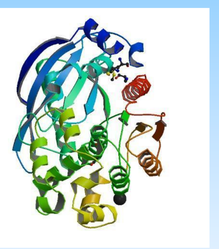

The Biochemical Characterization of the Protein Data Bank Structure 4EZI by *In Silico* and *In Vitro* Quantitative Methods

Kaitlin Hart[1], Alison Horn[1], Gregory Dodge[1], Herbert Bernstein[2], Paul Craig[1]
[1]Chemistry and Material Science, Rochester Institute of Technology, Rochester, NY, [2]Mathematics and Computer Science, Dowling College, Oakdale, NY

Abstract

This study is part of the Structural Biology Extensible Visualization Scripting Language, SBEVSL, project, the purpose of which is to propose and confirm the function of uncharacterized protein structures, in the Protein Data Bank, by use of *in silico* and *in vitro* quantitative techniques. For over 3000 structures in the PDB, no function has been assigned. These structures were first screened with the PyMOL molecular graphics system, along with the ProMOL plugin, which compares query structures against a library of enzyme active sites. The most promising alignments were further refined by sequence and structure comparison using BLAST and Dali. A structure of considerable interest is presented here: 4EZI. When 4EZI was queried with PyMOL/ ProMOL, the best alignment was with the dipeptidyl peptidase IV, 1ORV. After analysis with BLAST and Dali, the alignments revealed 4EZI matched closely to the esterase family, specifically 3H2H. This match to the esterase family was confirmed experimentally. The protein was purified using methods of transformation and column chromatography. Then, the predicted molecular weight was confirmed using SDS-PAGE. Protein concentrations were determined by a BCA assay and enzymatic activity was confirmed using *p*-nitrophenol esters of 2C and 10C carboxylic acids. In conclusion, 4EZI has been characterized as an esterase based both on *in silico* and *in vitro* quantitative methods. Funding was provided by RIT, Dowling College, and NIGMS 2R15GM078077-02 & 3R15GM078077-02S1.

Introduction

As of February 27, 2014, there were 3975 proteins with unknown functions in the Protein Data Bank [PDB; 1]. ProMOL, a plugin for PyMOL, performs template-based alignments of query protein structures with enzyme active sites, as described in the Catalytic Site Atlas. Results are reported as Levenshtein distance [a residue-for-residue comparison; [2], RMSDs for all atoms, C_α, and $C_\alpha C_\beta$ and visual alignments of active sites in ProMOL. Typical results for an alignment are shown in Figures 1-3.

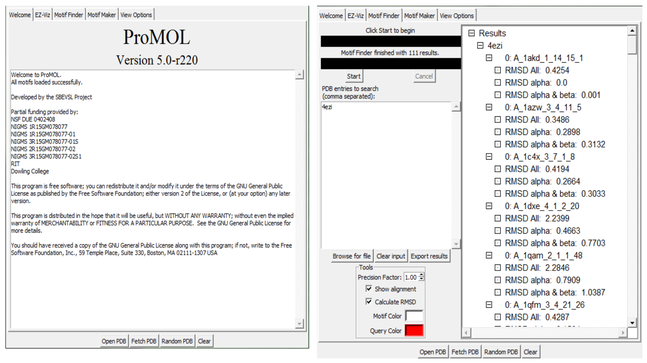

Figure 1. The ProMOL welcome page and Motif Finder tab. To analyze a PDB entry, the PDB ID is entered into the search box under the Motif Finder tab and the Start button is clicked. The RMSD values can be determined by selecting Calculate RMSD in the start options. The lower the RMSD value, the closer the alignment of the active sites of the motif and the query.

During the summer of 2013, >3000 of the proteins with unknown functions in the Protein Data Bank were screened through ProMOL. A "good hit" is considered to be a protein which aligns with three or more residues of another known protein, and has an RMSD value below 2.5 for three residues or below 4 for more than three residues. Two other alignment programs, BLAST and DALI, are also used to narrow down the exact function of the protein. BLAST [3] searches databases for sequences that are homologous to the target, while DALI [4] searches the PDB for full backbone alignments of proteins. 4EZI, a protein with no known function in the PDB, was identified as a good hit, and further pursued to discover that it is part of the hydrolase family and more specifically an esterase or lipase.

Acknowledgments

We would like to thank Dr. Lea Michel for allowing us to use her lab. Funding was provided by RIT, Dowling College, and NIH (2R15GM078077-02 and 3R15GM078077-02S1). Lastly, we would like to thank the Joint Center for Structural Genomics for making the plasmid containing the gene for PDB entry 4EZI available to the scientific community.

Methods and Results

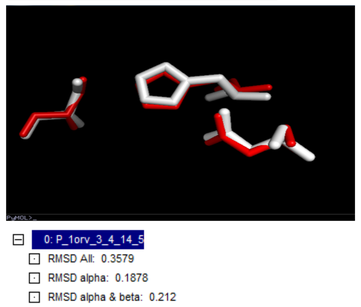

Figure 2. ProMOL alignments of 1ORV (white) with 4EZI (red) showing active site comparison. 1ORV is a hydrolase. The RMSD value of 0.3579 indicates that the active sites of the motif and the query are very close. The residues in the alignment are aspartate (ASP- 708), histidine (HIS-740) and serine (SER-630) of 1ORV and aspartate (ASP-334), histidine (HIS-337), and serine (SER-195) of 4EZI.

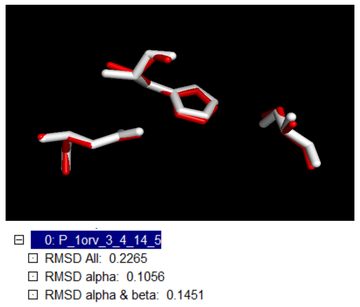

Figure 3. ProMOL alignments of 1ORV (white) with 3H2H (red) showing active site comparison. The RMSD value is 0.2265, suggesting that 3H2H is also a homologue of 4EZI. The residues in the alignment are aspartate (ASP- 708), histidine (HIS-740) and serine (SER-630) of 1ORV and aspartate (ASP- 336), histidine (HIS-337) and serine (SER-176) of 3H2H.

A list of proteins with unknown functions was screened against a library of >300 motif templates of proteins with known function in ProMOL. The PDB entry 4EZI aligned well with 3 residues of 1ORV which is classified as a dipeptidyl peptidase, a type of hydrolase.

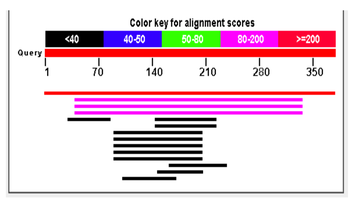

Figure 4. BLAST alignment of 4EZI. Based on sequence comparison, 4EZI matched best with the esterase family.

Description	Max score	Total score	Query cover	E value	Max ident	Accession
Chain A, Crystal Structure Of A Hypothetical Protein (Lpg1103) From Legionella Pneumophila Sub	748	748	100%	0.0	97%	4EZI_A
Chain A, Crystal Structure Of G231f Mutant Of The Rice Cell Wall Degrading Esterase Lipa From X	91.3	91.3	78%	3e-20	27%	3H2H_A
Chain A, Crystal Structure Of N228w Mutant Of The Rice Cell Wall Degrading Esterase Lipa From	90.5	90.5	78%	6e-20	27%	3H2I_A
Chain A, Crystal Structure Of A Rice Cell Wall Degrading Esterase Lipa From Xanthomonas Oryza	90.1	90.1	78%	1e-19	27%	3H2G_A

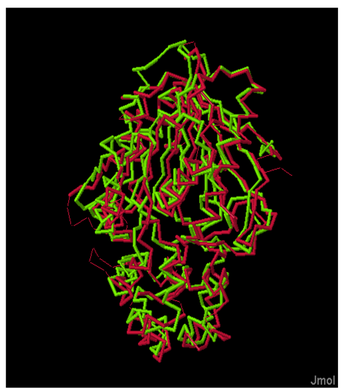

Figure 5. 3D alignment of 4EZI (green) with 3H2H (red, a known esterase). Created via Jmol and the Dali Structural Alignment Server.

PDB entry 4EZI was then screened with NCBI's BLAST algorithm to determine substrate specificity for the active site. The results showed that 4EZI had high structural similarity with proteins of the esterase subclass of hydrolases (Figure 4), with the top hit being the PDB entry 3H2H. Following BLAST, the DALI server was used to visualize the 3D alignment between 4EZI and the known esterase 3H2H (Figure 5). The active sites in both showed strong similarities (Figure 3).

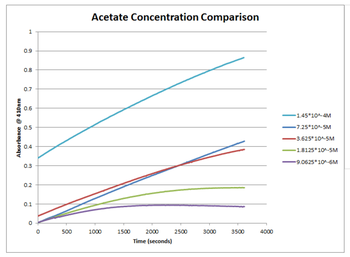

Figure 6. p-Nitrophenyl Acetate activity assay . Since 4EZI can hydrolyze the short ester of a 2-carbon carboxylic acid, it is acting as an esterase.

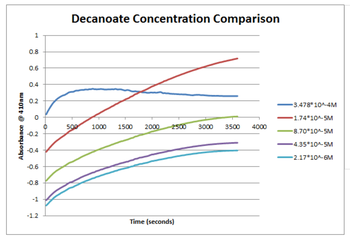

Figure 7. p-Nitrophenyl Decanoate activity assay. 4EZI possesses also demonstrates lipase activity, as evidenced by the hydrolysis for the ester of a 10-carbon carboxylic acid.

Activity assays were used for *in-vitro* characterization of 4EZI [5]. Both short chain and long chain p-nitrophenol esters were used as substrates to be hydrolyzed by 4EZI. Since 4EZI hydrolyzed the esters from both 2C and 10C carboxylic acids, further kinetic characterization is necessary.

Conclusion

- ProMOL was used as a tool for the initial insights into enzyme classification of 4EZI as a member of the hydrolase family.
- BLAST and DALI were then used to provide evidence that 4EZI belongs to the esterase subclass.
- *In vitro* methods which followed computer analysis, confirmed that 4EZI is an esterase or lipase, because 4EZI cleaved p-nitrophenyl esters of linear carboxylic acids.
- While 4EZI is clearly a hydrolase, it is unclear whether it is an esterase or a lipase.

Future Plans

- *In vitro* testing to provide more in depth kinetic analysis
- Further *in silico* characterization to more provide more insight into substrate specificity for 4EZI
- Autodock analysis for substrate fitting

References

1. Bernstein, F C, T F Koetzle, G J Williams, E F Meyer Jr, M D Brice, J R Rodgers, O Kennard, T Shimanouchi, and M Tasumi. "The Protein Data Bank: a Computer-based Archival File for Macromolecular Structures." *Journal of Molecular Biology* 112, no. 3 (May 25, 1977): 535–542.
2. Levenshtein, Vladimir I. (February 1966). "Binary codes capable of correcting deletions, insertions, and reversals". *Soviet Physics Doklady* **10** (8): 707–710.
3. Altschul, S F, T L Madden, A A Schäffer, J Zhang, Z Zhang, W Miller, and D J Lipman. "Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25, no. 17 (September 1, 1997): 3389–3402.
4. Holm, Liisa, and Päivi Rosenström. "Dali Server: Conservation Mapping in 3D." *Nucleic Acids Research* 38, no. suppl 2 (July 1, 2010): W545–W549. doi:10.1093/nar/gkq366.
5. Shang, Yu-Shuan, Xian-En Zhang, Xu-De Wang, Yong-Chao Guo, Zhi-Ping Zhang, and Ya-Feng Zhou. "Biochemical Characterization and Mutational Improvement of a Thermophilic Esterase from Sulfolobus Solfataricus P2." *Biotechnology Letters* 32, no. 8 (April 13, 2010): 1151–1157. doi:10.1007/s10529-010-0274-0.