-4 July 2017 - Budapest, Hungary. Hungarian Chemical Society. Poszter prezentáció. Megjelenés: Magyarország. http://www.ecbs2017.eu/
3. ECKER, J.; FÜLÖP, L. (2017): CHARMM-kompatibilis atrazin paraméterek számítása kvantumkémiai módszerekkel. In: Magyar Kémikusok Egyesülete Vegyészkonferencia 2017, *Program és előadásösszefoglalók*, Budapest, Magyarország: Magyar Kémikusok Egyesülete (MKE). 49. p. https://docplayer.hu/111080154-Vegyeszkonferencia-2017.html
4. ECKER, J.; FÜLÖP, L. (2017): CHARMM-kompatibilis atrazin paraméterek számítása kvantumkémiai módszerekkel. *Vegyészkonferencia 2017*, Hajdúszoboszló, 2017. június 19-21. Magyar Kémikusok Egyesülete (MKE). Poszter prezentáció. Megjelenés: Magyarország. https://www.mke.org.hu/vegykonf2017/
5. ECKER, J.; FÜLÖP, L. (2017): Lignin-peroxidáz ligandum csatornájának molekuladinamikai vizsgálata atrazin jelenlétében. In: Magyar Kémikusok Egyesülete Vegyészkonferencia 2017, *Program és előadásösszefoglalók*, Budapest, Magyarország: Magyar Kémikusok Egyesülete (MKE). 50. p. https://docplayer.hu/111080154-Vegyeszkonferencia-2017.html

6. ECKER, J.; FÜLÖP, L. (2017): Lignin-peroxidáz ligandum csatornájának molekuladinamikai vizsgálata atrazin jelenlétében. *Vegyészkonferencia 2017*, Hajdúszoboszló, 2017. június 19-21. Magyar Kémikusok Egyesülete (MKE). Poszter prezentáció. Megjelenés: Magyarország. https://www.mke.org.hu/vegykonf2017/
7. ECKER, J.; FÜLÖP, L. (2016): Extracelluláris lignin-peroxidáz atrazin bontásának modellezése. *MKE Szerves- és Gyógyszerkémiai Szakosztályának QSAR és Modellezési Szakcsoportja és az MTA Szegedi Akadémiai Bizottságának Kemometria és Molekulamodellezés Munkabizottsága által szervezett KeMoMo-QSAR 2016 szimpózium*, MTA MAB Székház (Miskolc, Erzsébet tér 3.) 2016. május 12.- május 13. [Előadás]. Megjelenés: Magyarország. http://www.chemicro.hu/QSAR/20160512.html

Additional publications in the topic of the dissertation

1. ECKER, J.; FÜLÖP, L. (2018): Structure of a lignin peroxidase ligand access channel-atrazine complex after molecular dynamics simulation. *Mendeley Data*. DOI: https://doi.org/10.17632%2Fb2wn6fdcj4.2
2. ECKER, J.; FÜLÖP, L. (2018): Structure of a lignin peroxidase ligand access channel-atrazine docking complex. *Mendeley Data*. DOI: https://doi.org/10.17632%2Fdvjms8448j.2