KBM Journal of Science Education (2010) 1 (1): 7-12 doi: 10.5147/kbmjse/2010/0013

# Teaching Principles of Enzyme Structure, Evolution, and Catalysis Using Bioinformatics

Pablo Sobrado

Department of Biochemistry, Virginia Tech, Blacksburg, VA 24060, USA

Received: February 10, 2010 / Accepted: July 25, 2010

### Abstract

he study of enzyme function has greatly benefited from the increase in the number of gene sequences and the development of bioinformatics tools that allow for the rapid search of these vast sets of data, the integration of their results, and the linking of information regarding enzyme families, function, chemical mechanism, and structures of homologous enzymes. Similarly, the development of freely available threading programs, which provide a 3-dimensional (3D) structural model of an enzyme of interest, have facilitated structural and functional studies that contribute to the understanding of the chemical mechanism of a target enzyme. These tools can also be used to teach basic principles of chemical biology, molecular evolution, and enzyme structure and function in silico. A training exercise for senior undergraduate students interested in working with enzymes is presented in this article. The exercise includes the use of bioinformatics programs to identify a target protein from a partial gene sequence, identify its family, and obtain a 3D model. The second part of the exercise consists of searching the literature database to provide a plausible chemical mechanism based on the available data from homologous enzymes.

**Keywords:** Attitudes, Beliefs, Teaching and Learning, Introductory Physics, Curriculum.

**Abbreviations:** 3D, three dimensional; PSI-BLAST, position specific iterative basic local alignment search tool; TMO, tryptophan 2-monooxygenase; At, Agrobacterium tumefaciens; Ps, Pseudomonas savastanoi; Phyre, protein homology/analogy recognition Engine; 3D-PSSM, 3D-position specific scoring matrix; PDB, protein data bank.

## Introduction

The increase in the number of gene sequences deposited in databases, along with advances in computational biology, have allowed the identification of protein families based on primary sequence relationships (Altschul et al., 1997). It is now possible to identify the corresponding family of a target protein based on similar functional regions or domains at the primary sequence level. Furthermore, the identification of conserved residues essential for catalysis and structural domains responsible for the regulation of enzyme activity can be identified using bioinformatics tools. This information is a major step towards understanding the function of the protein of interest. It is clear that bioinformatics has greatly aided the research of enzyme function and has become an essential tool in today's biochemistry laboratories. Thus, basic knowledge of the computational tools available to analyze enzyme function is essential for biochemistry students.

In most biochemistry programs at doctoral/research universities, students are required to do one or two semesters of research for their senior thesis/independent research. Students taking part in these research projects have completed most of their core biochemistry courses and, therefore, have a strong theoretical knowledge of general chemistry, organic chemistry, physics, and biochemistry (Voet et al., 2003; Zimmerman, 2003; Zimmerman et al., 1993). The main goal of requiring students to perform research projects is to teach them how to use knowledge acquired in the classroom to solve problems in the laboratory and to provide hands-on research experience to prepare them for industry/academic jobs or graduate/professional school. It is important that students begin applying the principles of biochemistry to their research as quickly as possible, since their time in the laboratory is limited. Advances in Web-based, free bioinformatics tools provides an opportunity for students to put

XBM Journal of Science Education - ISSN 1949-145X. Published By KBM Scientific Publishing, LP (www.kbm-scientific-publishing.org)

into practice what they have learned in the classroom within the context of biochemistry; specifically, the study of the mechanism of enzyme-catalyzed reactions. A simple one-week research exercise can be used to teach undergraduate students how to integrate what they have learned in terms of protein structure, evolution, and chemical catalysis. In this exercise, students follow a series of steps that leads them from a partial gene sequence to the proposal of a chemical mechanism for the target enzyme (Table 1). During this process, students learn to use literature databases and bioinformatics tools available on the internet (Table 2). This exercise can be modified to fit research being conducted in many laboratories and can easily be adopted as a training exercise for new students or as a homework assignment in an upper-level Biochemistry course.

### **Background Information**

The purpose of the exercise presented here is to allow the student to become familiar with the use of bioinformatics to study molecular evolution, enzyme structure, and catalysis. The instructor should begin by providing a short description of the different bioinformatic programs available to the students and their web addresses. Then, students must identify a protein from a partial (or modified) gene sequence. To minimize possible confusion, the gene sequence should translate into a protein product that will yield a high score and the lowest E-value1 in the PSI-BLAST search. The protein product should not have a 3D structure, but should contain significant primary sequence homology to a related enzyme with a known 3D structure, such that a 3D model can be obtained. Finally, the chemical mechanism for the particular enzyme should not be well understood; however, the mechanism of the homologous enzyme must be known in significant detail.

As an example, we use the truncated gene for tryptophan 2-monooxygenase (TMO) from Agrobacterium tumefaciens (AtTMO) to perform this exercise. TMO catalyzes the oxidative decarboxylation of tryptophan to indoleacetamide, carbon dioxide, and water (Figure 1). This enzyme is 49% identical and 69% similar to TMO from Pseudomonas savastanoi (PsTMO). Studies with PsTMO using the same tools listed in Table 1 allow for the identification of this enzyme as a member of the amine oxidase family [5]. Additionally, a 3D model structure can be obtained for PsTMO based on the structure of L-amino acid oxidase (LAAO), also a member of the amine oxidase family of enzymes. Further investigation into the role of amino acids in the active site of PsTMO support the assignment of PsTMO as a homologous enzyme to LAAO [5-7]. Ultimately, students are expected to be able to identify AtTMO as a member of the amine oxidase family of enzymes and build a 3D model using various bioinformatics tools. After searching the literature database, they should be able to obtain information about the reaction mechanism of homologous enzymes, in particular PsTMO and LAAO. The students are then asked to use this information in addition to the structural information obtained from the bioinformatics exercise to postulate a mechanism for the reaction catalyzed by AtTMO.

#### From DNA Sequence to Protein Primary Structure

In the initial part of the experiment, the student is given a partial gene sequence of a gene of interest. Their first step is to determine the open reading frame (ORF). We use the program Translate, which is available at the ExPASy proteomics server (Table 1). This program translates the gene sequence into all six reading frames. Students should provide explanations of why the program provides these results and explain why they selected a particular sequence. The student must choose the correct



You are here: ExPASy CA > Tools > DNA -> Protein > Translate

### Translate Tool - Results of translation

Please select one of the following frames:

#### 5'3' Frame 1

I-TLRNSCLVPCRTSTPCTTTVLSWLVPVVTWVPCRLTVLVTLVSLSLVLVLVVWLLLAN CCVLVLVSWLFSKLTRLVLVVVCCLSLSTRTTRTCWLKWVLCVSRRLRLCTTTWNVSVS TKNSVSLPLPSRTRVLLTPNCTTVVNVTTGRLVKNRRLCSVVFTKVGSLCCVTVTCWKV VLWLLKWTSLLRCGTVPSTTSLLPLKLSLNCLSTVKFSVRLVSVPVVGTLTSRTLCWK SSVLL-PTATNSRTLSWKSCVLSPLLTTTVVSLVVLSCRCVCGPTPRPVWLTGRTVP LFVL-TVVLPVSTVFRWVT-SNWVPRLPVFLLTVLVLL-PSKVPPRNSLVSLVLSLPLTV LLRSLSVWLRVVKPT-PVLLNCSC-PTNSGSPTVCRLPSLTVSLVVFTAWTTNRTTWT VVVLFCCLLTFGKTTTTNCCLLLTKNSVANVSLKNWLPSTRNKLVTWFRLLVTTPLFCTT TG-PTRTLLVLSLGKTRTSTVLSNSTSRVTS-PSSCTTTCCHLTRLPTRTCTSLVNL TLSTVVGPKVLSPPV-TLLPLLSNLWTVPC-STRTR-PTRVRVTT

#### 5'3' Frame 2

YERSGIPVSYPAVHRHPVRLRCFPGSYRWSPGYPAG-RCS-RSYRYRWCWCWWSGCCLRT AACWCSSAGCFRS-PGSCWWSSAVSVYQPGPPAPAG-NGCYAFPAVSGCSVPLPGTFRYR QXIRYLYLYHLPGPGCC-HRTALPW-TLLLAGW-KTAGSVPSCLRRLAVSAA-RLPAGRW FSGCSAGHRCSAVGRPYLLRLRCYL-SFR-TVFLPP-SFRSGWFYRWLGL-LPELYAGN LPCCYDQLRRIPELYPGNPACCYHRC-RPPPWYRWFSAAAASVDRHPDRYGSLAGRYL CSCSERWCYPFQRCSAG-PDQTGYPGYRCFC-LWFWCCCSDHQRYHPEILSCYRCYLSPF CSDLYRSGSGS-NPPDRFF-TVHADP-QILDHQRSAGYHPV-RFRSWCLLPGLRTGQPGR SWCCSAVLHLGRRRSQTAVCC-QKTALRTFR-RTGYLPFGTGSSPGSGCW-LRHLCSAPR LADRPALFWCYRLGRRTVPVRCFQIQLPG-RHPDPAAVLQLPAG-QPVYRPVPVLRW-IL LFPRWLDRRCYHHRSERCYRCYQISGRYPADQPEPADRPGSGLQL

#### 5'3' Frame 3

MNAPEFLSRTLPYIDTLYDYGAFLARTGGHLGTLPADGARDARIAIVGAGVGGLVAACEL LRAGARQLVVFEANPARVGGRLLSQSINPDHPHLLAEMGAMRFPPSQAALYHYLERPGID KKFGISTSTTPEDEGVVDTELHYRGKRYYMPAGKKPPALFRRVYEGMQSLLRDGYLLEG SLVAPLDIAAPLWDDRFYDFVATSKAFAKLSFYHREVFQQVGFGGWDSDFPNSMLEI FRVVMTNCDEFPNSILEILRVVITAADDHHRGIVGGSQQLPLRLWTDTPTGMAHWPDGTS VRALNGGATRFNGVPLGNLIKLGTPVTGVSVDSGSGVVALTIKGTTQKFSRAIVATSHRS AQISIGLAPGRETHLTGSSKLFMLTRDKFWITNGLPVTIQSGFARGVYCLDYEDNLDG RGVVLLSYTWEDDAHKLLSVADKKQRCERFVEELATFHPELARHLVPAAGNYDTSVLHHD WLTDPHSSGAIAWENEFYQYGAFKFNFPGDDLTQQLYVNYLLANNPSTDPYLYFAGESY SFHGGWTEGAITTGLNAATAVIKSLDGTLLINPKFLTDPGGGYNY

**Fig. 1.** Results of samples from the Translate program. The program provides the translation of the DNA of interest in all 6 reading frames. Only the translation of the leading strand is shown.

ORF (in this exercise, the one with no stop codons) and then take the predicted amino acid sequence and perform a PSI-BLAST (Figure 1) (Altschul et al., 1997).

### Determination of Protein Function from Sequence Homology

Since in the beginning of the exercise the students are given only a partial sequence of the gene, once the PSI-BLAST is performed, the target protein sequence must be identified and the complete amino acid sequence obtained. In addition, the students should take advantage of the conserved protein domains annotation that is given in the PSI-BLAST results. A link to the Protein Family database (Pfam) is also available from the conserved protein domain report page. In this site, more information about the members of the protein family can be found, including the related superfamily, sequence alignments, available 3Dstructures, and more (Figure 2). At this point, the student has very important information about the enzyme of interest: the protein family and its predicted function. In this example, the student identifies the protein as tryptophan 2-mooxygenase, based on the E-value and percent homology (Figure 3). The Pfam report shows that TMO is a member of the amino oxidase family of enzymes. The student should recognize (or be led by an instructor/mentor) the structure-function relationship between families of enzymes.

### From Amino Acid Sequence to 3-D Structure

Homologous enzymes with similar functions usually share similar structures (Gerlt and Babbitt, 2000, 2001). Since the struc-



Fig. 2. Results from the BLAST program showing the conserved domain of the protein of interest.

Home Reci	ent Results Saved S	Basic Local Alignment Search Tool trategies Help					My NC [Sign.]
/ BLAST/ blast	tp suite/ Formatting Rea	sults - 3EFRDT9B014					
dit and Resut	mit Save Search Str	ategies > Formatting options > Download					
		PSI blast Iteration					
otein Seque	ence (459 letters)						
	. ,						
	ID lcl 57002		ase Name				
	ion None vpe amino acid	D			SenBank CDS translat ental samples from W		ssProt+PIR+PRF
Query Leng				BLASTP 2.2.24+ >		os projecta	
Other reports	s: >Search Summary	Taxonomy reports] [Distance tree of results] [Multiple alignment]					
Graphic St							
Descriptio	ns						
Legend for link	is to other resources:	UniGene 🖪 GEO 🖸 Gene 🧧 Structure 🛄 Map Viewer 👫 PubChem BioAssa	iv.				
NEW - align	ment score below the th	reshold on the previous iteration	·				
	ment was checked on th						
	iteration 2 with max 500						
Run PSI-Blast	iteration 2 with max 500						
▼ Seque	ences producing sig	nificant alignments with E-value BETTER than threshold					
▼ <u>Sequ</u>	ences producing sig	nificant alignments with E-value BETTER than threshold Description	Max score	Total score	Query coverage	E value	Links
▼ <u>Sequ</u>			Max score	Total score 951	100%	E value	Links
V Seque	Accession	Description tryptophan monoxygenase [Agrobacterium tumefaciens] tryptophan monoxygenase [Agrobacterium tumefaciens str. C58]	<u>951</u> <u>946</u>	951 946	100% 100%	0.0 0.0	
V Seque	Accession PBAA76346.1 PAAD30489.1 NP 396528.1	Description tryptophan monooxygenae (Agrobacterium tumeFaciens) tryptophan monooxygenase (Agrobacterium tumeFaciens str. C58) tryptophan 2-monooxygenase (Agrobacterium tumeFaciens str. C58) >pb)AAKK	951 946 945	951 946 945	100% 100% 100%	0.0 0.0 0.0	Unks
V Seque	Accession MRAA76346.1 MAD30489.1 MP_396528.1 MCAB44640.1	Description tryptophan monooxygenase (Agrobacterium tumefaciens str. CSB) tryptophan monooxygenase (Agrobacterium tumefaciens str. CSB) tryptophan 2-monooxygenase (Agrobacterium tumefaciens)	951 946 945 945	951 946 945 945	100% 100% 100% 100%	0.0 0.0 0.0 0.0	G
V Sequi	Accession MRA26346.1 MAD30489.1 MNP 396528.1 MCAB44640.1 MNP 053423.1	Description trystophan monosoryense [Aprobactinum tumificients] rystophan monosoryense [Aprobactinum tumificients at: C38] rystophan 2-monosygense [Aprobactinum tumificients at: C38] - sp3/AAKK trystophan 2-monosygense [Aprobactinum tumificients at: C38] - sp3/AAKK trystophan 2-monosygense [Aprobactinum tumificients] - sp3/BAKK hypothetical pretext p15-6AURA_2185 [Aprobactinum tumificients] - sb3/BAKK	951 946 945 945 943	951 946 945 945 943	100% 100% 100% 100% 100%	0.0 0.0 0.0 0.0 0.0	
▼ <u>Sequ</u>	Accession MBAA76146.1 MAA30489.1 MP 398528.1 MP 053423.1 MP 053423.1 MP 25017.1	Description     Vrystophan monocoryperase [Aprobactivit Lainefacture]     trystophan monocoryperase [Aprobactivit Lainefactures at: CSB]     trystophan A-monocoryperase [Aprobactivit Lainefactures at: CSB] > p3(AAS)     trystophan A-monocoryperase [Aprobactivit Lainefactures at: CSB] > b3(AAS)     trystophan A-monocoryperase [Aprobactivit Lainefactures at: CSB] > b3(AAS)     trystophan A-monocoryperase [Aprobactivit Lainefactures at: CSB] > b3(AAS)     trystophan A-monocoryperase [Aprobactivit Lainefactures at: CSB]     trystophan A-monocoryperase [Aprobactivit Lainefactures at: CSB]     trystophan A-monocoryperase [Aprobactivit Lainefactures]	951 946 945 945 943 895	951 946 945 945 943 895	100% 100% 100% 100% 100%	0.0 0.0 0.0 0.0 0.0 0.0	6
▼ <u>Sequi</u>	Accession MAA76146.1 MA276146.1 MNP.395528.1 MNP.395528.1 MNP.053423.1 MNP.053423.1 MP25012.1 MP25012.1	Description tryptophan monocorypanse (Aprobactinum turindfacters) tryptophan monocorypanse (Aprobactinum turindfacters) tryptophan 2-monocypanse (Aprobactinum turindfacters str. C38) - sp3/ARM systophan 2-monocypanse (Aprobactinum turindfacters) - sb3)-sb3/ARM phytophan 2-monocypanse (Aprobactinum turindfacters) - sb3)-sb3/ARM Rectlaner (All-Thytophan 2-monocypanse s-psi(TAAVVI tryptophan 2-mo turindfacters) - sb3/BAM Rectlaner (All-Thytophan 2-monocypanse s-psi(TAAVVI tryptophan 2-mo	951 946 945 945 943 895 881	951 946 945 945 943 895 881	100% 100% 100% 100% 100% 100%	0.0 0.0 0.0 0.0 0.0 0.0 0.0	G
▼ <u>Sequi</u>	Accession DAA20348.1 DA203482.1 DA203482.1 DA2044540.1 DA244540.1 DA244540.1 DA25012.1 DA25012.1 DA25012.1 DA25012.1 DA25012.1	Description     Tryptophan monocorygenese [Aptrobactivum Limerfactions]     tryptophan monocorygenese [Aptrobactivum Limerfactions at: CSB]     tryptophan 2-monocorygenese [Aptrobactivum Limerfactions at: CSB]     tryptophan 2-monocorygenese [Aptrobactivum Limerfactions]     tryptophan 2-monocorygenese [Aptrobactivum Visit]     Lam (Aptrobacterium Limerfactions]     tryptophan 2-monocorygenese [Aptrobactivum Visit]	951 946 945 945 945 943 895 881 881	951 946 945 945 943 895 881 881	100% 100% 100% 100% 100% 100% 100%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	6
▼ Sequi	Accession @LAA76346.1 @AA030489.1 @VP_35528.1 @VP_053423.1 @VP_053423.1 @VP_00367411.1 @C2394498.1 @D0A3V3.1	Description tryptophan monocovyprase [Aprobactinum turinflacients] tryptophan anoncovyprase [Aprobactinum turinflacients (z. C38) tryptophan 2-monocyprase [Aprobactinum turinflacients str. C38) spBARSK hypothetical proteint pTi-SARIAB, 2185 (Aprobactinum turinflacients) 2-bB[BAR Actiliante: (All-Thyptophan 2-monocyprase p2)(FIAASCH Typtophan 2-mon lam (Aprobactinum turinflacients) xpBAR25038.11 [am (Aprobactinum turine tryptophan 2-monocyprase [Aprobactinum vita] Redmen: (All-Typtophan 2-monocyprase p2)(AbARSCH Typtophan 2-monocyprase (Aprobactinum turinflacients) xpBAR25038.11 [am (Aprobactinum turine tryptophan 2-monocyprase p2)(AbARSCH Typtophan 2-monocypra	951 946 945 945 943 895 881 881 881 881 875	951 946 945 945 943 895 881 881 881 875	100% 100% 100% 100% 100% 100% 100% 100%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© ©
▼ Sequi	Accession @hAA763146.1 @hAD30489.1 @hAD30489.1 @cAB44640.1 @cAB44640.1 @rc.801967413.1 @rc.801967413.1 @rc.801468.1 @rc	Psychophan monocorygenase [Aprobactivum Limethickens] tryptophan monocorygenase [Aprobactivum Limethickens] tryptophan 2-monocorygenase [Aprobactivum Limethickens str. CSB] > pb]AAK0 tryptophan 2-monocorygenase [Aprobactivum Limethickens] > cb]BAAK0 thypothetical protein p1=5-6KU8A, p185 [Aprobactivum Limethickens] > cb]BAAK0 tabutane (1-k1-hypothan 2-monocorygenase p3/(DAAK07 Hypothan 2-mo	951 946 945 945 943 895 881 881 881 875 875	951 946 945 945 943 895 881 881 875 875	100% 100% 100% 100% 100% 100% 100% 99% 99%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	6
▼ Seque 1000 1000 1000 1000 1000 1000 1000 10	Accession State 2016 Accession	Description tryptophan monoceryptessa (Aprobactinum turinfacters) tryptophan amonoceryptessa (Aprobactinum turinfacters) tryptophan 2-monocyptessa (Aprobactinum turinfacters) tryptophan 2-monocyptessa (Aprobactinum turinfacters) hypothetical protein 97-56/UAR, 2185 (Aprobactinum turinfacters) 2-bij(IAAR Machaner, 11-II-1796/Jaha, 2-monocyptess 2-9/III(IAAR/III) 11-16/14/2006 tryptophan 2-monocyptessa (Aprobactinum turinfacters) tryptophan 2-monocyptessa (Aprobactinum turinfacters) tryptophan 2-monocyptessa (Aprobactinum turint) Recharms; Fall-1796/pana 2-monocyptessa (Aprobactinum turint) hypothetical protein 91.004 (Aprobactinum turint) hypothetica	951 946 945 945 943 895 881 881 881 875 875 875 874	951 946 945 943 895 881 881 881 875 875 875	100% 100% 100% 100% 100% 100% 100% 100%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© ©
V Seque	Accession MaA25346.1 Ma203489.1 Mr2.39538.1 Mr2.95328.1 Mr2.053423.1 Mr2.053423.1 Mr2.053423.1 Mr2.053675.1 Mr2.053675.1 Mr2.053675.1 Mr2.053675.1	Execution  Verystophan monocorypense [Aprobactivit Interfacions]  verystophan monocorypense [Aprobactivit Interfacions]  verystophan Z-monocorypense [Aprobactivit Interfacions at: CSB]  verystophan Z-monocorypense [Aprobactivit Interfacions]  ketchem [NI-Thytophan Z-monocorypense aprox  verystophan Z-monocorypense [Aprobactivit Interfacions]  ketchem [NI-Thytophan Z-monocorypense  verystophan Z-monocorypense  verystophan Z-monocorypense  verystophan Z-monocorypense  verystophan  ver	951 946 945 945 945 943 895 881 881 875 875 875 874 589	951 946 945 943 895 801 881 875 875 875 875 874 689	100% 100% 100% 100% 100% 100% 100% 99% 99% 99%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© ©
V Seque	Ассезон Філалезіна, 1 Філалознал, Філалознал, Філалознал, Філалознал, Філалозна, Філалозна, Філалоза, Філалоза, Філалоза,	Description Inspitophan monoceryptesse (Aprobatinum tuninfactores) Inspitophan monoceryptesse (Aprobatinum tuninfactores (z. C38) Inspitophan Promoceryptesse (Aprobatinum tuninfactores str. C38) Inspitophan Promoceryptesse (Aprobatinum tuninfactores) 2-bil)AM Achamer, Enil-Prospitan Promoceryptesse 2-pil(DAMOST Unspitan Pro- tem (Aprobatinum tuninfactores) 2-bil)AM Achamer, Enil-Prospitan Promoceryptesse 2-pil(DAMOST Unspitan Pro- prophote-Arrowspites) Aprobatinum Vibi) Achamer, Enil-Prophotenia - Promoceryptesse 2-pil(AAAOST 1) Achamer, Enil-Prophotenia - Promoceryptesse 2-pil(AAAOSTSD) 1 (Ins 1 (Pasmic Inspitophan Pro-moceryptesse (Aprobatinum Vib) Achamer, Enil-Prophotania - Promoceryptesse Unspitophan Promoceryptesse Aprobatinum Vibi) Dipolohan Promoceryptesse Aprobatinum Vibi)	951 946 945 945 945 943 895 881 881 875 875 875 874 689 634	951 946 945 943 895 881 881 881 875 875 875 875 874 689 634	100% 100% 100% 100% 100% 100% 100% 99% 99% 99% 99% 99%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© ©
▼ Seque	Accession ชักล.2014/6.1 ชักล.2014/6.1 ชัก.2014/6.1	Executipation Verystophan monocorystemes [Aptivabatimum tumefactions] verystophan monocorystemes [Aptivabatimum tumefactions] verystophan Z-monocorystemes [Aptivabatimum tumefactions str. CSB] > ebj/AKO verystophan Z-monocorystemes [Aptivabatimum tumefactions] > ebj/AKO katterins (1-H1*Prophan Z-monocorystemes spill(MAXOVTPUttophan Z-monocorystemes [Aptivabatimum tumefactions] > ebj/AKO strategistophan Z-monocorystemes [Aptivabatimum tumefactions] > ebj/AKO strategistophan Z-monocorystemes [Aptivabatimum tumefactions] > ebj/AKO strategistophan Z-monocorystemes [Aptivabatimum tumefactions] katchane (1-H1*Prophan Z-monocorystemes aptivabatimum tumefactions] katchane (1-H1*Prophan Z-monocorystemes aptivabatimum tumefactions] verybothan Z-monocorystemes [Aptivabatimum vita] verybothan Z-monocorystemes verybothan Z	951 946 945 945 945 805 881 875 825 874 689 634 625	951 946 945 943 895 881 881 881 875 875 874 689 634 625	100% 100% 100% 100% 100% 100% 100% 9% 9% 9% 9% 9%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© ©
V Seque	Accession 耐みな2026年1 増から2028月1 増から2028月1 増から2052月1 増から2052月1 増から2052月1 増から2052月1 増から2012月1 増から2012月1 増から2012月1 増から2012月1 増から2015月1 増から2015月1	Description Inspirate (Aprobativity Landschurg Landschurg Inspirate Aprobativity Landschurg Landschurg Inspirate Amonosoyense (Aprobativity Landschurg Landschurg Inspiraten Amonosoyense (Aprobativity Landschurg Inspiraten Amonosoyense) (Aprobativity Inspiraten Amonosoyense (Aprobativity Inspiraten Amonosoyense) (Aprobativity Inspiraten Amonosoyense (Aprobativity Inspiraten Amonosoyense) (Aprobativity Inspiraten Amonosoyense) (Aprobativity Inspiraten Amonosoyense (Aprobativity Inspiraten Amonosoyense) (Aprobativity Inspiraten Amonosoyense) (Aprobativity Inspiraten Amonosoyense (Aprobativity Inspiraten Amonosoyense (Aprobativity Inspir	951 946 945 943 895 881 825 825 825 825 825 824 609 634 625 622	951 946 945 945 945 895 811 875 875 875 875 875 874 689 634 625 622	100% 100% 100% 100% 100% 100% 100% 99% 99% 99% 99% 99%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© © ©
▼ Seque 100 100 100 100 100 100 100 100 100 10	Accession 物ム023081,1 物ム023081,1 物子、296528,1 物子、296528,1 物子、29522,1 物子、29522,1 物子、29522,1 物子、29522,1 からいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいいい	Execution     Execution     Supportain monocorypense [Aprobaction tumefactions]     tryptophan monocorypense [Aprobaction tumefactions ar: CS3] > pb]AAG     tryptophan Z-monocorypense [Aprobaction tumefactions ar: CS3] > pb]AAG     tryptophan Z-monocorypense [Aprobaction tumefactions] > db]BAA     Addmin: Fill-Fill-Fillippina Z-monocorypense [Aprobaction tume]     Supportain Z-monocorypense [Aprobaction tume]     Supportain Z-monocorypense Aprobaction tume[     tryptophan Z-monocorypense     Aprobaction tumefactions]     tryptophan Z-monocorypense     tryptophan Z-monocorypense     tryptophan     Theorypense     Theorype	951 946 945 945 945 983 881 881 881 882 882 874 689 644 645 644 645	951 946 945 945 881 881 881 875 875 875 874 689 634 625 622 468	1 00% 1 00% 1 00% 1 00% 1 00% 1 00% 9% 9% 9% 9% 9% 9% 9%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	6 6 6
▼ Seque 80 80 80 80 80 80 80 80 80 80 80 80 80	Accession 耐みな2025年1 耐みな2025月1 物かし2025月1 れたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたし	Description Inspirate Applications Limited Constitution (Styptophan monocorprises (Aprobativitum LimitedConst) (Styptophan Primocorprises (Aprobativitum Vita) (Styptophan Primocorprises (Aprobativitum (Styptophan (St	951 946 945 943 855 801 875 875 875 875 875 875 875 875 874 609 604 605 604 605 602 468 411	951 946 945 945 945 881 875 875 875 875 875 874 689 634 625 622 468 411	100% 100% 100% 100% 100% 100% 100% 99% 99% 99% 99% 99% 99% 99%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	© © ©
▼ Seque 199 199 199 199 199 199 199 199 199 19	Accession 物ムの2348-1 物ムの2348-1 物子、295528.1 かテ、295528.1 かテ、295528.1 かテ、295528.1 かテ、295528.1 かテ、201877411.1 かテ、201877411.1 かテ、201877411.1 かテ、20187741.1 からいたい。 がのたいたいたいたいたいたいたいたいたいたいたいたいたいたいたいたいたいたいたい	Execution Instruction Instruct	951 946 945 945 955 801 805 805 805 805 805 805 805 805 805 805	951 946 945 945 943 895 875 875 875 875 874 689 634 625 622 468 411 410	1 00% 1 00% 1 00% 1 00% 1 00% 1 00% 1 00% 9% 9% 9% 9% 9% 9% 9% 9%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	6 6 6
▼ Security 1997 1997 1997 1997 1997 1997 1997 199	Accession 耐みな2025年1 耐みな2025月1 物かし2025月1 れたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたしたし	Description Inspirate Applications Limited Constitution (Styptophan monocorprises (Aprobativitum LimitedConst) (Styptophan Primocorprises (Aprobativitum Vita) (Styptophan Primocorprises (Aprobativitum (Styptophan (St	951 946 945 943 855 801 875 875 875 875 875 875 875 875 874 609 604 605 604 605 602 468 411	951 946 945 945 945 881 875 875 875 875 875 874 689 634 625 622 468 411	100% 100% 100% 100% 100% 100% 100% 99% 99% 99% 99% 99% 99% 99%	0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0	6 6 6

**Fig. 3.** Sample of the results obtained from the BLAST search. From these results, students can identify TMO from A. tumefaciens as the target protein.

ture of TMO is not available, a 3D model based on the structure of a homologous enzyme can be obtained using threading computer programs. The program's 3D-Position Specific Scoring Matrix (3D-PSSM) or Protein Homology/analogY Recognition Engine (Phyre) can be used to find a structural enzyme homologous to TMO and obtain a 3D model. These programs use structurally equivalent residues derived from sequence alignments from homologous proteins with similar 3D structures along with secondary structure predictions and solvation potentials to identify structural homology between the target protein and structures in the database (Kelley et al., 2000; Kelley and Sternberg, 2009). The programs thread the amino acid sequence of the target

#### From Structure to Mechanism

The last step in this exercise involves searching the literature database. Students must investigate what is known about PsT-MO, LAAO, and other related amino acid oxidases. The structure of the native form of LAAO was obtained by x-ray crystallography with the competitive inhibitor, amino benzoate (AB), bound in the active site [11]. In the active site of LAAO, residues R90 and Y379 are involved in hydrogen bonding interactions with the carboxylate of AB. Although there is no information about the role of the residues in the active site of LAAO, the role of the homologous residues in PsTMO were probed by site directed



Fig. 4. Phyre results showing that TMO contains structural similarity to several flavin-dependent amine oxidases.

protein onto the 3D structure of a homologous protein and the results are sent as a protein data bank (PDB) file via e-mail.

Using the amino acid sequence of TMO as a probe, 3D models based on the structures of LAAO, monoamine oxidase, and lysine specific histone deaminase were obtained with high scores and sequence identity (Figure 4). All of these enzymes are members of the amine oxidase family of enzymes. The selection of the correct model can not only be based on the E-value, but functional similarities must also be investigated. In the case of TMO, the results show that the E-value of members of this family of proteins are similar. The students (directed by the Instructor) must then look further into the function of these enzymes. By looking at the specific chemical reaction for members of the amino oxidase family, they should be able to select LAAO as the homologous enzyme. LAAO also binds and oxidizes L-amino acids just like TMO; therefore, the active sites should be conserved (Figure 5).



**Fig. 5.** Chemical reaction catalyzed by tryptophan 2-monooxygenase (TMO) and L-amino acid oxidase (LAAO).

mutagenesis (Sobrado and Fitzpatrick, 2003a,b). It was shown that these residues are essential for binding the carboxylate of tryptophan, as predicted by the 3D model obtained from the homology to LAAO. In addition, pH studies with mutant variants of PsTMO at R98 (R90 in LAAO) showed that the unprotonated form of the substrate is the preferred form for catalysis (Sobrado and Fitzpatrick, 2002, 2003b). The active site of LAAO is also conserved in a related flavoenzyme, D-amino acid oxidase (DAAO). These enzymes catalyze the same chemical reaction, but have different enantiomeric selectivity. The overall folds of DAAO and LAAO are not conserved, but the active sites are related by a mirror-symmetrical architecture (Pawelek et al., 2000). In the chemical mechanism of DAAO, the amino acid substrate binds in the unprotonated form and the reaction occurs via the direct transfer of a hydride from the  $C\square$  position

to the flavin cofactor. Consistent with the 3D structure of DAAO, this mechanism does not evoke the need for an active site base that would deprotonate the amino group of the amino acid substrate. Recognizing that DAAO, TMO, and LAAO are all members of the same family, catalyze similar chemical reactions, and share similar active sites, the student can propose a chemical mechanism for AtTMO and LAAO based on the reports for DAAO and the information available about the ionization state of the substrate in the R98 mutants in PsTMO.



**Fig. 6.** Results obtained from this exercise. The numbers in parentheses represent the steps from Table 2. The sequence alignment is a modified alignment obtained from the PSI-BLAST results with L-amino acid oxidase. The structure shown in step 3 is the model of TMO obtained from 3D-PSSM and the inset is the superposition of the active site residues of TMO and LAAO. In the final step of the exercise (Step 4), a proposed mechanism for the oxidation of amino acids by TMO is shown. Arg98 and Tyr413 are implicated in binding the unprotonated from of the substrate and stabilizing the transition state for hydride transfer to form the imino acid. This chemical transformation is followed by reaction with molecular oxygen, where FAD gets reduced and the imino acid intermediate is converted to its keto-acid, carbon dioxide, and water.

11

# Conclusion

Students should be required to write a report outlining the results obtained from each step in the exercise. Their final conclusion should include a proposed chemical mechanism based on the information obtained from PSI-BLAST, Pfam, 3D modeling, and the literature search. In the exercise with tryptophan 2-monooxygenase, students provided a report that is summarized in Figure 2. One learning outcome is that the students are exposed to the important resources available on the internet for studying enzyme function. The principles of molecular evolution and the relationship between enzyme structure and function are also applied. A more important lesson is the importance of critically reading the scientific literature, which can provide valuable information about the particular research the students are performing.

Some questions that the students should answer as they perform the exercise are: Does the DNA sequence code for a protein?, Are there homologous proteins with known function?, Can a 3D model of the unknown target protein be built?, What folds are present in this structure?, What family does it belong to?, and Can a chemical mechanism for the reaction be proposed? If the structure and the mechanism of a particular enzyme is known, similar exercises can be performed utilizing DNA sequences from homologous enzymes.

# References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PS-BLAST: A new generation of protein database search programs, Nucleic Acids Res. 25, 3389-402.

- Gerlt, J. A., and Babbitt, P. C. (2000) Can sequence determine function?, Genome Biol 1, reviews0005.1–0005.10.
- Gerlt, J. A., and Babbitt, P. C. (2001) Divergent evolution of enzymatic function: Mechanistically diverse superfamilies and functionally distinct suprafamilies, Annu. Rev. Biochem. 70, 209-46.
- Kelley L.A, Sternberg M.J. (2009) Protein structure prediction on the web: a case study using the Phyre server Nature Protocols 4, 363 371
- Kelley, L. A., Maccallum, R. M., and Sternberg, M. J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM, J. Mol. Biol. 299, 499-520.
- Pawelek, P. D., Cheah, J., Coulombe, R., Macheroux, P., Ghisla, S., and Vrielink, A. (2000) The structure of I-amino acid oxidase reveals the substrate trajectory into an enantiomerically conserved active site, EMBO J. 19, 4204-15.
- Sobrado, P., and Fitzpatrick, P. F. (2002) Analysis of the roles of amino acid residues in the flavoprotein tryptophan 2-monooxygenase modified by 2-oxo-3-pentynoate: Characterization of His338, Cys339, and Cys511 mutant enzymes, Arch. Biochem. Biophys. 402, 24-30.
- Sobrado, P., and Fitzpatrick, P. F. (2003a) Identification of Tyr413 as an active site residue in the flavoprotein tryptophan 2-monooxygenase and analysis of its contribution to catalysis, Biochemistry 42, 13833-8.
- Sobrado, P., and Fitzpatrick, P. F. (2003b) Analysis of the role of the active site residue arg98 in the flavoprotein tryptophan 2-monooxygenase, a member of the l-amino oxidase family, Biochemistry 42, 13826-32.
- Voet, J. G., Bell, E., Boyer, R. Boyle, J., O'leary, M., Zimmerman, J.K., (2003) Recommended curriculum for a program in biochemistry and molecular biology, Biochem. Mol. Biol. Educ. 31, 161-162.
- Zimmerman, J. K. (2003) Implementing the American Society Biochemistry and Molecular Biology recommended curriculum in a biochemistry and molecular biology degree program in a genetics and biochemistry department, Biochem. Mol. Biol. Educ. 31, 375-377.
- Zimmerman, J. K., and Boyer, R. F. (1993) ASBMB educational affairs committee studies undergraduate biochemistry and molecular biology degrees, FASEB J. 7, 733-4.