

# 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

**T**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

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-value<sup>1</sup> 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 (LAO), 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 LAO [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 LAO. 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

The screenshot displays the ExPASy Proteomics Server interface. At the top, there's a search bar and navigation links. The main content area shows the 'Translate Tool - Results of translation' page. It prompts the user to 'Please select one of the following frames:' and lists three frames. The first frame, '5' Frame 1', is selected and shows a long protein sequence. The second frame, '5' Frame 2', and the third frame, '5' Frame 3', also show protein sequences. The sequences are displayed in a monospaced font, typical of code or data output.

**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, includ-

ing the related superfamily, sequence alignments, available 3D-structures, 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-

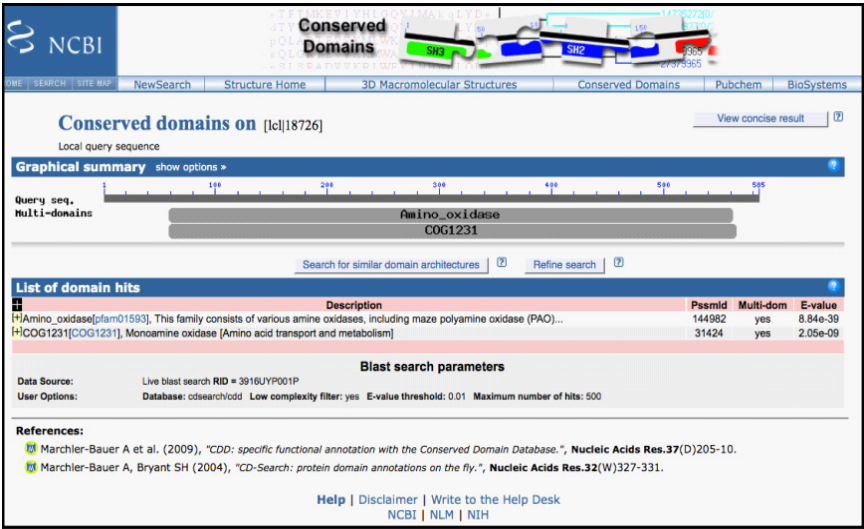


Fig. 2. Results from the BLAST program showing the conserved domain of the protein of interest.

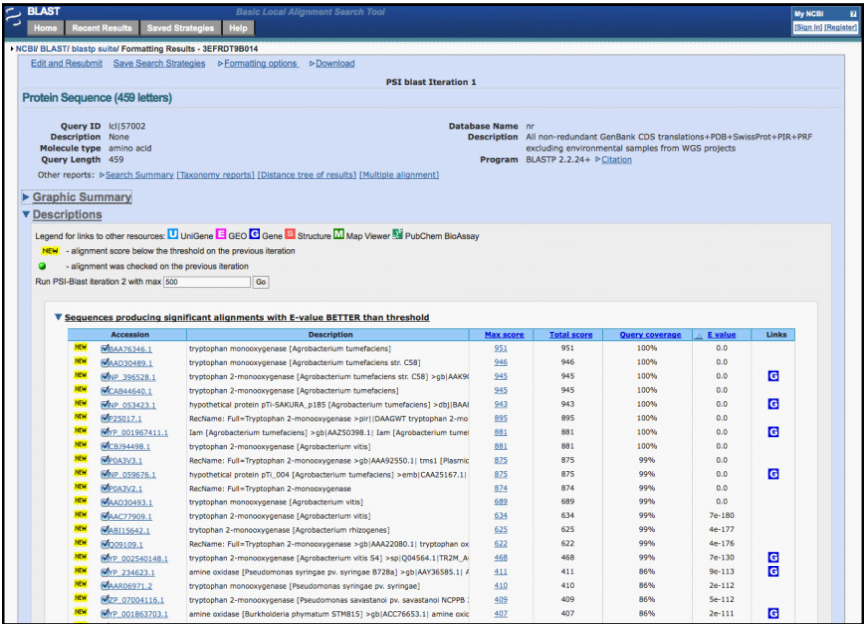


Fig. 3. Sample of the results obtained from the BLAST search. From these results, students can identify TMO from A. tumefaciens as the target protein.



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 PsTMO, 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

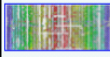
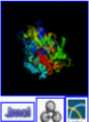

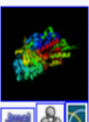

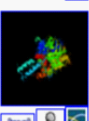



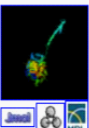
Fold Recognition								
View Alignments	SCOP Code	View Model	E-value	Estimated Precision	BioText	Fold/PDB descriptor	Superfamily	Family
	c2c70B (length:520) 15% i.d.		2.1e-43	100 %	n/a	PDB header:oxidoreductase	Chain: B; PDB Molecule:amine oxidase (flavin-containing) b;	PDBTitle: functional role of the aromatic cage in human monoamine2 oxidase b: structures and catalytic properties of tyr4353 mutant proteins
	c2ldD (length:498) 17% i.d.		4.8e-43	100 %	n/a	PDB header:oxidoreductase	Chain: D; PDB Molecule:l- amino-acid oxidase;	PDBTitle: structure of l-amino acid oxidase from calloselasma2 rhodostoma in complex with l-phenylalanine
	c2ib3B (length:489) 13% i.d.		7.9e-43	100 %	n/a	PDB header:oxidoreductase	Chain: B; PDB Molecule:l-amino acid oxidase;	PDBTitle: the structure of l-amino acid oxidase from rhodococcus2 opacus in complex with o-aminobenzoate
	c2iw5A (length:666) 13% i.d.		8.9e-43	100 %	n/a	PDB header:oxidoreductase/transcription regulator	Chain: A; PDB Molecule:lysine-specific histone demethylase 1;	PDBTitle: structural basis for corest-dependent demethylation of2 nucleosomes by the human lsd1 histone demethylase
	c2v1dA (length:730) 13% i.d.		2.7e-42	100 %	n/a	PDB header:oxidoreductase/repressor	Chain: A; PDB Molecule:lysine-specific histone demethylase 1;	PDBTitle: structural basis of lsd1-corest selectivity in histone h32 recognition

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).

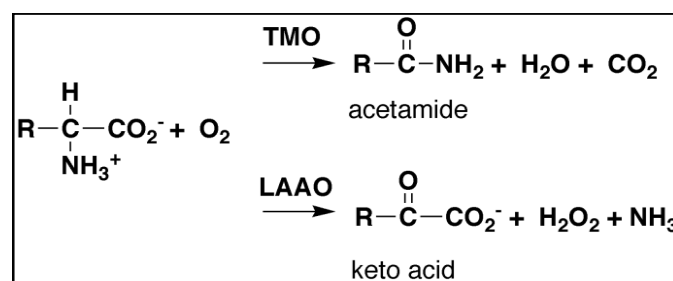
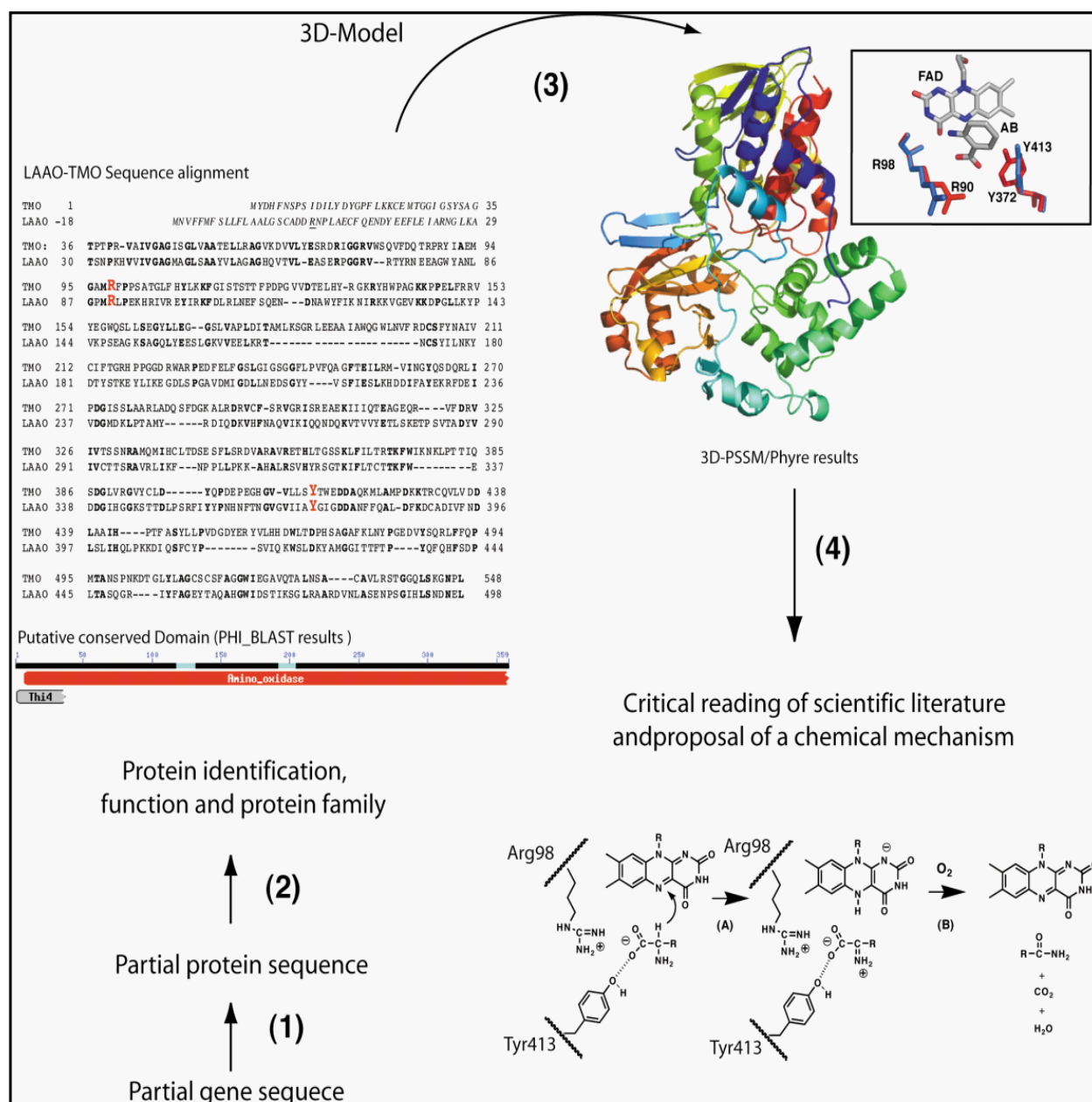


Fig. 5. Chemical reaction catalyzed by tryptophan 2-monoxygenase (TMO) and L-amino acid oxidase (LAAO).

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 $\alpha$  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 form 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.

## 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 L-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.