



Structural Genomics of ORFan Genes from *Halobacterium* NRC-1

Shmuely, H.¹, Chehanovsky, N.¹, Dahan, I.¹, Fischer, D.², Eichler, J.¹ and Shaanan, B.¹

¹ Department of Life Sciences, Ben Gurion University

² Department of Computer Sciences, Ben Gurion University

With the sequencing of the human genome and the completion of other genome projects, it has become clear that functional knowledge is lacking for proteins encoded by the majority of ORFs. Identifying shared structural motifs offers a strong predictive tool for describing the function of a protein. Unfortunately, our ability to predict the structure of a protein from its sequence alone remains limited. This obstacle is further magnified by the limited number of solved protein structures available for comparative purposes. Structural genomics tries to fill in this gap by solving the structure for as many proteins as possible. In this project we are attempting to solve the structure of ORFs for which no sequence homologs exist (i.e. ORFans) from the halophilic archaeon *Halobacterium sa-*

linarium. In doing so, we hope to reveal known folds in proteins with no sequence similarity or to describe novel protein folds. Accordingly, we have identified 42 ORFans that can be divided into 15 paralogous groups. The structure of each protein was predicted using different bioinformatics web servers. Upon cloning these genes into *Escherichia coli*, encoded proteins have been expressed and purified. After refolding in high salt conditions, the structure of the bacterially-expressed proteins will be determined using X-ray crystallography. In addition, those genes being heterologously expressed in the haloarchaeon *Haloferax volcanii* will also be crystallized. Based on structural information revealed, we will try to identify the function and biochemical parameters of the proteins.