

Structural Genomics of “ORFan” Genes:

The complete sequences of more than 60 microbial and four eukaryote genomes are already available in the public domain and many more genomic projects are in progress throughout the world. Despite this accumulation of data, newly sequenced microbial genomes continue to reveal up to 50% of functionally uncharacterized “anonymous” genes, including a significant number of putative ORFs without clear similarity to previously determined protein-coding sequences. This set of unique - apparently species specific - sequences are referred to as “ORFans”. Most genes found in databases have only been predicted by computer methods and have never been experimentally validated. While theoretical evolutionary arguments support the reality of genes when homologues are found in a variety of distant species, this is not the case for ORFans. The biochemical and structural analysis of ORFan products is thus both of evolutionary and functional interest.

Our laboratory started to study 25 *E. coli* orfan genes in 2000 [1], and used RT-PCR to demonstrated that transcripts could be identified for a total of 19 orphan genes, with 2 genes found to be expressed in only one of the two tested growth conditions (exponential and stationary phase). Our results suggest that a vast majority of *E. coli* ORFs presently annotated as “hypothetical” correspond to bona fide genes. By extension this implies that randomly occurring “junk” ORFs have been actively counter selected during the evolution of the dense *E. coli* genome. The same year, we reported [2] the cloning, expression and crystallization of *E. coli* ORFan *ykfE* gene, followed by the functional characterization of the encoded protein [3]. Under physiological conditions, the protein is an homodimer with a strong affinity for lysozyme, as revealed by co-purification. Activity measurements and fluorescence studies demonstrated that the *ykfE* gene product is a potent lysozyme inhibitor, with a 1:1 stoichiometry. To denote this newly assigned function we changed the gene name *ykfE* into *ivy*, for “Inhibitor of Vertebrate LYsozymes”. Crystals of Ivy as well as crystals of its complex with Hen Egg White Lysozyme have been produced and the 3-D structure determined [PDB1GPQ]. Using the structure of the IVY/HEWL complex we were able to identify potential orthologues genes in other organisms. The corresponding *Pseudomonas aeruginosa* gene was cloned, expressed, crystallized and functionally tested for its activity as an inhibitor of type-C lysozymes. The structure of the complex IVYP/HEWL [PDB 1HKE] was also elucidated allowing to demonstrated the existence of a new bacterial family of proteins responsible of the type C lysozyme inhibition. This family of sequences are intriguing due to the high rate of divergence in their sequence and the conservation of their anti-lysozyme function. The main difference between the *E. coli* and the *P. aeruginosa* proteins is the dimeric nature of the *E. coli* protein while the *P. aeruginosa* one is a monomer.

References

1. Alimi J.-Ph., , Poirot O., , Lopez F., , and Claverie J.-M., (2000) Reverse Transcriptase-Polymerase Chain Reaction Validation of 25 “Orphan” Genes from *Escherichia coli* K-12 MG1655. *Genome Res* 10: (7) 959- 966.
2. Abergel C., Monchois V., Chenivresse S., Jeudy S. and Claverie J.-M. (2000) Crystallization and preliminary crystallographic study of b0220, an “ORFan” protein of unknown function from *Escherichia coli*. *Acta Cryst. D56* : 1694-1695
3. Monchois V., Abergel C., Sturgis J., Jeudy S. and Claverie J.-M. (2001). *Escherichia coli* YkfE “ORFan” gene encodes a potent inhibitor of C-type lysozyme. *J. Biol. Chem.* 276 : (21) 18437-18441