COMPUTER-BASED MODELING FOR SUGAR PREFERENCES OF AN OXIDOREDUCATSE FROM ARTHROBACTER NICOTINOVORANS PAO1 PLASMID

MARIUS MIHĂŞAN[1,2*], VLAD ARTENIE[1]

Keywords: *computer-modeling, oxidoreductase, binding site, sugars*
Abstract: The pAO1 megaplasmid of Arthrobacter nicotinovorans encodes two different pathways: one for nicotine metabolism and an putative sugar catabolic pathway. Using computational methods, the preferences for sugar of an oxidoreductase from the megaplasmid was addressed. The experimental and modeling data showed that the enzyme is tetrameric, has an ligand binding place close to the EKP conserved sequence and might prefer as substrate ketohexoses.

INTRODUCTION

Plasmids are simple genetic elements, which give to the bacterial cells great metabolic versatility. Several plasmid-encoded pathways were described (ex: for metabolism of phthalate or naphthalene), allowing the bacteria to be present in many environments as natural autochthonous microflora with a high potential for bioremediation of pollutants. Such a microorganism is the gram positive soil bacteria *Arthrobacter nicotinovorans.* The presence of the 165- kb pAO1 megaplasmid inside the cell allows this bacteria to grow on nicotine as sole carbon and nitrogen sources.

The sequence of this plasmid was established and two putative pathways could be described : on one hand the nicotine-degrading pathway, almost fully described by and on the other hand an yet unknown putative sugar-catabolic pathway. The overall GC content of the pAO1 plasmid indicates that nicotine-catabolism gene clusters are a new acquisition, being attached during the evolution to an older plasmid, containing the sugar-catabolic pathway.

This latter pathway is comprised of several genes, among which an putative cellulase, an ABC-transporter system gene cluster and a cluster of several dehydrogenases and oxidoreductases. This last cluster probably encodes the last steps of the pathway, connecting it to the general metabolism of the cell. A part of this cluster are ORF39, an putative succinate-semialdehyde dehydrogenase and ORF40, an putative oxidoreducase. The ORF39 was found to encode an monomeric-aldehyde dehydrogenase with a broad substrate specificity.

Our current study is focused on the ORF40 protein and its possible role in the cell. Using computational methods, an possible tridimensional structure was generated and the protein was scanned for possible binding site for ligands. Two possible binding sites were found and the substrate specificity of this enzyme is discussed.

MATERIAL AND METHODS

Protein purification was achieved using standard IMAC techniques on Fast-Flow Ni-chelating Sepharose (Amersham Biosciences, Sweden). **Native molecular weight determination** was done using gel permeation chromatography on an HiLoad 16/60 Superdex 200 column connected to an AKTA Basic FPLC system.

Computational methods. Molecular coordinates of the protein were generated by *EasyPred3D Web Server* 1.0 using the *orf40* aminoacid sequence derived from the *in-silico* translated sequence of pAO1 (accession number in NCBI Nucletide database: gi|25169022|). Ligand molecules were part from the Protein Data Bank and part were generated using ZINC database as shown in table 1.

Both the receptor molecule and the ligands were prepared for docking with the Chimera Antechamber , sphgen and grid programs . Molecular docking of the ligand and the receptor were done with the default options using the DOCK6 program (UCSF) running on an Linux machine (kernel 2..6.24)

RESULTS AND DISCUSSIONS

The putative structure. Using as input the aminoacid sequence of ORF40, the EasyPred3D Web Server 1.0 returned a putative tridimensional model of the protein. The model was build on the structure of glucose-fructose oxidoreductase, subunit A (GFOR, PDB id 1H6D) from Zymomonas mobilis. The two proteins share 19.5 % similarity at the sequence level and the same general organization. GFOR monomer consists of two well defined domains: the N-terminal domain involved in the binding of NADP and the C-terminal domain involved in the oligomerization.

The nucleotide binding fold from GFOR is different than the common Rosmann fold. The $NADP^+$ is totally buried in the protein and forms 21 potential H-bonds with its sourounding aminoacids, being this is way very tightly bound to the protein and it not released during catalysis. The purified recombinant ORF40 protein was analyzed for the NADP/NAD content by

TLC, but the none of coenzymes could be detected. An sequence based alignment of the two proteins shows that only 7 aminoacids involved in the formation of H-bonds with NADP are conserved in the both GFOR and ORF40. Probably, the binding of the coenzyme differ in this two enzymes, GFOR being the only known case of an strong un-covalently bound $NADP^+$ containing enzyme. An other interesting point is the fact that the Lys69 from GFOR, the aminoacid responsible for the coenzyme preference , is also conserved in the ORF40, indicating that the latter enzyme might prefer $NADP^+$.

One more particular thing about N-terminal domain in the GFOR protein is the presence of an N-terminal arm rich in prolines. This arm is frequent in tetramer proteins like dog fish muscle LDH, contributing to the stabilization of the quaternary structure. GFOR is also tetramer and proteolitic cleavage of this arm leads to dimer formation. Still, reports could be found indicating that the N-terminal arm is not required for stabilization, many bacterial LDHs lacking this feature being tetrameric. The arm is missing in the computer-generated structure of ORF40 protein, but the C-terminal domain is similar to the β sheet structure found in the structure of GFOR. This terminal β sheet is directly involved in the quaternary structure formation.

Gel permeation chromatography with the homogeneous purified ORF40 protein was performed in order to established its oligomeric state. The relative molecular weight determined by SDS-PAGE was approximatively 47 kDA and determined by gel filtration was of approximatively 163 kDa, indicating that the protein is a tetramer in solution.

Ligand binding. Analyzing the aminoacid sequence of ORF40, one partially conserved consensus sequence could be found: 119-AGKHIFTEKP-128, which is quite similar to AGKHVxCEKP motif found in sugar-dehidrogenases.. In order to establish the best candidates s substrate for the enzyme, *in-silico* experiments were conducted using the dock6 program, the putative model of ORF40 as receptor and 29 sugars as ligands. All the tested ligands were found to bind to the protein, but the binding site differers. Two ligands (O1-methyl-glucose and L-xylulose-P) are predicted to bind at the surface of the molecule and we excluded them. The rest of the ligands are clustered into two different distinct regions (fig.1), hereby noted A and B.

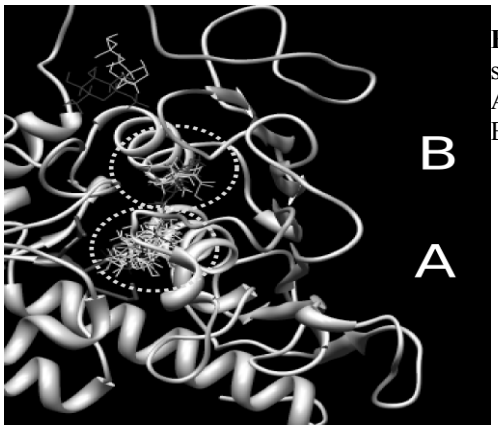

Fig. 1. The proposed binding sites in the putative structure of ORF40.
A – the binding site for 23 out of 29 tested sugars
B – the binding site for 4 out of 29 tested sugars

The model molecule (GFOR protein) acts on two substrates by an ping-pong mechanism: glucose is bound, transformed into gluconolactone and released, then fructose is

bound and reduced to sorbitol. Surprisingly, both the substrates are bound in in the same place , close to the 128-EKP-130 motif. The three aminoacids are directly involved in the catalysis and in the binding of $NADP^+$. In the modeled structure of $NADP^+$ only site A is close to the equivalent conserved sequence 126-EKP-128 (shown with ball and stick representation in figure 1). This, and also the fact that 23 out of 29 tested sugar bind in this site lead us to the conclusion that the site A is the real binding site of the ORF40 proposed model.

Substrate preferences. Aside from the prediction of binding place of different ligands, the dock6 program allows us to quantify the force of interaction by calculating the interaction energy (docking score). This kind of scores were used previously to test the binding preferences of lectin (2, 11). The results are shown in table 1, as well as the source of ligand molecule and the binding site.

Table 1. Ligands used and scores obtained in the *in-silico* docking experiments

Ligand	Source (pdb id or zinc database)	Docking score	Predicted binding site
D-tagatose	tag	-31.34	A
D-psicose	fru	-28.81	A
L-sorbose	ZINC	-28.81	A
D-xylose	xls	-28.47	A
L-xilulose	xul	-28.45	A
D-glucose	bgc	-27.62	A
D-Fructose	fru	-27.45	A
L-xylulose	xul	-27.24	A
D-galactose	gal	-26.74	A
Anhidro-L-galactose	ZINC	-26.53	A
D-lyxopyranose	ldy	-26.47	A
L-xylose	ZINC	-25.93	A
D-sorbose	ZINC	-25.86	A
L-arabinos	ara	-25.13	A
deoxiribose	ZINC	-24.56	A
L-ribose	ZINC	-24.48	A
L-arabinose.pdb	arb	-24.36	A
D-mannose	bma	-23.72	A
Annidro-D-galactose	ZINC	-23.23	A
eritrulose	ZINC	-23.09	A
L-manopiranose	ZINC	-22.31	A
erithrose	ZINC	-22.12	A
L-threose	ZINC	-22.12	A
gliceraldehide	ZINC	-18.4	A
D-allose	aos	-27.22	B
L-mannose	ZINC	-23	B
L-glucose	BLGL	-22.6	B
D-beta-glucose	glu	-22.33	B
L-xylulose-5-phosphate	lx1	-34.21	outside
O1-methyl-glucose	MGL	-25.29	outside

As it can be observed from the docking scores the enzyme is predicted to prefer tagatose, psicose and sorbose.

This is in very good accordance with our previous findings and with the general organization of the operon in which ORF40 is situated. We have shown that the previously gene in this operon, *ofr30* is an aldehyde dehydrogenase . The dehydrogenation of ketohexoses by ORF40 leads to the formation of an 2C shorter coresponding aldehyde, which is further dehydrogenated by ORF39 and turned into an acid, which is further introduced into the Krebs cycle. The starting point of this small chain of reactions could not be an aldohexose, as the product of dehydrogenation would be directly an acid. We tested the purified his-tagged ORF40 enzyme for GFOR activity using the assay described by Zacharoiu but no activity could be detected.. We are currently undergoing test for sugar-dehidrogenase activity, but under the our experimental conditions no activity could be seen.

CONCLUSIONS

The computer predicted structure of ORF40 protein was build using the GFOR enzyme as model, following the same double-domain organization. Although the N-terminal arm is missing from the structure, our experimental data showed that the *Arthrobacter nicotinovorans* enzyme is an tetramer in solution and it does not contain tightly bound NADP. The docking experiments revealed two possible binding place for the ligands, one of which is in close proximity with the EKP conserved sequence. The docking score showed that the enzyme might prefer ketohexoses as substrates. Further investigation are required in order to establish the substrate of this enzyme.

REFERENCES

Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. & Rossmann, M.G. 1987, *J. Mol. Biol.* 198, 445-67.
Adam J., Kriz., Malinovska L,. Lameignere E., Imberty A., Koca J., Wimmerova M. (2008) FEBS Journal 275, 157.
Ausubel M Frederick, Brent Roger, Kingston E Robert , Moore D David,Seidman J G,Smith A John, Struhl Kevin *2002 Short Protocols in Molecular Biology,* John Wiley & Sons,.
Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliland, G., Weissig, H .2000. *Nat. Struct. Biol*. 7 Suppl, 957-9.
Brandsch Roderich 2006. *Appl. Microbiol. Biotechnol* 69, 493-498.
Eaton A. 2001 *Journal of Bacteriology* 183, 3689-3703.
Igloi G.L., Brandsch R., 2003 *J. Bacteriol.* 185, 1976-86.
Irwin, J.J., Shoichet, B.K.2005 *J Chem Inf Model* 45, 177-82.
Jackson, R.M., Gelpi, J.L., Cortes, A., Emery, D.C., Wilks, H.M., Moreton, K.M., 1992, *Biochemistry* 31, 8307-14.
Kingston, R.L., Scopes, R.K., Baker, E.N. 1996. *Structure* 4, 1413-28.
Kriz J., Adam J., Prokop M., Wimmerova M., Koca J., 2008 . FEBS Journal 275, 169.
Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E.,1982 *J,. Mol. Biol*. 161, 269-88.
Lambert, C., Léonard, N., De Bolle, X. Depiereux, E. 2002 *Bioinformatics* 18, 1250-6.
Mihasan M., Artenie V., Brandsch R.,2008. *FEBS Journal* 275, 285.
Nurizzo, D., Halbig, D., Sprenger, G.A., Baker,E.N. 2001 *Biochemistry* 40, 13857-67().
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. 2004 *J Comput Chem* 25, 1605-12.
Rosselló-Mora, R.A., Lalucat, J., García-Valdés, E. 1994. *Appl. Environ. Microbio*l. 60, 966-72.
Wang, J., Wang, W., Kollman, P.A., Case, D.A. 2006, *J. Mol. Graph. Model.* 25, 247-60().
Wiegert T., Sahm H., Sprenger G.A.1997 . *The Journal of Biological Chemistry*, 272, 13126-13133.
Zachariou, M. & Scopes, R.K.,1986 *J. Bacteriol.* 167, 863-9

.

Acknowledgments
Part of this work was supported by the CNCSIS TD-236 research grant.

1. „Alexandru Ioan Cuza" University, Iaşi, Biology Faculty, Molecular and Experimental Biology Depertament
2. National Institute for Biological Sciences, Iaşi, Romania
* mariusmihasan@gmail.com