

**Ph.D. THESIS**

**Sarolta Pilbák**

**Study and modelling of  
protein-ligand interactions**

**Supervisor: Dr. László Poppe**



**Budapest University of Technology and Economics**  
Department of Organic Chemistry and Technology

Budapest

2008

# 1. INTRODUCTION

Proteins, the building blocks of our organism, play a part in the build-up of heart, muscles and other tissues, but are also important constituents of the enzymes, hormones and materials responsible for the immune system. Proteins play an important role in animals and plants too. The important biological role is supported by the fact that they participate in every process taking place in cells.

By the recently modern amino acid sequence determination techniques, the number of protein sequences deposited in databanks are growing rapidly. While the number of known sequences is exponentially growing, functional information only for a few proteins can be established. Therefore the most important criteria in protein structure determination is the structure based function assignment.

There are many experimental techniques for protein structure determination. The most outstanding of these is the X-ray crystallography, for which protein crystallization is required. Enzymes are usually soluble proteins, which can be easily crystallized. But for the transmembrane proteins the crystal growing is difficult, since the natural medium of the proteins dissolves and the protein will unlikely adopt again his original structure. In the absence of experimentally determined structures, proteins can be modelled.

The genes of our organism encode in 20% membrane proteins, which have a significant role in cell communication. The G-protein coupled receptors (GPCRs) are the most important transmembrane proteins, since they assist in regulation of many physiological processes. The sensory receptors, which transmit the light energy, have a prominent role. The opsins, for example, convert the light to cell responses during a photoisomerization reaction. The melanopsin is the photopigment of the retinal ganglion cells involved in the entrainment of the circadian clock, while rhodopsin is the photopigment of the rod cells responsible for the

night vision.<sup>1</sup> My aim was to model the 3D-structure of the melanopsin. With such a prediction and analyzing the ligand binding, melanopsin and rhodopsin can be compared by the aspect of functional physiology.

Enzymes are formed in living cells and catalyze well-defined chemical reactions (biocatalysts). L-phenylalanine ammonia-lyase (PAL) is the starting point of the phenylpropanoid pathway in plants and yeast catalyzing the ammonia elimination from L-phenylalanine. A similar reaction is catalyzed by the histidine ammonia-lyase (HAL) found in humans and bacteria and in the bacterial tyrosine ammonia-lyase (TAL).<sup>2</sup> The reaction is due to a post-translationally formed prosthetic group, named MIO.<sup>3</sup> Two significantly different mechanisms were suggested for these reactions.<sup>2</sup> The first one postulated an *N*-MIO intermediate between the substrate amine group and MIO methylene. From here three pathways are possible: E<sub>1</sub>cB, E<sub>1</sub> or E<sub>2</sub>. Later, the E<sub>1</sub> and E<sub>2</sub> routes were excluded, and in the E<sub>1</sub>cB case the abstraction of the corresponding  $\beta$ -proton remained an open question. To solve the problem, Rétey suggested an alternative mechanism involving a Friedel-Crafts (FC) type attack at the aromatic ring of the substrates by the electrophilic prosthetic group (MIO), which can acidify properly the  $\beta$ -position. The PAL activity of the *R*sTAL His89Phe mutant<sup>4</sup> proved that the TAL and PAL reactions have the same mechanism. A second goal during my work was the study of the ammonia-lyase structures and their reaction mechanism by modelling and computations within the enzymic environment.

Enzyme catalyzed processes are widely used for synthetic purposes, among others for synthesis of enantiopure compounds. *Candida antarctica* Lipase B (CaLB) is a remarkable

<sup>1</sup> Erren, T. C., Erren, M., Lerchl, A., Meyer-Rochow, V. B., *Naturwissenschaften*, **2008**, *95*, 273-279.

<sup>2</sup> Poppe, L., Rétey, J., *Angew. Chem. Int. Ed.*, **2005**, *44*, 3668-3688.

<sup>3</sup> Schwede, T. F., Rétey, J., Schulz, G. E., *Biochemistry* **1999**, *38*, 5355-5361.

<sup>4</sup> Louie, G. V., Bowman, M. E., Moffitt, M. C., Baiga, T. J., Moore, B. S., Noel, J. P., *Chem. Biol.* **2006**, *13*, 1327-1338.

biocatalyst in the stereoselective esterification of secondary alcohols.<sup>5</sup> The prediction of the stereoselectivity by calculations within the enzymic environment constitutes a real challenge, which is a third goal during my work. But the task is getting difficult, because in case of a racemic alcohol, not only the two enantiomers should be calculated. According to the mechanism of the lipases, a tetrahedral intermediate (THI) is formed between the acyl enzyme and the alcohol, while a new asymmetry centre is formed.<sup>6</sup>

Traditionally, proteins and other large biological molecules have been out of the reach of electronic structure methods. The QM/MM computational technique allows the study of large molecules by defining two or three layers within the structure that are treated at different levels of accuracy. In a typical two-layer approach, the active site is treated using an electronic structure method, whilst the rest of the system is modelled with molecular mechanics.<sup>7</sup>

<sup>5</sup> (a) Paizs, C., Tosa, M., Bódai, V., Szakács, G., Kmech, I., Simándi, B., Majdik, C., Novák, L., Irimie, F.-D., Poppe, L., *Tetrahedron: Asymmetry*, **2003**, *14*, 1943-1949. (b) Szatzker, G., Móczár, I., Kolonits, P., Novák, L., Huszthy, P., Poppe, L., *Tetrahedron: Asymmetry*, **2004**, *15*, 2483-2490.

<sup>6</sup> Haeffner, F., Norin, T., Hult, K., *Biophys. J.*, **1998**, *74*, 1251-1262.

<sup>7</sup> Vreven, T., Morokuma, K., *Theor. Chem. Acc.* **2003**, *109*, 125-132.

## **2. CALCULATION METHODS**

### ***2.1. Homology modelling***

The homology models were built by the aid of automated homology modelling servers (Swiss-Model, Robetta).

### ***2.2. Molecular mechanics and molecular dynamics***

The resulted models were relaxed in molecular mechanics force field (Hyperchem, protein part: amber99, non-protein part: MM+), where I carried out the molecular dynamics (MD) simulations too. MD simulations were performed on the loop (outside the transmembrane) region of the melanopsin model and in case of PAL enzymes on the catalytically essential Tyr containing loop.

### ***2.3. Conformational analysis within the rigid enzyme environment***

The possible arrangement of the ligands bound to the enzyme was studied by conformational analysis within the rigid enzymic environment. For this, during the reaction a covalent (ammonia-lyases: MIO-ligand, CaLB: tetrahedral) intermediate formation is needed.

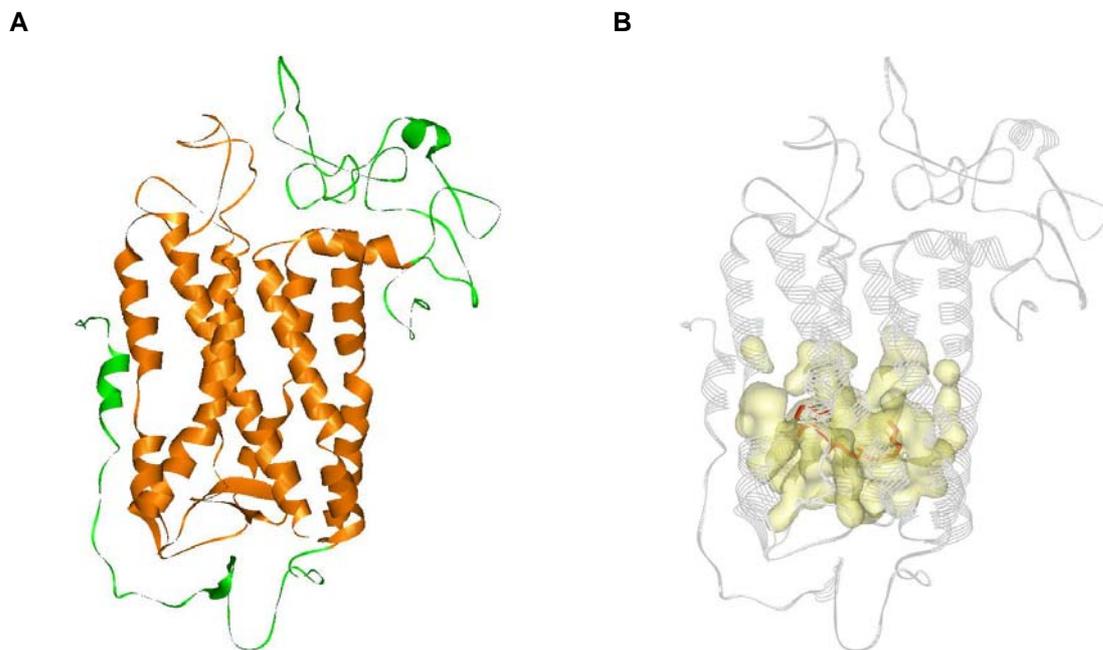
### ***2.4. QM/MM calculations***

The QM/MM (combined quantum mechanics/molecular mechanics) calculations were performed on a  $\sim 10$  Å sphere cut from the enzyme structure. The QM part contains the ligand and the catalytically important amino acids, whereas the rest of residues remained in the fixed MM part (calculations of ammonia-lyases, CaLB)

### 3. RESULTS

#### 3.1. *The first 3D-structure of the melanopsin protein*

I found good similarity between the hamster melanopsin and bovine rhodopsin sequences, I used therefore homology modelling to obtain the melanopsin structure (Fig. 1A). I analyzed the binding of 11-*cis*-retinal in the melanopsin model comparing to the rhodopsin crystal structure (Fig. 1B). I found that the stabilization of the Schiff-base formed during the chromophore binding to the protein is weaker in melanopsin. This suggests that the photoisomerization process (involved in the entrainment of the circadian clock) is slower in melanopsin, than in rhodopsin, which is responsible for the night vision. [1]

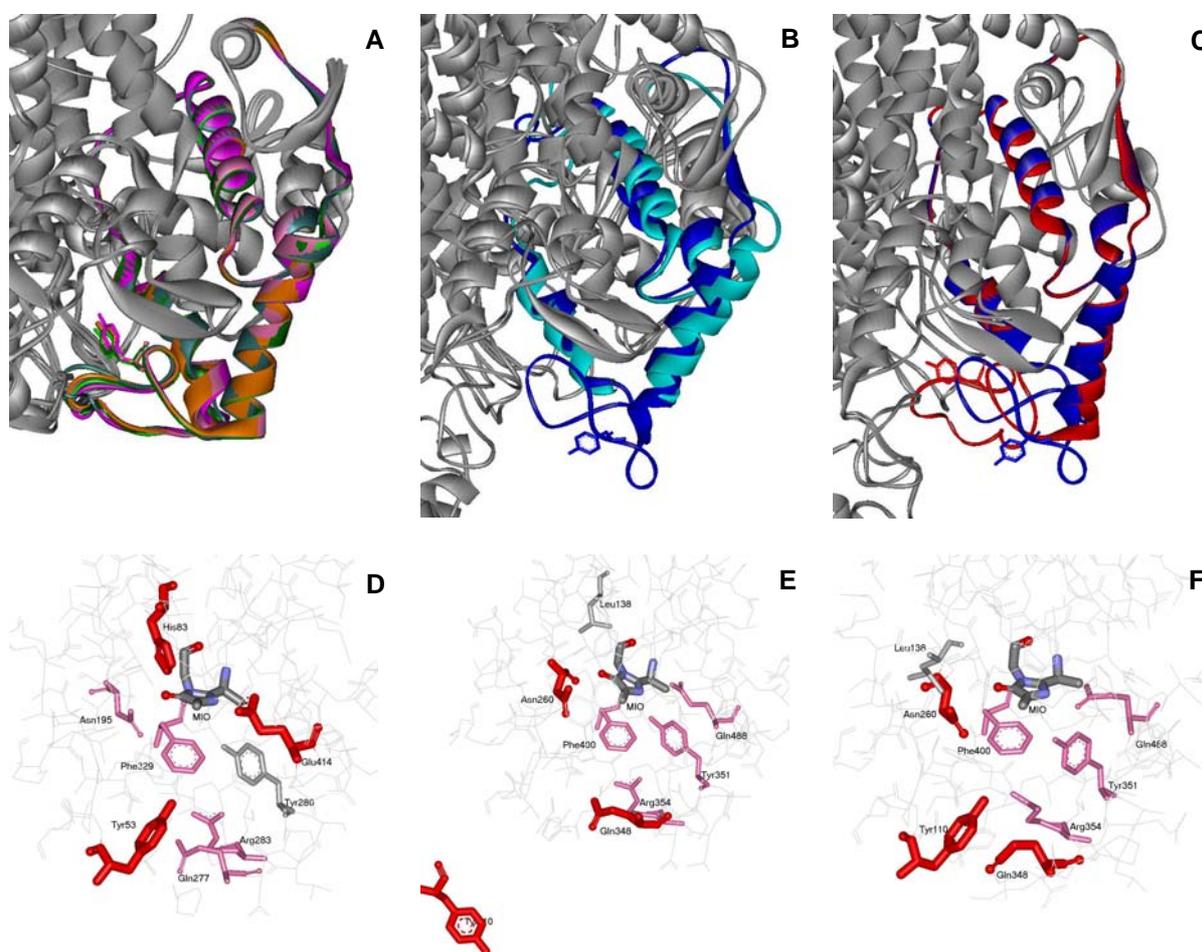


**Fig. 1.** (A) The composite homology model of the melanopsin protein: the transmembrane region (orange) was obtained from Swiss-Model, while the loop regions (green) are from Robetta server. (B) The 8 Å region (yellow) of the chromophore (red stick) in the protein.

### 3.2. Study of the ammonia-lyases

#### 3.2.1. The build-up of the active phenylalanine ammonia-lyase (PAL) structure

The catalytically essential Tyr110 (parsley PAL numbering) was missing from the bacterial PAL (PDB code: 1T6P) structure (Fig. 2B, cyan) and it was in a quite remote position from the active centre in the parsley PAL (PDB code: 1W27) structure (Fig. 2B, blue; Fig. 2E). Modifying the parsley PAL (1W27) crystal structure I build up a PAL model in which Tyr110 is close to the active centre like in histidine ammonia-lyase (HAL) structures (Fig. 2A, 2D). [2]



**Fig. 2.** (A) Bacterial HAL structures. (B) Bacterial (cyan) and parsley (blue) PAL crystal structures. (C) The experimental and modified crystal structures of parsley PAL. Active centre models of (D) HAL (1B8F), (E) PAL (1T6P) and (F) a PAL homology model.

I compared the modified PAL structure with the experimental structure. The PAL model was better than the crystal structure as a result of comparison of the energies (the model is more stable with ~640 kJ/mol) and protein structure characterizing Ramachandran plot analysis [2]. Performing molecular dynamics simulations I studied the stability of the model [2]. The simulations on 300 and 370 K confirmed the stability (with a movement of 1-1,5 Å) of the Tyr-in loop structure. This movement is sufficient for the substrate to enter and leave the active centre without the enzyme adopting the Tyr110-loop-out conformation present in the PAL crystal structure.[2]

### 3.2.2. The study of the tyrosine ammonia-lyase (TAL) mechanism by QM/MM calculations

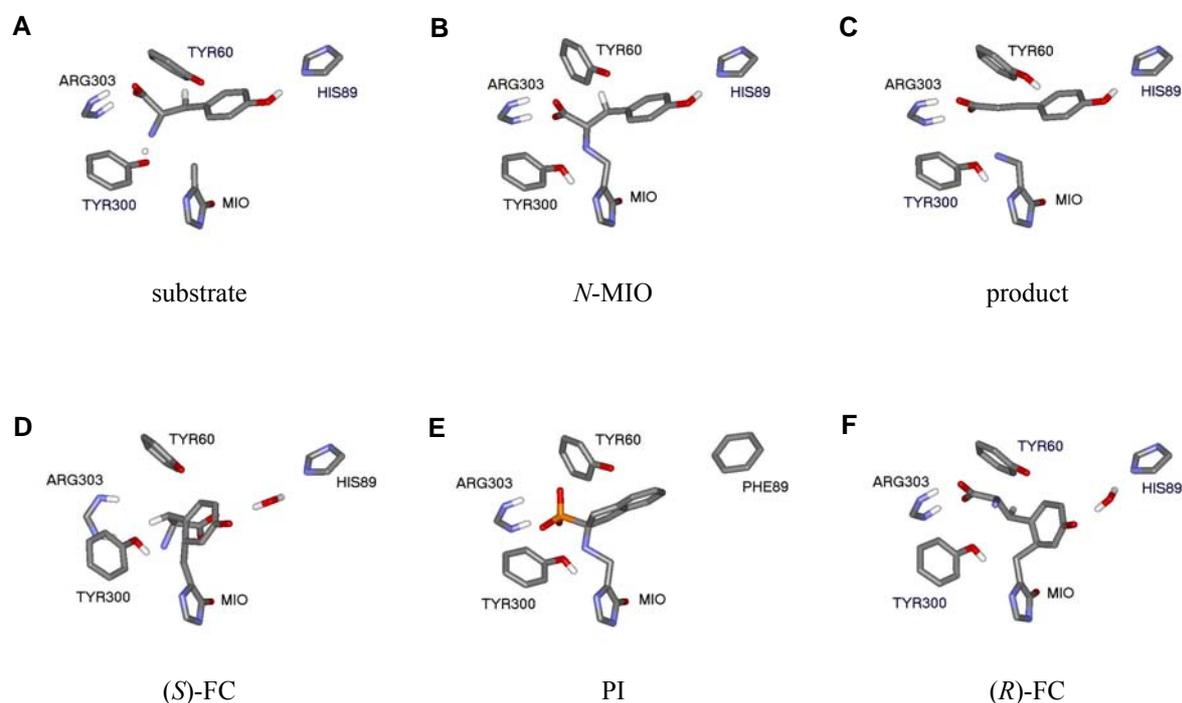
The recently published crystal structures of *Rhodobacter sphaeroides* tyrosine ammonia-lyase (*RsTAL*) contain the essential Tyr60 (Tyr100 in parsley) residue and the MIO prosthetic group in a close active centre.

Using the QM/MM method within the enzymic environment, I calculated the arrangement of the possible covalent intermediates and also the substrate and product states. The *N*-MIO intermediate has a more than 120 kcal lower energy than the favourable FC diastereomer [(*R*)-FC]. Therefore I concluded that the TAL reaction can proceed only *via* an *N*-MIO intermediate state. The energy values (Table 1) are in full agreement with experimental results since the TAL (and also PAL and HAL) reaction under normal conditions, is practically irreversible.[5]

**Table 1.** The potential energy (kcal/mol) difference for five different *RsTAL* active centre models

Model	PM3/UFF <sup>a</sup>	HF(3-21+G**)/UFF <sup>b</sup>	HF(6-31+G**)/UFF <sup>c</sup>
substrate + MIO	47,3	57,8	49,3
( <i>S</i> )-FC	> 10000	> 10000	> 10000
( <i>R</i> )-FC	136,0	145,0	139,8
<i>N</i> -MIO	7,5	0,0	0,0
product + H <sub>2</sub> N-MIO	0,0	4,9	4,9

<sup>a</sup>Relaxed at PM3/UFF level; <sup>b</sup>Relaxed at HF(3-21+G\*\*)/UFF level; <sup>c</sup>The geometries were relaxed at HF(3-21+G\*\*)/UFF level and then single point calculations were performed at HF(6-31+G\*\*)/UFF level



**Fig. 3.** Comparison of the structures [(A) substrate, (B) *N*-MIO intermediate, (C) product, (D) (*S*)-FC and (F) (*R*)-FC intermediates] calculated by HF(3-21+G\*\*)/UFF method within the *RsTAL* active site with (E) the experimentally determined arrangement of *N*-MIO bound 2-aminoindane phosphonate inhibitor (PI).

I compared the calculated structures with the phosphonate inhibitor (PI) containing *TAL* crystal structure. I found that the arrangement of the *N*-MIO intermediate is in full agreement with the PI structure (the aromatic moiety points towards His89 and the carboxylate group is to close Arg303). The elimination reaction *via* the *N*-MIO intermediate can proceed according to the least-motion principle, whereas substantial displacements would be required for a pathway even *via* the less distorted (*R*)-FC intermediate. [5]

### 3.3. Prediction of the product absolute configuration for the *Candida antarctica* Lipase B (CaLB) enzyme

I used the CaLB (PDB code: 1LBT) crystal structure for interpretation of the stereoselectivity of the products formed from esterification of secondary alcohols. [3,4]

I studied by QM/MM methods within the CaLB enzyme the absolute configuration of the products from the stereoselective acylation of four alcohols [1-(benzothiazol-2-yl)ethanol, BTZ2; 1-(benzothiophen-2-yl)ethanol, BTF2; 1-(benzofuran-3-yl)ethanol, BZF3 and 1-(benzothiophen-3-yl)ethanol, BTF3].

Conform the mechanism of serine hydrolases, a tetrahedral intermediate (THI) is formed between the acyl enzyme and alcohol, which results in forming of a new asymmetry centre. Therefore for each alcohol I studied four THI states (Fig. 4.)

The covalent bound reaction intermediate allowed systematical conformational analysis within the rigid enzyme, which constitutes the bases for the QM/MM calculations. Since the absolute configuration of the BTF2 product was unknown, I studied therefore with more sophisticated methods the 2nd substituted cases (BTF2, BTZ2), the rest of it were just references (Table 2.,3.). [4]

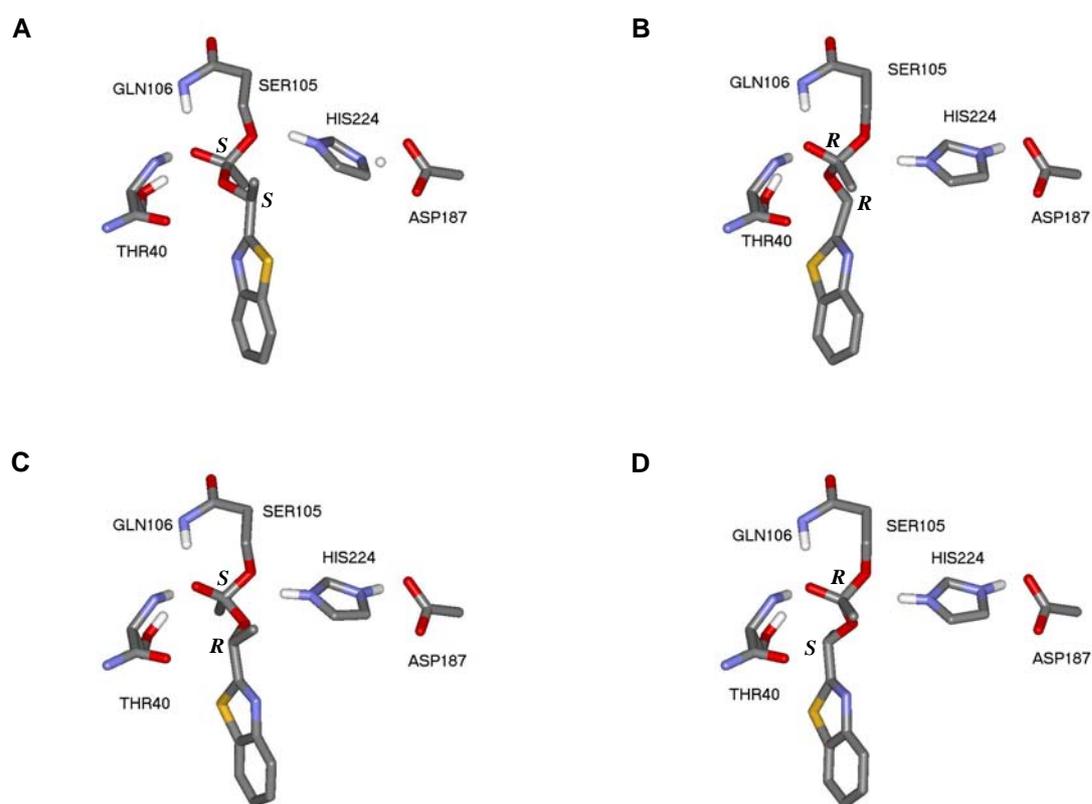
**Table 2.** Comparison of the THI states of acylation of 1-heteroarylethanols within the active site of CaLB determined by PM3/UFF method

substrate	<i>R</i> -alc- <i>R</i> -THI E (kcal/mol)	<i>R</i> -alc- <i>S</i> -THI E (kcal/mol)	<i>S</i> -alc- <i>R</i> -THI E (kcal/mol)	<i>S</i> -alc- <i>S</i> -THI E (kcal/mol)
<b>BTZ2</b>	20,1	<b>0,0</b>	<b>8,4</b>	21,2
<b>BTF2</b>	37,5	<b>0,0</b>	<b>14,8</b>	35,3
<b>BZF3</b>	16,1	<b>0,0</b>	10,8	<b>10,0</b>
<b>BTF3</b>	21,4	<b>0,0</b>	<b>10,7</b>	16,8

**Table 3.** Comparison of the THI states of acylation of 1-heteroarylethanols within the active site of CaLB determined by HF(3-21+G\*\*)/UFF method

substrate	<i>R</i> -alc- <i>S</i> -THI E (kcal/mol)	<i>S</i> -alc- <i>R</i> -THI E (kcal/mol)
<b>BTZ2</b>	<b>0,0</b>	<b>8,3</b>
<b>BTF2</b>	<b>0,0</b>	<b>6,8</b>

The QM/MM calculations show that in each case the (*R*)-alcohol-(*S*)-THI is the lowest energy intermediate, from which the (*R*)-acetate is the most likely forming product. Our results are in full concordance with experimental data. [4]



**Fig. 4.** Tetrahedral intermediate states within the active site of CaLB determined by PM3/UFF calculations for BTZ2

## 4. THESIS

### *4.1. The first 3D-structure of the melanopsin protein*

I build up the first 3D structure of the melanopsin protein, in which comparing to the rhodopsin structure I analyzed the chromophore binding. The chromophore stabilization in melanopsin is weaker than in rhodopsin. [1]

### *4.2. Ammonia-lyase structures*

For the study of PAL and TAL mechanism first I modified the arrangement of the catalytically essential Tyr110 containing loop in PAL crystal structure (of which stability I analyzed with molecular dynamics simulations) [2], then in the recent TAL structure I calculated the possible covalent intermediates formed during the reaction [5].

### *4.3. Product absolute configuration prediction by the QM/MM method in CaLB*

Using QM/MM methods within the *Candida antarctica* Lipase B (CaLB) enzyme I calculated the energies of the four possible tetrahedral intermediates (acyl enzyme – alcohol complex) formed during the acylation of heterocycle containing secondary alcohols. In each case the (*R*)-alcohol-(*S*)-THI state is the most stable, which favors the (*R*)-acetate formation. [3,4]

## 5. APPLICATIONS

Homology modelling can be used as a method for 3D structure determinations of proteins. A hamster melanopsin model gives the possibility to interpret photoisomerization processes. The G-protein coupled receptors, as proteins which bind ligands outside cells, can be good targets in drug design.

Since the phenylalanine ammonia-lyase plays an important role in plant metabolism, PAL enzyme is a common target of herbicides and is one of the most intensively studied plant enzymes. The mechanism of the ammonia-lyases can lead to understanding their function. Referring to our work, the thermic stability of the shorter chain PAL enzymes and the Tyr-loop-in state in the active PAL was experimentally demonstrated.<sup>8</sup>

The *Candida antarctica* Lipase B (CaLB) is a commonly used biocatalyst for the esterification of secondary alcohols. The absolute configuration of the product formed during the stereoselective process can be predicted by calculations performed within the enzymic environment.

The starting point of the calculation strategy which I developed is the systematic conformational analysis of the covalent ligand-enzyme intermediate. In addition, by QM/MM methods the intermediates, substrates, products states can be calculated within the enzyme. Owing a crystal structure, the QM/MM methods can be used for elucidating other enzyme mechanisms or to predict product configurations.

<sup>8</sup> Wang, L., Gamez, A., Archer, H., Abola, E. E., Sarkissian, C. N., Fitzpatrick, P., Wendt, D., Zhang, Y., Vellard, M., Bliesath, J., Bell, S. M., Lemontt, J. F., Scriver, C. R., Stevens, R. C., *J. Mol. Biol.*, **2008**, *380*, 623-635.

## 6. PUBLICATIONS

### 6.1. Publications related to the thesis:

#### *Appeared publications:*

1. Hermann, R., Poppe, L., **Pilbák, S.**, Boden, C., Maurer, J., Weber, S., Lerchl, A., Predicted 3-D structure of melanopsin, the non-rod non-cone photopigment of the mammalian circadian clock, from Djungarian hamsters (*Phodopus sungorus*), *Neurosci. Lett.*, **2005**, 376, 2, 76-80. [IF: 2,09; I: 1]
2. **Pilbák, S.**, Tomin, A., Rétey, J., Poppe, L., The essential Tyr-containing loop conformation and the role of the C-terminal multihelix region in eukaryotic phenylalanine ammonia-lyase, *FEBS J.*, **2006**, 273, 5, 1004-1019. [IF: 3,29; I: 6]
3. Szatzker, G., **Pilbák, S.**, Tőke, E. R., Bódai, V., Poppe, L., Stereoselectivity of *Candida antarctica* Lipase B reaction: transition states calculated by QM/MM methods, *FEBS J.*, **2005**, Suppl. 1, 272, 114-114. [IF: 3,29; I: 1]
4. Toşa, M., **Pilbák, S.**, Moldovan, P., Paizs, C., Szatzker, G., Szakács, G., Novák, L., Irimie, F.-D., Poppe, L., Lipase catalyzed kinetic resolution of racemic 1-hetero-arylethanol - Experimental and QM/MM study, *Tetrahedron: Asymm.*, **2008**, 19, 1844-1852. [IF: 2,47]

#### *Publications under preparation:*

5. **Pilbák, S.**, Rétey, J., Poppe, L., Mechanism of the tyrosine ammonia-lyase reaction, *FEBS J.*, **2008**, under preparation.

#### *Further publications and posters:*

6. **Pilbák, S.**, Poppe, L., A melanopsin fehérje homológia modellezése, *MKE QSAR, Kemometria és Molekulamodellezési Szakcsoporthok Előadói Ülése*, 25-26 March **2004**, Szeged, Hungary, *LECTURE*.

7. **Pilbák, S.**, Tomin, A., Poppe, L., Rétey, J., Lerchl, A., Homológia modellezés enzimek mechanizmusok vizsgálatára, *MTA Terpenoidkémiai és Elemorganikus Munkabizottság előadói ülése*, 2 April **2004**, Budapest, Hungary, *ELŐADÁS*.
8. **Pilbák, S.**, Poppe, L., Lerchl, A., Protein modeling as a tool for enzyme mechanistic studies, *Minisymposium on Cobalamins and Mimics*, 22-23 May **2004**, Antwerpen, Belgium, *LECTURE IN ENGLISH*.
9. **Pilbák, S.**, Rétey, J., Poppe, L., Enzimatisz folyamatok működésének tanulmányozása QM/MM módszerekkel, MKE Vegyészkonferencia, 19-21 June **2008**, Hajdúszoboszló, Hungary, *LECTURE*.
10. **Pilbák, S.**, Rétey, J., Poppe, L., A tirozin ammónia-liáz reakció mechanizmusa, *MKE QSAR, Kemometria és Molekulamodellézési Szakcsopoztok Előadói ülése*, 8-9 May **2008**, Szeged, Hungary, *LECTURE*.
11. **Pilbák, S.**, Noel, J. P., Poppe, L., A tirozin ammónia-liáz reaktív köztitermékének szerkezet-meghatározása QM/MM módszerekkel , *13<sup>th</sup> International Conference of Chemistry*, ISSN: 1843-6293, 8-10 November **2007**, Cluj-Napoca, Romania, *LECTURE*.
12. **Pilbák, S.**, Tomin, A., Rétey, J., Poppe, L., Role of the essential Tyr-containing loop in MIO-containing phenylalanine ammonia-lyase, *European Chemical Societies, 13<sup>th</sup> FEChem Conference on Heterocycles in Bioorganic Chemistry*, 28-31 May **2006**, Sopron, Hungary, *POSTER*.
13. **Pilbák, S.**, A fenilalanin ammónia-liáz vizsgálata számítógépes módszerekkel, *Dokozans Konferencia, BME Vegyész Kar*, 7 February **2006**, Budapest, Hungary, *LECTURE*.
14. Poppe, L., **Pilbák, S.**, Tomin, A., Conformation of the Y100 loop in phenylalanine ammonia-lyase. Loop correction of the x-ray structures by modeling and molecular dynamics, *Molecular Modeling in Chemistry and Biochemistry, Workshop*, 21-23 April **2005**, Cluj-Napoca, Romania, *LECTURE IN ENGLISH*.
15. **Pilbák, S.**, Forró, E., Fülöp, F., Poppe, L., A CaLB sztereoselektivitásának számítása, *MKE QSAR, Kemometria és Molekulamodellézési Szakcsopoztok Előadói ülése*, 27-28 April **2006**, Szeged, Hungary, *LECTURE*.
16. **Pilbák, S.**, Szatzker, G., Tőke, E. R., Bódai, V., Poppe, L., Stereoselectivity of *Candida antarctica* lipase B: QM/MM calculations for trans-cyclohexane-1,2-diol derivatives, *Biotrans Conference*, 3-8 July **2005** Delft, Holland, *POSTER*.

## 6.2. Publications not directly related to the thesis:

17. **Pilbák, S.**, Croft, A., Poppe L., QM/MM calculations on the rearrangement reaction of isobutyryl-CoA mutase, *FEBS J.*, **2005**, Suppl. 1, 272, 113-113.
18. **Pilbák, S.**, Croft, A., Poppe, L., Influence of the model truncation for the isobutyryl-CoA mutase reaction - Partial or full protonation?, *Minisymposium on Cobalamins and Mimics*, 25-29 May **2005**, Obergurgl, Austria, *LECTURE IN ENGLISH*.
19. **Pilbák, S.**, Croft, A., Poppe, L., Degree of model-truncation – Influence on the results of DFT calculations for the isobutyryl-CoA mutase reaction, *Molecular Modeling in Chemistry and Biochemistry, Workshop*, 21-23 April **2005**, Cluj-Napoca, Romania, *LECTURE IN ENGLISH*.
20. **Pilbák, S.**, Poppe, L., Rétey, J., Croft, A., Az izobutiril-CoA mutáz reakció tanulmányozása homológia modellezés és QM/MM számítások segítségével, *10<sup>th</sup> International Conference of Chemistry*, ISBN 973-7840-00-3, 291-293, 12-14 November **2004**, Cluj-Napoca, Romania, *LECTURE*.
21. **Pilbák, S.**, Croft, A., Poppe, L., The isobutyryl-CoA mutase reaction - a study by homology modelling and QM/MM calculations, *Minisymposium on Cobalamins and Mimics*, 4-5 November **2004**, Marburg, Germany, *LECTURE IN ENGLISH*.
22. **Pilbák, S.**, Poppe, L., Croft, A., The isobutyryl-CoA mutase reaction - a study by homology modelling and QM/MM calculations, Part I, *2nd Central European Conference "Chemistry towards Biology"*, 26-29 September **2004**, Seggau, Austria, *LECTURE IN ENGLISH*.
23. **Pilbák, S.**, Poppe, L., Synthetic efforts towards glutamate mutase reaction mimics, *Minisymposium on Cobalamins and Mimics*, 29-30 December **2003**, Stockholm, Sweden, *LECTURE IN ENGLISH*.