IN-SILICO IDENTIFICATION OF KEY RESIDUES FOR SHIFTING THE COENZYME SPECIFICITY OF AN ALDEHYDE-DEHIDROGENASE

MARIUS MIHĂŞAN[1*], VLAD ARTENIE[1], ZENOVIA OLTEANU[1]

Keywords: *A.nicotinovorans*, aldehyde-dehidrogenase, co-enzyme, mutagenesis, homology modelling

Abstract. The aldehyde dehydrogenase enzyme coded by the pAO1 megaplasmid of *A. nicotinovorans* has an 35% identity at the sequence level with the succinic-semialdehyde dehydrogenase from *E. coli*, the latter being a suitable template for homology modeling experiments. The computer generated model of AlDH helped at identifying the residues implicated in catalysis (C266, E232, R143, Q266, S423) as well as in co-enzyme specificity (K158, Ser161, G215). Several single mutants were constructed and docked with NAD^+ and $NADP^+$ in an attempt to identify the key residues for shifting the NAD/NADP preference of the pAO1 coded enzyme.

INTRODUCTION

Glutaraldehyde (GA, dialdehyde glutaral) based instrument sterilants have been widely used for over 30 years as effective high-level disinfectants in hospitals, doctors’ offices, and dental offices throughout the world. The widespread use of glutaraldehyde-based instrument sterilants can be attributed to their efficacy, speed of disinfection, ease of use, cost-effectiveness, and lack of corrosivity (Jordan *et al.*,1996). Studies concerning the toxicity (Cosmetic Ingredient Review Expert Panel,1996) and carcinogenicity (Clair *et al.*,1991; Vergnes & Ballantyne,2002) of this product indicated an potential dangerous product and raised the question on inactivation procedures for GA contaminated products. Traditionally, only chemical inactivation of GA is used, ether by reaction with ammonia to form a nonmicrobiocidal GA–NH_2 complex or by treatment with sodium bisulfite (Jordan *et al.*,1996). A third method of inactivation, only usable for lower concentrations of GA have been was developed, in which the pH is raised to 11 or 12 in order to form GA polymers, and then neutralized to pH 7 prior to disposal. All of this methods produce by-products which, although are not toxic for the microflora, apparently are are dangerous for the algae.

Recent studies on the genetic organization of the pAO1 catabolic megaplasmid of *Arthrobacter nicotinovorans* has lead to the identification of a novel tagatose catabolic pathway (Mihasan,2010). Part of this pathway is a monomeric aldehide-dehidrogenase (AlDH) with a wide substrate specificity which was recently purified and partiality characterized (Mihasan *et al.*,2009). Interesting, AlDH is able to degrade with significant speed also GA among others aldehydes, being this way a possible alternative for GA inactivation. Furthermore, considering a double dehidrogenasion of GA as described previously by (Aghaie *et al.*,2008) the product would be glutaric acid, a useful compound in plastic industry. Unfortunately the biotechnological applicability of this oxidoreductase is very limited due to the fact that it prefers NADP, a very expensive to regenerate coenzyme (Schneider, Wubbolts, Sanglard & Witholt,1998; Torres Pazmiño, Winkler, Glieder & Fraaije,2010).

Pyridine-nucleotide-dependent enzyme families are generally noted for strict specificity for either NAD(H) or NADP(H). Enzymes specific for NAD generally act in oxidative, catabolic reactions while those using NADPH generally play reductive, anabolic roles. Thus, simple phosphorylation of the hydroxyl group at the 2 H position of the adenine ribose of NAD, to create NADP, effectively yields a distinct coenzyme with nonoverlapping uses. But, as with most general statements, there are exceptions.

No fewer than 13 distinct families comprise the aldehyde dehydrogenase (ALDH) extended family (Perozich J., Nicolas H., Wang, Bi-Cheng, Lindahl R., Hempel J.,1999). Most ALDH families are specific for NAD, a few are specific for NADP, and others have yet to be sufficiently characterized in this regard. Class 3 ALDHs are notable as the only ALDH family with a well- established ability to use either NAD or NADP. Regardless of their co-enzyme specificity, NAD(P)-dependent ALDH bind the co-enzyme using the same scaffold, namely a six-stranded open α/β `Rossmann' domain, in which a characteristic spacing of glycine residues is expected to be predictive for the location of the turn between β-1 and α- A (Bellamacina R. Cornelia,1996). One hallmark of the Rossmann fold in NAD-specific dehydrogenases is the acidic residue at the end of β-2. This spatially corresponds to the location of the 2'- adenine ribose hydroxyl in both ALDHs and in traditional Rossmann folds, explaining the absence of an acidic residue at this location in NADP-specific enzymes (Branden & Tooze,1999).

So apparently, as both NADP and NAD dependent enzymes use the same scaffold for nucleotide binding, changing the preference for the co-enzyme would resume at changing several residues in key locations on the scaffold. This work aims at identifying the crucial aminoacids involved NADP/NAD specificity of the AlDH from *A. nicotinovorans* by means of *in-silico* mutagenesis, molecular modeling and docking in an attempt to further enlarge the biotechnological applicability of this newly found enzyme.

MATERIALS AND METHODS

Comparative modeling. The AlDH protein sequence (GenBank GI:25169061) was used for three-dimensional model generation using SWISS-model (Arnold *et al.*,2006). Model validation was achieved by using Procheck (Laskowski *et al.*,1993). Structural alignments were performed using DaliLite (Holm & Park,2000). Root mean square (' RMS ') value calculations between the model and the templates used for comparative modelling, using the Cα fitting, as well as all images were generated using the Swiss-PdbViewer v4.0.1 (Guex & Peitsch,1997). Pockets inside the proteins structures were detected and measured using Pocket-Finder (Laurie & Jackson,2005).
***In-silico* mutagenesis** was performed using the TRITON software package (Prokop, Damborský & Koca,2000; Prokop *et al.*,2008), (Damboský, Prokop & Koca,2001) interfaced with MODELLER 9v8.(Eswar *et al.*,2007; Eswar *et al.*,2007; Sali & Blundell,1993). This approach uses the method of the satisfaction of spacial restraints for model building with the structure from the structural database as a template and the amino acid sequence of the studied protein with the desired substitution as the target sequence.
Ligands. The 3D structures of NAD and NADP wer**e** downloaded from PubChem (CID 15938971 and CID 15938972 respectively) and transformed in tridimensional coordinates using FROG v.1.01 – free on-line drug conformation generation (Leite *et al.*,2007).
Docking was performed by the academic software AutoDock 4(Morris *et al.*,1998) using the TRITON software as am interface. For the protein partial atoms charges, the Kollman united atom charges were used as suggested in the AutoDock user manual. The Lamarckian genetic algorithm was used for the search of the energetically favored binding modes. A box set containing the whole NADP binding site (its center on the C226, the dimensions being 64 × 68× 72 grid points with 0.375 Å spacing), with 100 runs of Lamarckian algorithm and 500 000 energy evaluations per run were used. The population size was 50 individuals. The evaluation of scoring function energies and the free energies using the cluster analysis with a 2 Å rmsd threshold was performed by AutoDock. Input data were prepared, and the results were visualized by TRITON software. All calculations were done on a Linux Kernel 2.6.32-24 machine.

RESULTS AND DISCUSSIONS

Model description and quality. The amino-acid sequence of AlDH was used to generate putative structures using the SWISS-MODEL server (Arnold *et al.*,2006). The Alignment mode tool from SWISS-MODEL identified as potential templates several structures (table 1).
Table 1.
Suitable templates for homology modeling of A. nicotinovorans AlDH as identified by SWISS-MODEL

PDB ID	Function	Sequence identity with AlDH
3efvA	Putative protein	44 %
3jz4A(Langendorf *et al.*,2010)	NADP dependent succinic semialdehyde dehydrogenase	35 %
3eklA	Putative protein, unpublished	34 %
3ifgE	Putative protein, unpublished	33 %
1bxs (Moore A.S., Baker M.H., Blythe T., Kitson K., Kitson M.T., Baker E. N.,1998)	Sheep liver cytosolic aldehyde dehydrogenase	32 %

The NADP dependent semialdehyde dehydrogenase from *E.coli* (PDB id. 3jz4A, SSADH) was the only protein with known function and significant sequence identity compared with *A. nicotinovorans* AlDH. The alignment between 3jz4A and AlDH was hand edited and used to generate the 3D model. In the the resulted putative model 90,2 % of non-glycine and non-proline residues have conformational angles (φ, ψ) in the most favored of the Ramachandran plot, 7.7 % fall in the 'additional allowed regions', 1,8% fall in the 'generously allowed regions' and 0.3 % fall in the 'disallowed region', as defined by Procheck (Laskowski *et al.*,1993). 17 out of 20 prolines and 35 out of 37 glycine residues are in permitted regions. The root mean square (RMS) value between the model and the template used for comparative modelling is 0.26 Å for 454 Cα atoms. This, as well as the fact that the sequence identity between the template and AlDH is 35% indicates a god quality model.

The computer generated AlDH model follows the same general structural organization as 3jz4A with a typical NAD(P)+ dehydrogenase fold with four β sheets (A–D) and 13 α helices (1–13) . The secondary structure assignments used in this report are similar to those described by (Langendorf *et al*.,2010) and are as shown in Figure 1. The L-shaped molecule consists of three domains: the catalytic domain, the cofactor binding domain and the oligomerisation domain. The catalytic domain consists of a central 7 stranded b-sheet (the D-sheet) flanked by 2 helices on one side and 3 helices on the other. The catalytic loop is located adjacent to the cofactor binding site. The cofactor binding domain would interact with NADP+ via two tandem Rossmann folds in a $(\beta\alpha)_4\beta$ formation. This is a variation of the classic Rossmann fold (Weiner & Hurley,2001), where the last β strand of first $(\beta\alpha)2\beta$ motif forms the first β strand of the second $(\beta\alpha)2$ β motif. The oligomerisation domain comprises an elongated 3-stranded β-sheet (the B-sheet), which interacts with two other monomers in the final tetrameric assembly.

Putative co-enzyme binding site. As shown in our previous work (Mihasan *et al.*,2009), AlDH is able to use as a co-enzyme both $NADP^+$ and NAD^+. By homology modeling and using the existing informations regarding the co-factor binding site in *E. coli* SSADH, the putative co-factor binding site could be describes also for AlDH. Its comprises of two pockets; one of which is close to the surface of the AlDH molecule and would accommodates the adenosine (adenine and the first ribose) and the 2’phosphate. The second pocket is located centrally in the active site and would accommodate the second ribose and the nicotinamide. The catalytic residues are located at the center of the molecule with two funnel-like openings on the surface of either side of the molecule. The larger opening functions to allow entry of the cofactor. The conserved catalytic residues (C288 and E254) and the active site residues (R164, R282 and S445) of *E. coli* SSADH structure superpose well with that of *A. nicotinovorans* AlDH respectively C266, E232, R143, Q266, S423 (Figure 1).

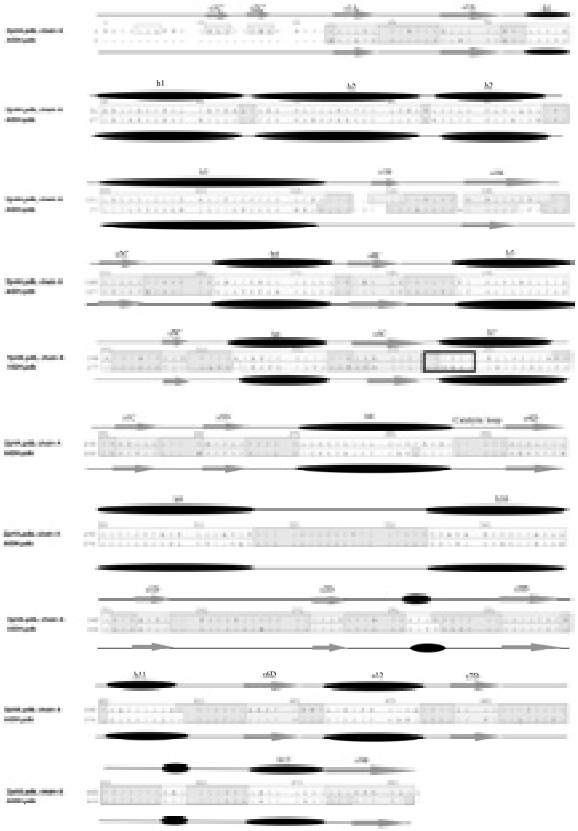

Figure 1. Alignment of *E. coli* SSADH with *A. nicotinovorans* AlDh SSADH. Conserved residues have been highlighted according to the ClustalX score. The secondary structure (*E. coli* SSADH above the sequence and *A. nicotinovorans* AlDh SSADH below the sequence) has been marked with either an arrow designating a β-sheet or a oval representing an α-helix. Structurally important regions have also been marked and labelled, catalytic loop and the GXXXXG motif (box) from the Rossmann fold.

Mutating K158 to H switches the NADP/NAD binding energies. The catalytic mechanism for aldehyde dehydrogenases is well characterised and is the same diregarding the preference for NAD or NADP. The first step of the reaction is nucleophilic attack by the catalytic C288 residue on aldehyde to give the hemithioacetal intermediate. Hydride transfer from this intermediate to $NAD(P)^+$ results in formation of the thioacyl enzyme intermediate and NAD(P)H. Lastly, the conserved E232 residue acts as a general base to deprotonate a water molecule prior to its attack on the thioacyl enzyme intermediate resulting in formation of the coresponding acid and regeneration of the C288 residue. The key difference between NAD and NAD preferring enzymes resides in the shape and size of the cofactor binding pocket. For example, human SSADH utilizes NAD^+ as a cofactor, whereas E. coli SSADH utilizes $NADP^+$. The structure comparison of this two proteins reveals the basis for this preference – in E. coli SSADH a three-residue deletion (of the human sequence 261RKN263) in the loop connecting s5C and h6 permits accommodation of the extra phosphate group of $NADP^+$ (2'phosphate). Consequently, the *E. coli* SSADH can utilize both NAD^+ and $NADP^+$ as a cofactor but the activity in the presence of NAD^+ molecule is only 1/20 of that of $NADP^+$ (Langendorf *et al.*,2010). In human SSADH, 261RKN263 occupies the space for the 2'phosphate of NADP+ and consequently, only NAD^+ but not $NADP^+$ can be utilized as a cofactor for this enzyme.

AlDH of *A. nicotinovorans* is able to use both $NADP^+$ and NAD^+, but the speed of the reaction is low with with the latter co-enzyme (Mihasan *et al.*,2009). The simple insertion of the above mention RKN residues would only impair the enzyme to use $NADP^+$. We focused on mutating single key residues to disrupt the catalytic binding site in order to both impair $NADP^+$ binding and to improve the protein affinity towards NAD^+. Based on sequence similarities with various aldehyde dehydrogenases three aminoacids were selected for mutation, as depicted in table 2.

Table 2. Key residues in AlDH of *A. nicotinovorans* selected for mutagenesis and their role in NAD/NADP binding

Residue	Role and reference	Mutation
K158	Interacts with 2' phosphate of NADP+ (Langendorf *et al.*,2010)	Changed to a bulky residue H, W as well as a neutral one A
Ser161	NAD+ enzymes are characterised by a dicarboxilic residue in this position, while NADP+ enzymes have a small residue (Perozich J., Kuo I., Wan B.C., Boesch J.S., Lindahl R., Hempel J.,,2000)	Mutated to a dicarboxilic residues E or D
G215	Part of the GxGxxG motif of NAD(P) utilizing enzymes. In NADP enzymes the last G is replaced by bulkyresidues (Bellamacina R. Cornelia,1996)	Mutated into a more bulky aminoacid S and P

Tridimensional models with the mutated proteins were build using the MODELLER 9v8.(Eswar *et al.*,2007; Eswar *et al.*,2007; Sali & Blundell,1993) and used for an round of *in-silico* docking experiments using both NAD^+ and $NADP^+$. The overview of the docking results can be seen in table 3.

Table 3. Docking results with the single mutants of AlDH of *A. nicotiniovorans*

Mutant	Free energy (kcal/mol)	
	NAD^+	$NADP^+$
K158W	-10,06	-9,74
K158H	-9,79	-6,76
G215P	-8,61	-5,71
S161D	-8,4	-8,05
K158A	-7,56	-6,83
WT	**-7,38**	**-8,26**
G215S	-6,98	-3,98
S161E	-5,58	-4,42

Noticeable, not all the selected mutations lead to an improvement in $NADP^+$ binding energy. Moreover, two

mutations G215S and S161E had as a result a significant drop in the binding affinity for both NAD^+ and $NADP^+$.

In the case of all the tested mutants, the predicted position and conformation of the NAD^+ in the catalytic site is the same as for the wt protein (figure 2, C). Nevertheless, the K158W, K158H, G215P, S161D, K158A mutants exhibit higher binding energies for NAD^+ compared to the wt, probably due to changes of the local hidrophobicity and charge in the catalytic pocket. Noticeable, in the case of all the mutants, the free energy score would indicate that the enzyme binds NAD^+ stronger then $NADP^+$. This is due to the situation depicted in figure 2, A and B, where in the wt the $NADP^+$ is bound close to the catalytic residues, at only 4,86 Å from the K158 residue. In the K158W, K158H due to the bulkiness of the histidine or triptophane, the 2' phosphate does not have enough space and the coenzyme is bound much further from the catalytic residues, at 11.95 Å, dropping severely the free energy values.

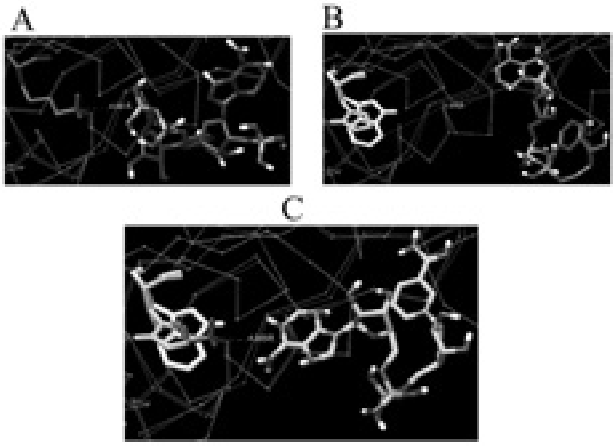

Figure 2. Superposition of the wt AlDH and the best scored mutants: K158W K158H. The protein backbone is represented as wire in the background. The mutated residues are represented as sticks K158 is in red, W in yelow and H in cyan. **A.** $NADP^+$ binding and position in the wt protein. **B.** $NADP^+$ binding is disrupted in the mutant proteins. **C.** NAD binds in the same maner in both the wt and the mutated proteins.

CONCLUSIONS

Mutating the selected residues in the AlDH model lead to various responses. Although in all the mutants the preference for NAD/NADP was shifted as desired, in some cases the affinity for NAD^+ was lower then the wt. The most favorable position for mutagenesis is K158, two of the mutants at this location K158W and K158H having an docking score for NAD higher than the score obtained by the wt protein. The G215 position is also suitable for mutagenesis; currently we are undergoing in-silico mutagenesis experiments for obtaining double mutants for this two positions.

REFERENCES

Aghaie Asadollah,, Lechaplais Christophe,, Sirven Peggy,, Tricot Sabine,, Besnard-Gonnet Marielle,, Muselet Delphine,, de Berardinis Veronique,, Kreimeyer Annett,, Gyapay Gabor,, Salanoubat Marcel,, Perret Alain, (2008): *New Insights into the Alternative d-Glucarate Degradation Pathway*. Journal of Biological Chemistry, 283(23), 15638-15646,

Arnold Konstantin,, Bordoli Lorenza,, Kopp Jürgen,, Schwede Torsten, (2006): *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics, 22(2), 195-201,

Bellamacina R. Cornelia (1996): *The nicotinamide dinucleotide binding motif: a comparison of nucleotide binding proteins*. FASEB Journal, 10(), 1257-1269,

Branden, C. & Tooze, J. (1999): *Introduction to Protein Structure*, 212-215, ed., vol. , Garland Publishing.

Clair1991. **Clair Mary Beth,, Bermudez Edilberto,, Gross Elizabeth,, Butterworth Byron,, Recio Leslie,, Carrano A V,** (1991): *Evaluation of the genotoxic potential of glutaraldehyde*. Environmental and Molecular Mutagenesis, 18(), 113-119,

Cosmetic Ingredient Review Expert Panel (1996): *Final Report on the Safety Assessment of Glutaral* . International Journal of Toxicology, 15(), 98-139,

Damboský J,, Prokop M,, Koca J, (2001): *TRITON: graphic software for rational engineering of enzymes*. Trends Biochem Sci, 26(1), 71-73,

Eswar Narayanan,, Webb Ben,, Marti-Renom Marc A,, Madhusudhan M S,, Eramian David,, Shen Min-Yi,, Pieper Ursula,, Sali Andrej, (2007): *Comparative protein structure modeling using MODELLER*. Curr Protoc Protein Sci, Chapter 2(), Unit 2.9,

Guex N,, Peitsch M C, (1997): *SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling*. Electrophoresis, 18(15), 2714-2723,

Holm Liisa,, Park Jong, (2000): *DaliLite workbench for protein structure comparison*. Bioinformatics, 16(6), 566-567,

Jordan S L,, Russo M R,, Blessing R L,, Theis A B, (1996): *Inactivation of glutaraldehyde by reaction with sodium bisulfite*. J Toxicol Environ Health, 47(3), 299-309,

Langendorf Christopher G,, Key Trevor L G,, Fenalti Gustavo,, Kan Wan-Ting,, Buckle Ashley M,, Caradoc-Davies Tom,, Tuck Kellie L,, Law Ruby H P,, Whisstock James C, (2010): *The X-ray crystal structure of Escherichia coli succinic semialdehyde dehydrogenase; structural insights into NADP+/enzyme interactions*. PLoS One, 5(2), e9280,

Laskowski RA,, Macarthur MW,, Moss DS,, Thornton JM, (1993): *{PROCHECK}: a program to check the stereochemical quality of protein structures*. J. Appl. Cryst, 26(), 283-291,

Laurie Alasdair T R,, Jackson Richard M, (2005): *Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites*. Bioinformatics, 21(9), 1908-1916,

Leite T Bohme,, Gomes D,, Miteva M A,, Chomilier J,, Villoutreix B O,, Tufféry P, (2007): *Frog: a FRee Online druG 3D conformation generator*. Nucleic Acids Res, 35(Web Server issue), W568-72,

Mihasan Marius, (2010): *In-silico evidence of a pAO1 encoded tagatose-derivate catabolic pathway in Arthrobacter nicotinovorans*. Biologia - Sect. Cellular and Molecular Biology, 65(), ,

Mihasan Marius,, Artenie Vlad,, Brandsch Roderich, (2009): *Purification of a novel aldehide-dehidrogenase with wide substrate specificity*. Analele Stiintifice ale Universitatii Alexandru Ioan Cuza, Sectiunea Genetica si Biologie Moleculara, X(2), 9-13,

Moore A.S., Baker M.H., Blythe T., Kitson K., Kitson M.T., Baker E. N. (1998): *Sheep liver citosolic aldehyde dehydrogenase: the structure reveals the basis for retinal specificity of class 1 aldehyde dehydrogenase*. Structure, 6(), 1541,

Morris Garrett M,, Goodsell David S,, Halliday Robert S,, Huey Ruth,, Hart William E,, Belew Richard K,, Olson Arthur J, (1998): *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*. Journal of Computational Chemistry, 19(), 1639-1662,

Perozich J., Kuo I., Wan B.C., Boesch J.S., Lindahl R., Hempel J., (2000): *Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase*. Eur. J. Biochem, 267(), 6197-6203,

Perozich J., Nicolas H., Wang, Bi-Cheng, Lindahl R., Hempel J. (1999): *Relationships whithin the aldehyde dehidrogenase extended family*. Protein Science, 8(), 137-147,

Prokop M,, Damborský J,, Koca J, (2000): *TRITON: in silico construction of protein mutants and prediction of their activities*. Bioinformatics, 16(9), 845-846,

Prokop Martin,, Adam Jan,, Kríz Zdenek,, Wimmerová Michaela,, Koca Jaroslav, (2008): *TRITON: a graphical tool for ligand-binding protein engineering*. Bioinformatics, 24(17), 1955-1956,

Sali A,, Blundell T L, (1993): *Comparative protein modelling by satisfaction of spatial restraints*. J Mol Biol, 234(3), 779-815,

Schneider Silke,, Wubbolts Marcel G.,, Sanglard Dominique,, Witholt Bernard, (1998): *Biocatalyst Engineering by Assembly of Fatty Acid Transport and Oxidation Activities for In Vivo Application of Cytochrome P-450BM-3 Monooxygenase*. Appl. Environ. Microbiol., 64(10), 3784-3790,

Torres Pazmiño DE,, Winkler M,, Glieder A,, Fraaije MW, (2010): *Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications*. Journal of Biotechnology, 146(1-2), 9-24,

Vergnes Jane S,, Ballantyne Bryan, (2002): *Genetic toxicology studies with glutaraldehyde*. J Appl Toxicol, 22(1), 45-60,

Weiner, H. & Hurley, T. (2001). NAD(P)+ Binding to Dehydrogenases in H. Weiner (Ed.), *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd, New-York, .pg. 1-7)

Acknowledgements. This work was supported by CNCSIS-UEFISCSU, project number PN II- RU 337/2010.

1 „Alexandru Ioan Cuza" University, Faculty of Biology, Department of Molecular and Experimental Biology, Iaşi, Romania,
E-mail: marius.mihasan@uaic.ro