Scaling up Protein Production for Functional Genomics: Study of new anti-fungal target proteins

Partners: Protein Expert - Structural and Genomic information laboratory

This project is granted by the French ministry of research and education N° 02 L 0194

Given the private status of this project we wish these results to remain confidential

There is growing recognition that a key way in which companies and basic science organizations are going to obtain maximum return on their investment in genomics is to include strong capabilities in the field of proteomics. The high throughput identification and quantification of protein expression, as well as studies of protein function, brings the researcher closer to the actual biology than studies of gene sequence or gene expression alone. Advances in analysis of protein-protein interactions, industrialization of protein expression, and the development of different approaches for more rapid and comprehensive approaches to protein function, all promise to pay off in accelerating diagnostic and therapeutic product development. Large scale protein expression projects have been initiated mostly in the context of "Structural Genomics" i.e. the massive production of new protein 3-D structure. Few years into these projects, the success rate has been unexpectedly low, with attrition rate below 20% from the number of initial targets to the number of effectively produced proteins. Large-scale attemps to produce proteins have clearly identified the main bottleneck at the level of producing sizable amounts of recombinant proteins in a soluble/stable form.

One of the objective of the project presented here, is to further explore the development of micro-titer plate level automated techniquesto improve the correlation with success when scaling up the production/purification processes. Our expertise in incomplete factorial experimental design (SAMBA software, Audic et al., (1997) *Proteins*) will also be put to use at this stage.

A second objective is to design, apply and validate new strategies of molecular engineering of the desired protein product that will strongly improve its soluble expression while maintaining its relevant biological and physico-chemical properties. This part of the project will merge Bioinformatics analysis (database searches, multi-alignment, structure prediction, ligand-binding motifs) at the 3-D and sequence level to guide directed mutagenesis of the target protein. This concept is at the center of Protein Expert business plan. The IGS laboratory has proven expertise in the area of motif detection and multiple alignment.

The third (but not the least) objective of this project is to generate a unique porte-folio of fungal proteins (to be valorized by Protein Expert) while obtaining the proof of the principles that will be developed during this study. Fungal infections have recently emerged as a growing threat to human health, especially in immuno-compromised patients. There is also a concern that the increasing use of antifungal drugs will lead to resistance, especially through the growing nosocomial (hospital-acquired) infections. Recent studies have documented resistance of *Candida* species to fluconazole and other azole drugs, used widely to treat patients with systemic fungal diseases. Our project thus timely concentrate on ~100 putative new antifungal drug target proteins, based on their evolutionary conservation. A comprehensive computational comparison of all available fungal genomes was thus an integral part of this project. A complete fungal protein database will thus be constituted and maintained during the course of this project.

We used S. cerevisiae, A. fumigatus, and C. albicans as reference genomes for target selection. we selected the target genes for protein expression in C. albicans and S. cerevisiae

genomes by taking into account their low number of introns. Out of the 150 genes selected, 54 from *Candida albicans* have been cloned and are currently studied. 14 of them are expressed as soluble during expression screening using incomplete factorial design (Abergel *et al.*, (2003) *J. Struct. Funct. Genomics*), 2 produced crystals, For one of them to crystal forms has been produced and correspond to two conformation states of the molecule depending on the absence or presence of a ligand in the binding site of the molecule. The 2 structures have been solved (one on FIP-BM30A, the other on ID29) and the bounded form is currently under study to identify the specific ligand of this target. For the *Saccharomyces cerevisiae* targets, out of 53 genes have been successfully cloned, 19 are expressed as soluble, 3 produced crystals. They have all been tested on FIP, datasets have been produced for one of them and the structure solving process is in progress using molecular replacement and homology modelling (CaspR server: http://www.igs.cnrs-mrs.fr/Caspr/index.cgi; Claude *et al.*, (2004) *Nucleic Acids Res.)*. The two other crystal forms do not produce usable diffraction. We are currently producing various crystals in the presence of co-factors and ligands in order to improve the diffraction. We also systematically perform desiccation process on the beamline in order to improve the diffraction (Abergel, (2004) *Acta Crystallogr. D*).

The CA3427 Structure

Description : The gene targeted in this study encodes a 299-amino-acids long protein with no significant homologues in the Protein Data Bank. However using the solved structure and the Dali Software ca3427 shares significant structural similarities to ion, and amino-acids binding proteins such as sulfate binding (1SBP, Jacobson & Quiocho, 1998); the *E. coli* Atp-Phosphoribosyltransferase (1H3D), the Lysyl-Ornithyl- Lysine binding protein, OppA (1B2H, Tame et al 1995) glutamate/glutamine-binding protein (1US5 Takahashi et al. 2004; 1GGG Sun et al. 1998), histidine binding protein (1HSL Yao et al. 1994) or lysine/arginine ornithine protein (1LST Oh et al. 1993). All these proteins are described as periplasmic binding proteins (PBPs). PBPs serve to scavenge or sense nutrient source in the environment to mediate substrat transport or induce chemotactism to nutrient source. The PBP gene module appears in multiple otherwise unrelated gene families througouth prokaryotes and eukaryotes. PBPs shown structural homologies with metabotropic and ionotropic glutamate receptors and of the bacterial regulatory protein lac repressor. PBPs may also served as chaperonnes (Richarme and Dantas Caldas, 1997). Substrates include a great variety of structure as amino-acids, oligopeptides, monosaccharides, oligosaccharides, oxyanions, cations and vitamins.

PBPs are stable monomeric proteins with a molecular mass of 22,000 to 59,000 Da, and have an unusual broad pH and ionic strengh optima for substrate binding. The binding affinity in terms of dissociation constant (Kd) is between 0.1µM for amino-acids and 1µM for sugars (Quiocho. 1990). The structure of several PBPs were known. They share an α/β secondary structural topology and have two similar globular domains connected by two or three short flexible peptide hinges. The two domains form a large cleft for substrate binding. Upon substrate binding, the two lobes twist and close thereby entrapping the ligand (Wolf and al. 1996).

The electronic density analysis of our structure allows to identify some extra density in the binding pocket. Further studies are necessary to identify this compound.

Data collection:

Crystal 1:

Fip BM30A. ESRF (Grenoble, France)

This structure was solved using the MAD method and the selenium (2 selenomethionines). Mosflm, Scala, Solve, Resolve and CNS were used to index, scale, solve, built and refine the structure
resolution: 2.34 to 25Å (Table 1)
A cavity is present in the protein binding site where residual density is observed (Figure 1: molecular surface of the complexed structure top, cavity marked by an arrow).

Crystal2:
ID29. ESRF (Grenoble, France)
This structure was solved by molecular replacement using the complexed form.
Mosflm, Scala, CaspR and CNS were used to index, scale, solve and refine the structure
Resolution: 2. to 25Å
No cavity inside the molecule and no ligand (Figure 1: molecular surface of the free structure bottom, binding pocket marked by an arrow)

Table 1 X-ray data collection and refinement statistics

Data collection				
Data set		MAD		
	λ1	λ2	λ3	λ4
Beam line		ESRF/BM30		ESRF/ID29
Wavelength (Å)	0.9798	0.9799	0.9537	0.9756
Space group		$P2_12_12_1$		$P2_12_12_1$
Unit cell dimensions(Å)		a=42.6 b=66.8 c=114.0		a=41.4 b= 65.7 c= 128.2
Resolution range (Å)	45-2.5	45-2.5	45-2.34	46-2.0
Observations	61514	61514	76662	147779
Unique reflections	10802	10802	13379	17895
Multiplicity[1]	5.5 (2.4)	5.5 (2.4)	5.5 (3)	8.0 (6.0)
Completeness[1]	98.8 (89.6)	98.8 (89.6)	99.1 (94.4)	76.7 (72.6)
$\langle I / \sigma I \rangle$[1,2]	7.6 (3.9)	7.6 (3.9)	4.1 (3.0)	6.5 (2.7)
R_{sym} (%)[1,3]	6 (16.5)	2.6 (9.1)	7 (22.7)	9.6 (26.5)
Refinement				
R_{cryst} (%)[4]	0.208			0.189
R_{free} (%)	0.256			0.239

[1] values in parentheses are for the highest resolution shell.

[2] $\langle I / \sigma I \rangle$, is the mean signal to noise ratio, where I is the integrated intensity of a measured reflection and σ is the estimated error in the measurement.

[3] $R_{sym} = \sum_h \sum_i \left| I_{h,i} - \langle I_h \rangle \right| / \sum_h \sum_i \left| I_{h,i} \right|$, where I is the integrated intensity of reflection h having i observations and $\langle I_h \rangle$ is the mean recorded intensity of reflection h over multiple recording.

[4] $R_{cryst} = \sum \left| |F_o| - |F_c| \right| / \sum |F_o|$, where F_o are observed and F_c calculated structure factor amplitudes. R_{free} is calculated from a randomly chosen 9.9% of reflections.

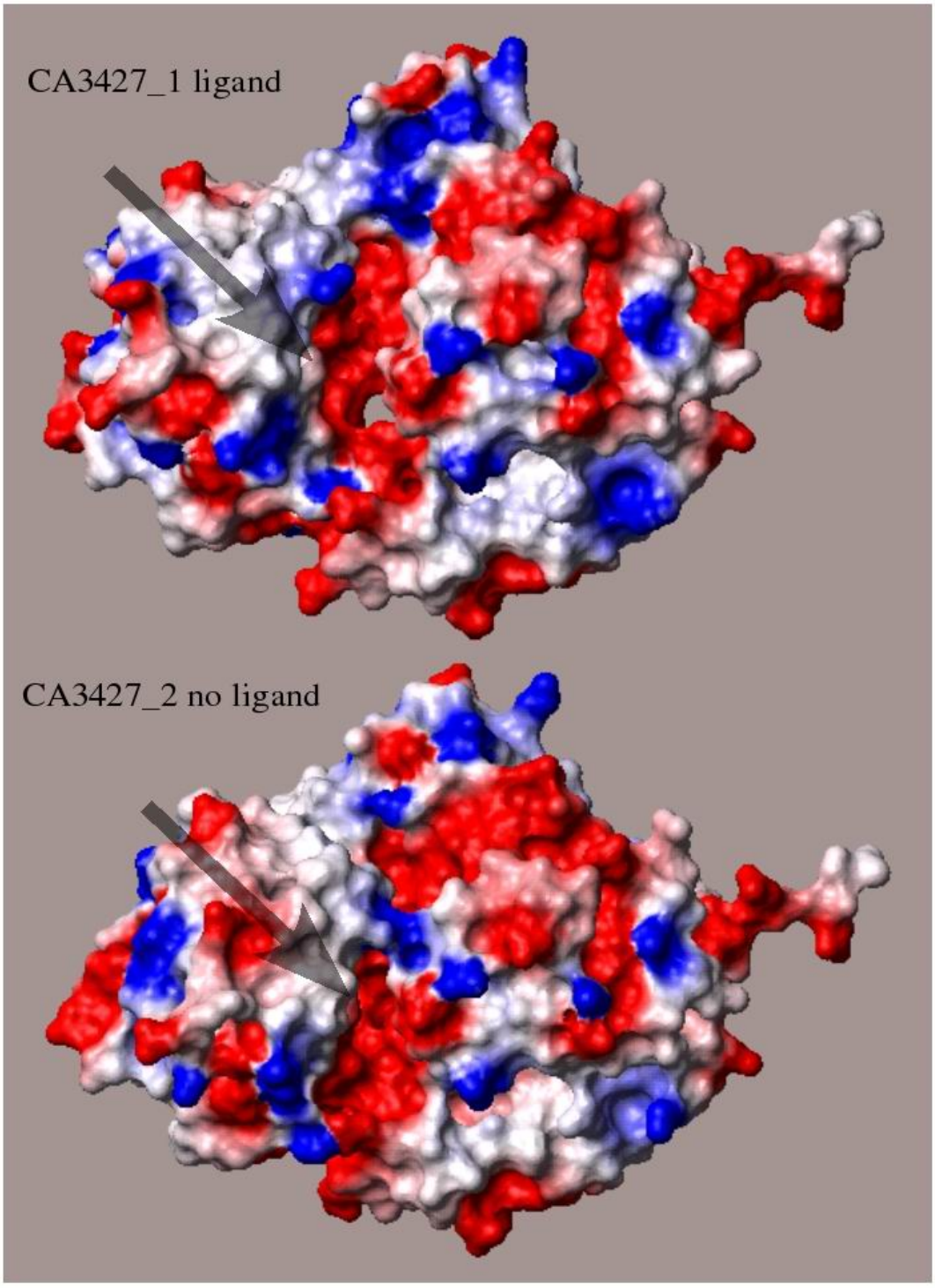

Figure 1

References :

Abergel C., Coutard B., Byrne D., Chenivesse S., Claude J-B., Deregnaucourt C., Fricaux T., Boutreux C., Jeudy S., Lebrun R., Maza C., Notredame C., Poirot O., Suhre K., Varagnol M. & Claverie J-M. 2003 Structural genomics of highly conserved microbial genes of unknown function in search of new antibacterial targets. *J. Struct. Funct. Genomics*; **4**:141-157.

Abergel C. 2004) Spectacular improvement of X-ray diffraction through fast desiccation of protein crystals. *Acta Crystallogr. D.*; **60**:1413-6.
Audic S., Lopez F., Claverie J.-M., Poirot O. and Abergel C. 1997 SAmBA: an Interactive Software for Optimizing the Design of Macromolecule Crystallization Experiments. *PROTEINS*.**29**:251-256.
Claude, J-B., Suhre K., Notredame C., Claverie J-M. & Abergel C. 2004 CaspR: a web-server for automated molecular replacement using homology modelling. *Nucleic Acids Res.*; **32**:W606-9.
Felder CB, Graul RC, Lee AY, Merkle HP and Sadee W. 1999. The Venus flie trap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. *AAPS PharmSci* 1(2) article 2.
Oh BH, Ames GF and Kim SH. 1994. Structural basis for multiple ligand specificity of the periplasmic lysine- arginine- ornothine-binding protein. *J. Biol. Chem*. 269: 26323-26330.
Oh BH ; Pandit J, Kang CH, Nikaido K, Gokcen S, Ames GF and Kim SH. 1993. Three-dimentionnal structures of the preriplasmic lysine/arginine/ornithine binding protein with and without ligand. *J. Biol. Chem*. 268: 11348-11355.
Quiocho FA. 1990. Atomic structures of periplasmic biding proteins and the high-affinity active transport systems in bacteria. *Philos Trans R Soc Lond B Biol Sci.* 326(1236):341-351.
Richarme G and Caldas TD. 1997. Chaperone properties of the bacterial periplasmic substrate-binding proteins. *J. Biol. Chem*. 272:15607-15612.
Sun YJ, Rose J, Wang BC and Hsiao CD. 1998. The structure of glutamine-binding protein complexed with glutamine at 1.94A resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol*. 278:219-229.
Takahashi H, Inagari E, Kuroishi C and Tahirov TH. 2004. Structure of the Thermus thermophilus putative periplasmic glutamate/ glutamine-binding protein. *Acta crystall*. D60:1846-1854.
Wolf A, Lee KC, Kirch JF and Ames GFL. 1996. Ligand-depend conformaional plasticity of the periplasmic histidine-binding protein HisJ. Involvement in transport specificity. *J. Biol. Chem*. 271:21243-21250.
Yao N, Trakhanov S and Quiocho FA. 1994. Refined 1.89-A structure of the histidine-binding protein complexed with histidine and its relationship with many other active transport/ chemosensory proteins. *Biochemistry* 33(16):4769-4779.