

US EPA ARCHIVE DOCUMENT

July 7, 2005

MEMORANDUM

SUBJECT: Transmittal of Minutes of the FIFRA Scientific Advisory Panel Meeting Held May 3-4, 2005: Scientific Issues Associated With TSCA Inventory Nomenclature for Enzymes and Proteins

TO: Charles M. Auer, Director
Office of Pollution Prevention and Toxics

FROM: Paul I. Lewis, Ph.D. Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Larry C. Dorsey, Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Clifford J. Gabriel, Ph.D. Director
Office of Science Coordination and Policy

Please find attached the minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, Virginia from May 3-4, 2005. These meeting minutes address a set of scientific issues being considered by the U.S. Environmental Protection Agency regarding TSCA inventory nomenclature for enzymes and proteins.

Attachment

cc:

Susan Hazen
Margaret Schneider
Mary Ellen Weber
Neil Patel
Oscar Hernandez
James Jones
Anne Lindsay
Janet Andersen
Debbie Edwards
Steven Bradbury
William Diamond
Richard Keigwin
Arnold Layne
Tina Levine
Lois Rossi
Frank Sanders
William Jordan
Douglas Parsons
Karen Chu
Anthony Britten
Enesta Jones
Vanessa Vu (SAB)
Greg Fritz
Mark Segal
Henry Lau

FIFRA SAP Members

Steven G. Heeringa, Ph.D.
Janice Elaine Chambers, Ph.D., D.A.B.T.
Stuart Handwerger, M.D.
Gary Isom, Ph.D.

FQPA Science Review Board Members

Ralph Bradshaw, Ph.D.
Richard Cammack, Ph.D., Sc.D.
Louis B. Hersh, Ph.D.
Micah Krichevsky, Ph.D.
William D. Nes, Ph.D.
Robert S. Phillips, Ph.D.
Joseph Spradlin, Ph.D.

SAP Report No. 2005-03

MEETING MINUTES

**May 3-4, 2005 FIFRA Scientific Advisory Panel Meeting,
held at the Holiday Inn-Rosslyn at Key Bridge**

***A Set of Scientific Issues Being Considered by the
Environmental Protection Agency Regarding:***

**TSCA Inventory Nomenclature for
Enzymes and Proteins**

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). This report has not been reviewed for approval by the United States Environmental Protection Agency (Agency) and, hence, the contents of this report do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP was established under the provisions of FIFRA, as amended by the Food Quality Protection Act (FQPA) of 1996, to provide advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP) and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act Science Review Board members serve the FIFRA SAP on an ad hoc basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at <http://www.epa.gov/scipoly/sap/> or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Paul Lewis, Designated Federal Official, via e-mail at lewis.paul@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by the Agency presenters, as well as information presented by public commenters. This document addresses the information provided and presented within the structure of the charge by the Agency.

TABLE OF CONTENTS

	Page
Participants.....	7
Public Commenters	8
Introduction.....	8
Summary of Panel Discussion and Recommendations	8
Panel Deliberations and Response to the Charge	11
References.....	33

SAP Report No. 2005-03

**MEETING MINUTES:
FIFRA Scientific Advisory Panel Meeting,
May 3-4, 2005, held at the Holiday Inn-Rosslyn at
Key Bridge, Arlington, Virginia**

***A Set of Scientific Issues Being Considered by the
Environmental Protection Agency Regarding:***

**TSCA Inventory Nomenclature For
Enzymes And Proteins**

Paul I. Lewis, Ph.D.
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: July 7, 2005

Steven G. Heeringa, Ph.D.
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel
Date: July 7, 2005

**Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
May 3-4, 2005**

TSCA Inventory Nomenclature for Enzymes and Proteins

PARTICIPANTS

FIFRA SAP Session Chair

Steven G. Heeringa, Ph.D., Research Scientist and Director for Statistical Design, Institute for Social Research, University of Michigan, Ann Arbor, MI

Designated Federal Official

Paul I. Lewis, Ph.D., FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, U.S. Environmental Protection Agency, Washington, D.C.

FIFRA SAP Members

Janice Elaine Chambers, Ph.D., D.A.B.T., William L. Giles Distinguished Professor and Director, Center for Environmental Health Sciences, Mississippi State University, Mississippi State, MS

Stuart Handwerger, M.D., Director, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati OH

Gary Isom, Ph.D., Professor of Toxicology, School of Pharmacy & Pharmacal Sciences Purdue University, West Lafayette, IN

FQPA Science Review Board Members

Ralph A. Bradshaw, Ph.D., Professor, Department of Physiology and Biophysics, School of Medicine, University of California at Irvine, Irvine, CA

Richard Cammack, Ph.D., Sc.D., Professor of Plant Biochemistry, Department of Life Sciences King's College London, University of London, London, United Kingdom

Louis B. Hersh, Ph.D., George Schwert Professor and Chair, Department of Molecular and Cellular Biochemistry, College of Medicine, University of Kentucky, Lexington, KY

Micah Krichevsky, Ph.D., Chairman, Bionomics International, Wheaton, MD

William D. Nes, Ph.D., Program Director, Molecular and Cellular Biosciences Division, National Science Foundation, Arlington, VA

Robert S. Phillips, Ph.D., Professor of Chemistry, Biochemistry & Molecular Biology,
Page 7 of 34

University of Georgia, Athens, GA

Joseph Spradlin, Ph.D., Adjunct Professor, University of Arkansas, Hot Springs, AR

PUBLIC COMMENTERS

Oral statements were made by:

Ms. Alice Caddow of Genencor International and John Carroll, Ph.D., of Novozymes on behalf of the Enzyme Technical Association

Dan Robertson, Ph.D., on behalf of Diversa Corporation

Mr. Brent Erickson on behalf of the Biotechnology Industry Organization

Ms. Martha Marrapese on behalf of Keller and Heckman, LLP

Written statements were provided by:

Biotechnology Industry Association

Diversa Corporation

Enzyme Technical Association

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of a set of scientific issues being considered by the Agency pertaining to its review of the Toxics Substance Control Act (TSCA) inventory nomenclature for enzymes and proteins. Advance notice of the meeting was published in the *Federal Register* on March 9, 2005 (docket number OPP-2005-0060). The review was conducted in an open Panel meeting held in Arlington, Virginia, May 3-4, 2005. The meeting was chaired by Steven G. Heeringa, Ph.D. Paul Lewis, Ph.D., served as the Designated Federal Official. Clifford Gabriel, Ph.D., Director, Office of Science Coordination and Policy, EPA, welcomed the Panel to the meeting. Mr. Neil Patel, Associate Director, Economics, Exposure and Technology Division, Office of Pollution Prevention and Toxics, EPA, provided introductory remarks. Greg Fritz, Ph.D., Chemist, Economics, Exposure, and Technology Division, Office of Pollution Prevention and Toxics, EPA, provided a background on the TSCA inventory system. Mark Segal, Ph.D., Microbiologist, Risk Assessment Division, Office of Pollution Prevention and Toxics, EPA, reviewed the Agency's proposed approach for enzyme identification on the TSCA inventory.

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

The FIFRA SAP considered each of the four data elements (function, sequence, source, and processing) the Agency is proposing to employ for a comprehensive listing and for distinguishing among enzymes on the TSCA Inventory. The Panel concluded that each of these four elements has merit for cataloging and distinguishing among enzymes, and that the Agency's proposed nomenclature system is useful and should be retained. However, each of the four elements do not necessarily carry equal weight, and some elements may not be known and likely

are difficult to obtain. The Panel believed that many, if not the majority of its responses were most relevant when applied to purified or isolated enzyme preparations, which as noted above is often not the case with commercial enzymes. The application of the four elements to impure enzyme preparations could be useful, but in many cases the information available is incomplete.

In considering the four elements, the Panel concluded that the identification of enzyme function is of prime importance. However, function, by itself may be too broad a criteria. Sequences can differ, yet activity remains the same. Source, although less useful, should not be ignored. Processing information can be of importance in certain cases.

The Panel, in preparing and reporting its deliberations, noted that commercial enzyme preparations generally lack homogeneity and indeed in most cases, there is no attempt by the manufacturer to purify the enzymatic activity. Since classifications using descriptors based on sequence, source or processing largely presuppose some degree of purity, their indiscriminate inclusion in any classification scheme would not serve either the agency or the commercial purveyors well. However, the Panel did concur that improvements in identification and classification were needed and reported that in many instances information from one or more of the proposed categories, when added to a functional description, could materially aid in developing and maintaining an improved classification scheme without overly burdening either industry or the government for collecting and categorizing such information.

Use of the functionally based International Union of Biochemistry and Molecular Biology (IUBMB) nomenclature system is recommended. The additional knowledge of reaction conditions and non-catalytic functions of specific enzymes was believed to be of limited use. On the other hand, knowledge of the binding specificity and catalytic mechanism were thought to be useful. Although some enzymes are multifunctional, these are relatively few in number. In the absence of post-translational modifications, knowledge of the amino acid sequence of an enzyme provides an exact chemical description of the molecule. However, large variations in the sequences of enzymes catalyzing the identical reaction makes sequence not as useful a criterion for describing function.

The Agency addressed the issue of the expected amount of variation in an enzyme amino acid sequence from various sources. When an enzyme is cloned from its gene, random errors can be introduced; however these are readily detected by DNA sequencing. Such variants could be altered in their stability, specificity, or catalytic efficiency. During production of an enzyme, variation in amino acid sequence can occur as a result of transcriptional as well as translational errors. However, these occur at a very low frequency and the Panel concluded they would not represent a significant fraction of the enzyme preparation. The Panel knew of no case in which an enzyme used in commerce or research changed its amino acid sequence over time.

The Panel addressed the Agency's question as to the level of maximum permissible overall amino acid sequence variations that could be determined when identifying a specific enzyme. The Panel concluded that one cannot predict the differences that might occur among enzyme variants. In some cases a single amino acid change can drastically alter the enzyme

while in other cases large sequence variants catalyze the same reaction. Changes at the active site or at specific functional motifs have a much greater effect than changes at other regions. The same can be said for deletions and/or excisions; the site at which they occur determines to what extent they will change enzyme activity. It would be difficult for an enzyme manufacturer to determine the location of the active site or other specific regions, as this requires techniques generally not used by enzyme manufacturers.

The Agency inquired as to the efficacy of existing sequencing technologies. The Panel pointed out that both protein and nucleotide sequencing are reliable ways of measuring amino acid sequence, but due to rapidity and cost most protein sequences are determined from their nucleotide sequence. In conjunction with amino acid sequence, knowledge of post-translational modifications is important as these can affect both structure and function.

Regarding knowledge of the source of an enzyme, the Panel concluded that knowing the original source as well as the production source will be of limited value in differentiating enzymes. If the original source was used as an identification element to discriminate among enzymes, the lowest taxonomic level available should be utilized. However, it was noted that scientific nomenclature codes only extend down to the subspecies level. Knowledge of the tissue or organ source will be of value in a relatively limited number of cases. Similarly, for enzymes only characterized by activity, the chemical, geographic, and/or environmental condition from which source organisms were isolated may be a useful descriptor. This is not the case when manipulating the enzyme's original source prior to gene transfer or manipulating the production source.

In reference to processing, the Panel concluded that processing techniques generally will not affect the structure of an enzyme. However there are documented cases when this is not the case. The use of detergents, and autolysis by endogenous or exogenously added proteases, could change the structure of a protein and its activity. For example, the milk activity referred to as xanthine oxidase, is actually xanthine dehydrogenase, and becomes an oxidase after treatment with thiol reagents or proteases. This is an example where a simple processing step gives you an enzyme with different protein structure and catalytic function. The Panel noted that although there are a host of processing techniques utilized in the research laboratory, most of these are not utilized for the majority of enzymes on the TCSA inventory. Changes in processing techniques are not likely to occur with industrial enzymes, and any such changes would not be expected to significantly alter an enzyme's chemical structure or its properties.

Other factors that might be beneficial to describe an enzyme include the minimum structural requirement for a substrate, the mechanism of catalysis, the presence or absence of regulatory sites, the assay used to define an enzyme's activity, and the identification of enzymes with multiple non-identical subunits.

PANEL DELIBERATIONS AND RESPONSE TO THE CHARGE

The specific issues to be addressed by the Panel are keyed to the Agency's background documents, references and Agency's charge questions.

EPA is proposing the use of four data elements (function, sequence, source, and processing) for comprehensively listing and distinguishing among enzymes on the TSCA Inventory. The following questions are intended to help the Agency make a final decision on how enzymes will be listed on the Inventory in the future.

General Comment

The Panel addressed the class of proteins with a catalytic function, namely enzymes, rather than proteins in general. However, many of the answers can be applied to proteins as a group.

Function

The FUNCTION of an enzyme refers to its catalytic activity. Internationally-accepted nomenclature conventions of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) describe and categorize enzymes based on their function. The NC-IUBMB assigns enzymes an Enzyme Committee (EC) code number based on the specific reaction(s) catalyzed by the enzyme, the nature of the bond involved, and the substrate acted upon. EPA intends to incorporate function into TSCA Inventory enzyme listings by using these EC codes and the systematic name for the specific catalytic activity. In the questions below, please identify the scientific merit for using function information to differentiate among enzymes and identify what level of detail regarding function would be scientifically appropriate for this purpose.

1. While the Agency recognizes the practical, historical advantages of using function to describe enzymes, in the context of the Agency's need for unique and unambiguous naming, what is the *scientific* rationale for identifying an enzyme based on the chemical reaction(s) it catalyzes?

The Panel qualified its response indicating that this question can best be addressed with regard to isolated enzymes (not necessarily pure). The Panel concluded that in terms of function, the Agency's proposed nomenclature system is useful and should be retained. Identification of an enzyme based on the chemical reaction(s) it catalyzes permits a useful description of an enzyme. The chemical reaction catalyzed by an enzyme is the most essential piece of information needed in any nomenclature system for enzymes. The function of an enzyme is more important, from a practical point of view, in naming an enzyme than any of its structural features. The Enzyme Commission (EC) list, maintained by the nomenclature committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB; IUBMB and IUPAC/IUBMB Joint Commission on Biochemical Nomenclature, <http://www.chem.qmw.ac.uk/iubmb/enzyme>) defines enzymes based on the reaction catalysed. Such a system represents a clear descriptor of the enzyme. In many cases, the IUBMB nomenclature system, which is based on function (reaction catalyzed), further classifies enzymes

based on mechanism (e.g. zinc metalloproteases) and provides insights into functional groups involved in catalysis. Grouping of enzymes based on their catalytic mechanism frequently permits insights into structural features common to the group. It also provides a starting point for defining physiological function. If one wanted to use an enzyme for some specific purpose, reference to the grouping based on the reaction it catalyzes would be an important starting point. The nomenclature used by the IUBMB provides a reference point where different enzymes catalyzing the same reaction (e.g. proteases) might be used for the same purposes, or provides a way to recognize differences in the reactions catalyzed by similar enzymes, i.e. proteases with different specificities.

While the Panel did recognize limitations with the IUBMB nomenclature system, the Panel was generally in support of the IUBMB approach. The IUBMB-EC system is usually based on measurements of enzyme functions in an aqueous system reacting with a particular substrate. In some cases, this may not be relevant to industrial uses. There are other complexities that make the IUBMB nomenclature of categorizing enzymes based on the reaction they catalyze awkward. For reasons of consistency, the system describes the reaction catalysed by a formalism that does not necessarily reflect the thermodynamically favored direction of the reaction nor the direction of the reaction in metabolism. Categorizing an enzyme based on the chemical reaction(s) it catalyzes doesn't define specificity in a precise manner since enzymes may react with multiple substrates. Furthermore, the enzyme name does not always reflect its commercial use (e.g. xylose isomerase is used for catalyzing fructose formation).

For historical reasons, the depth of coverage of enzymes with broad and narrow specificities is not consistent throughout the nomenclature list. The peptidases (proteases) are listed in great detail, including in some cases structure and species, whereas other enzymes are listed only as broad classes. For example, the restriction endonucleases, are not described in the IUBMB system of naming enzymes by the reaction they catalyze, but rather reference is made to lists in other databases. Another potential problem is that some activities have two or more EC numbers depending on whether the enzymes have broad or narrow specificity. It is estimated that about 20% of the enzyme names in the literature are incorrect.

The IUBMB number tells us basic functionality. More information such as specificity would enhance the value of the classification. Overall, there is more of an interest in function than structure. However, without the IUBMB number, or other description of the reaction catalyzed, no useful information is available. Bioinformatics databases treat the IUBMB number as an identifier and by compiling all the data, get an optimum description or definition of an enzyme.

The one property that all enzymes have in common other than being proteins is that they catalyze at least one chemical reaction. Clearly no list of enzymes should omit this parameter. However, it needs to be recognized that there can be ambiguity involved, particularly with enzymes whose activity is very general such as a non-specific protease. It is also clear that none of the databases or classification systems is complete, although all have some value.

2. How precise is the IUBMB EC categorizing system for describing enzyme function? For

example, in addition to the EC function category to which an enzyme belongs, what additional information about enzyme structure and/or chemical properties, if any, would be gained by a more detailed functional description that included

a. enzyme reaction conditions (e.g., pH range, reaction temperature range)?

The Panel concluded that at this time little would be gained by including enzyme reaction conditions. The properties listed, including pH range, etc. are standard properties that are generated during a typical characterization of the protein in order to establish its activity assay. These properties can vary for a given enzyme or amongst enzyme systems, depending on its source. The Panel concluded that knowledge of additional information about enzyme structure and/or chemical properties in order to provide a more detailed functional description would be of limited use. The difficulty of comparing enzyme preparations at different times and in different places has been recognized. Hicks and Kettner (2003) described efforts to establish internationally agreed Experimental Standard Conditions of Enzyme Characterizations for the measurement (assay) of enzyme activity in an initiative led by the Beilstein Institute.

b. non-catalytic enzyme functions that are not represented by EC codes (e.g., binding properties)?

The issue of non-catalytic enzyme functions does not appear to be important to the inventory issue. Additional information such as allosteric modulators (effectors) can be a useful indicator of protein function. Interacting proteins can also be an important indicator of effector action.

Information about the binding properties of the active site is important in predicting what other substrates an enzyme will use besides the originally described substrate. The binding properties are also needed to predict the “reversion” products formed when the catalyzed reaction is at or near thermodynamic equilibrium. In hydrolyses, the size distribution of polymer fragments depends heavily on the structure of the binding site.

c. other additional information about function that could be used to differentiate enzymes (please specify what would be of value)?

Other useful information about an enzyme's function would be its catalytic mechanism. Enzymes that use the same substrate and produce the same products may do so by totally different catalytic mechanisms (e.g. chymotrypsin and pepsin).

3. The Agency is trying to gauge the probable comprehensiveness of enzyme catalytic function descriptions for subsequent enzyme reporting.

a. How common are multifunctional enzymes?

Multifunctional enzymes represent a relatively small percentage of the total known enzymes. Less than 1% of total enzymes are known to be multifunctional (including orthologous

enzymes in different organisms). A key word search for “multifunctional enzyme” in the National Center for Biotechnology Information (NCBI) protein data base resulted in 1954 hits, in Swiss-Prot 1666 hits of 180652 entries (0.9%) and in TrEMBL 4011 hits out of 1689375 entries (0.2%).

Some multifunctional enzymes are multienzyme complexes with separate chains for each activity. Other multifunctional enzymes are single chains with multiple catalytic sites. In some cases, a multienzyme complex in one organism may be a single chain in another organism. Some enzymes are promiscuous, and can catalyze multiple activities in a single site. It is important to keep in mind the distinction between enzymes with broad specificity (e.g. cytochromes P450) and truly multifunctional enzymes (such as ribulose biphosphate carboxylase/oxygenase). The simplest way to make that distinction is by catalytic activity. Enzymes may have unrelated structural or regulatory functions, in addition to catalytic activity. These can include protein machines (e.g., proteasomes) that contain both catalytic and non-catalytic components. It may also be useful to distinguish multisubstrate enzymes and truly multifunctional enzymes.

b. How frequently are new catalytic functions for existing enzymes discovered?

The discovery of new catalytic functions for enzymes is an infrequent event. Based on the low abundance of multifunctional enzymes (less than 1% of total), the reporting of a new catalytic function for an enzyme would be expected to be an infrequent event. However a search of PubMed for 2005 publications shows three new activities described for previously identified proteins.

In the past, the discovery of multifunctional enzymes has often been related to the elucidation of metabolic pathways. Enzymes that catalyze sequential steps in metabolism may be fused into a single chain. If the activities are not related, then the discovery of multifunctional enzymes is usually due to accident. Researchers investigating different activities may find that after purification and sequence determination, both enzyme activities are due to the same protein.

c. How good are existing models to assess the likelihood that an enzyme may have several catalytic functions?

There are no general models to discover new activities in known enzymes. Sequence alignment of a known enzyme with genome databases may show homology to another enzyme sequence, which may help to indicate another activity. If a new sequence shows homology to two different enzyme sequences, that may provide an indication that there are two activities in a single enzyme. However, homology is not a reliable predictor of a particular activity.

d. What information is required to utilize such models?

There is no particular piece of information that can be used to assess the likelihood that an enzyme may have several catalytic functions. Proteomics has opened the field for finding interacting proteins, and thus it may be possible to make predictions of activity based on protein-protein interactions. Promiscuous enzymes and those with additional regulatory or structural functions are particularly difficult cases for the prediction of novel activities or functions from sequence data. In an impure mixture, such as those used in an industrial setting, it will be impossible to determine if two different enzyme activities are properties of a multifunctional enzyme or two different enzymes.

Sequence

The AMINO ACID SEQUENCE of an enzyme is known as its primary structure. It is a systematic representation of the linear sequence of amino acids that are connected via amide bonds to form a polypeptide. In the questions below, please consider what scientific support there is for using sequence information to differentiate among enzymes and what level of detail would be scientifically appropriate for this purpose.

4. What information about an enzyme could be gained by identifying it based on its amino acid sequence?

The amino acid sequence of an enzyme defines its covalent structure (minus any co/post translational modifications and prosthetic groups). This is analogous to the structure of any other organic molecule and can be determined very accurately. Notwithstanding genetic variation, this can be viewed as the 'fingerprint' of this molecule. Clearly, an enzyme could be uniquely identified by its sequence. There is a good reason that most biotechnology patents on matter are written on sequences of either the protein or the DNA encoding it. Sequence information also often provides structural and functional information of possible interest in the commercial application of enzymes. Since the enzyme sequence, if properly folded, will virtually always lead to the structure that produces its function(s), in this case catalysis, it is a more absolute descriptor than any other information available.

It should be noted however that there are sequence variations of the same enzymes with the same catalytic function within an organism and among organisms. Thus one might accumulate a long list of sequences for the same functional enzyme that might vary in a few amino acids or might vary by a significant number of amino acids. As a means to address this, deciding when enzyme structural relationships can be ascertained from sequence identity and sequence similarity calculations is helpful. These do not necessarily define function, although they may define a common reaction mechanism used by a group of related enzymes performing different functions. In this regard, phylogenetic analyses can be a useful adjunct in using this information as a descriptor.

Thus although sequence information will clearly define an enzyme chemically, the large variations expected among enzymes catalyzing the identical reaction, and thus having the same function, makes this potentially a daunting task. This is probably not such a problem with genetically engineered enzymes, as the changes made will have been defined. On the other hand

any listings of enzymes based entirely on exact sequence would be large and cumbersome if one considers both natural and engineered mutations.

5. The Agency is trying to assess the expected amount of variation in an enzyme amino acid sequence due to various causes in spite of current quality control standards.

When an enzyme is cloned from one source for expression in the same or a heterologous source (i.e. a gene from one organism expressed in another), the preparation of the gene can introduce random errors based on the fidelity of the system used to copy it. This can lead to random point mutations in the enzyme, which can vary widely in their resultant effect. On the one hand these mutations may have no effect or they can lead to an enzyme with reduced or no activity, or an enzyme with a lower stability. However, in the absence of DNA sequencing, the errors can go undetected.

When an enzyme is produced in batches it is usually generated from its cDNA with an appropriate expression vector. In this case, there are then two general sources of errors that can lead to a variation in an enzyme amino acid sequence; nucleotide changes (mutations) in the gene that arise from transcription errors and misincorporation of amino acids into the enzyme through translational errors.

a. How much and what type of variation (including substitutions, deletions, and additions) can be expected in the amino acid sequence of an enzyme *produced in multiple batches that will arise due to unintended differences in production conditions*? Estimate a percentage, number of residues, or other quantifiable measure of variation.

As noted above, variations in the amino acid sequence of an enzyme produced in multiple batches can arise from mutations introduced during transcription and through translation errors. By far the most common change will be amino acid substitutions. Deletions and additions would more likely lead to unstable proteins. The more replications of a cloned gene, the greater the likelihood of a mutation. Since the mutation rates are low, if the enzyme were produced in multiple batches, one would expect a relatively low number of amino acid sequence changes to arise, and these would be different in each batch of enzyme. It is difficult to quantitate the number of expected changes since they are dependent upon a number of variables including the enzyme itself, the expression system used, and the growth conditions. When multiple batches are used, the stable variations will continue to accumulate from batch to batch. Manufacturers will avoid long-term deterioration by re-starting the fermentation from stock cultures.

b. How much and what type of variation (including substitutions, deletions, and additions) can be expected in the amino acid sequence of an enzyme *within a given sample of a single production batch due to individual-level variation in an enzyme-producing population*? Estimate a percentage, number of residues, or other quantifiable measure of variation.

The same type of variation observed in multiple batches (including substitutions, deletions, and additions) will be expected in the amino acid sequence of an enzyme within a single production batch. However since these will not be cumulative they will be present at a much lower frequency relative to an enzyme produced in multiple batches.

The error rate of transcription is generally low but significant (about 1 in 10^7). The more replications of a cloned gene, the greater the likelihood that it will become mutated.

c. How much and what type of variation (including substitutions, deletions, and additions) can be expected in the amino acid sequence of an enzyme *across multiple samples collected over time* (e.g., in microbial cultures stored for extended periods) *due to changes in an enzyme-producing population*? Estimate a percentage, number of residues, or other quantifiable measure of variation.

i. Over what time scale will such variation arise? That is, is there a predictable relationship between the amount of variation and the length of time in culture?

Differences in protein sequence can arise from mutations in the gene or from translational errors. Mutations in the gene causing differences in protein sequence arise at the mutation frequency of the system used to express it. For *E. coli* there are between 10^{-6} and 10^{-7} mutations per gene per generation and assuming any one of three mutations could affect a given amino acid, this would turn out to be a maximal value of 3×10^{-6} to 3×10^{-7} amino acid changes per generation. However, due to the fact that many base changes do not affect amino acid coding, this number will be significantly smaller. If any of these random mutations caused instability, the protein would likely be cleared from the cell by proteolysis. Thus, if the culture were not under continuous culture for many generations, one would expect a negligible number of amino acid sequence changes to arise from mutations, and these would be different in each batch of enzyme.

Translational errors occur at approximately 10 times the frequency as transcriptional changes (Rosenberger 1994a). During recombinant protein synthesis, translational errors will occur at specific sites in the protein with each translational error occurring at an error rate of 2×10^{-3} to 2×10^{-4} (Rosenberger 1994a) and thus representing a small but significant percentage of the total. Rosenberger (Scorer et al. 1991; Rosenberger 1994b) measured the error rate in synthetic recombinant mouse epidermal growth factor (EGF) produced in *E. coli* by measuring the phenylalanine content in this non-phenylalanine containing protein. It was found that phenylalanine was mis-incorporated into 1.1% of all the amino acids present and 2.6% of those amino acids whose codons differed by a single base.

However, the translational error rate can be protein dependent. Weickert and Apostol (1998) measured isoleucine incorporation into coexpressed di-alpha-globin and beta-globin expressed in *E. coli*. They found ≤ 0.2 mol of isoleucine per mol of hemoglobin which corresponds to a translation error rate of ≤ 0.001 . They concluded that this is not different from typical translation error rates found for other *E. coli* proteins. Two different expression systems that resulted in accumulation of globin proteins to levels equivalent to $\sim 20\%$ of the level of *E. coli* soluble proteins also resulted in equivalent translational fidelity.

In yet another study, Kane et al. (1992) found that about 2% of recombinant bovine

placental lactogen (bPL) produced from *E. coli* using rare arginine codons exhibited an altered trypsin digestion pattern. This was not the case when a preferred codon was used. They proposed a model in which translational pausing occurred at the arginine residues encoded by an AGG codon because the corresponding arginyl-tRNA species is reduced by the high level of bPL synthesis, and a translational hop occurs from the leucine residue 85 TTG codon to the leucine residue 87 TTG codon. Thus misincorporation of amino acids may depend on the protein itself as well as codon usage and the system utilized.

Misincorporation of amino acids occurs at an error rate ~10 times that of stop-codon read-through errors and frameshift errors, the latter being estimated to occur at about 1/2 the frequency as stop-codon read-through errors (Rosenberger 994a). The above numbers represent single batch data, and thus in multiple batches the number of errors would increase.

The use of multiple samples collected over time will reduce the accumulation of mutations in the culture. Due to the variables noted above there is not an easily quantifiable relationship between the amount of variation and the length of time in culture.

ii. What kinds of changes might occur to an enzyme preparation if naturally occurring variants become the dominant component (e.g., changes in rates of activity, reactions catalyzed, substrate range, response to environmental conditions)?

If naturally occurring variants became the dominant component, any one of a number of changes could occur that could include changes in the rate in which the enzyme catalyzes the reaction (usually slower), subtle changes in the enzyme's substrate specificity, possible changes in the enzyme's response to environmental conditions, and changes in stability. However it would be very unlikely that the reaction mechanism would change.

The production of commercial industrial grade enzymes is usually if not always started from a frozen stock culture and rarely or never from a sample from a previous batch. This is done to ensure that the enzymes produced from batch to batch are the same and that any variations that may occur during a production run are never passed on to future production runs. Very large numbers of frozen cultures are produced from one highly controlled fermentation and then used in production over several years. This minimizes the number of replications and their cumulative mutations. Therefore, the amino acid sequence can remain unchanged for decades. Finally, manufacturers will avoid long-term deterioration by re-starting the fermentation from stock cultures.

iii. Have any enzymes in commerce or research been known to change in amino acid sequence over time? Have any been known to remain unchanged in amino acid sequence for a year/decade or longer?

In the drug field, it is well established that biologics do not change significantly over periods of at least ten years. The Panel could not think of any specific examples of enzymes in commerce or research that are known to have changed in amino acid sequence over time. However, the Panel pointed out that the whole process of evolution leads to changes in amino acid sequence over time.

6. EPA is trying to judge whether a scientifically appropriate level of maximum permissible *overall* amino acid sequence variation could be determined when identifying a specific enzyme.

a. What types of differences may exist among enzyme variants that differ by a single amino acid change? What types of differences may exist among enzyme variants that differ in amino acid composition by 0.5%? 1%? 10%? etc.?

The Panel could not specify the differences among enzyme variants that vary by a single amino acid change, because the consequences of such changes vary enormously. In some cases, a single amino acid change, e.g. in the active site can virtually eliminate activity. On the other hand, two proteins having only 20% homology can have a closely similar protein folding and activity. Usually a single or few amino acid differences have only a minor effect on activity or specificity; the principal effect is on enzyme stability.

b. How much does the region of the enzyme in which the variation occurs matter? For example, how important are changes in the amino acid sequence of the active site versus the rest of the molecule? Are there other regions of the enzyme that are considered important, i.e., where sequence is generally conserved?

The region of an enzyme in which variation occurs is very important. Changes at surface residues generally will have little effect while changes in the active site will have dramatic effects. It is noted in the literature (Carter and Wells, 1988) that mutation of the active-site serine of subtilisin, for example, causes activity to decrease by a factor of 10^6 . The activity is still greater however than the uncatalysed reaction. Amino acid changes in the hydrophobic core of a protein are more likely to affect stability than those on the surface of the molecule.

In addition to the active site, there are other regions of the enzyme that are considered important and where amino acid changes will likely have a dramatic effect. Certain arrangements of amino acids, known as motifs, can be recognized in the amino-acid sequences of enzymes, which in their folded state, have specific properties such as the HXXXH motif as a zinc binding motif. There are many such motifs for binding specific substrates and for binding cofactors or effectors.

c. How important are deletions and/or excisions in determining differences between enzymes?

As noted by the Panel in response to question 6a above, the site of deletions and/or excisions is critical. On the one hand, a small deletion may have a dramatic effect on activity, while a long deletion may leave some activity intact, with decreased stability.

d. How easy would it be for a typical enzyme manufacturer to determine the location of the active site or other specific regions mentioned in 6b?

There are relatively few simple and inexpensive ways for an enzyme manufacturer to determine the location of the active site or other specific regions of a protein other than through DNA sequencing coupled with recognition of conserved motifs. There are however a number of techniques commonly used by researchers. For example, changes in molecular weight caused by a mutation can be readily detected using electrospray ionization (ESI) mass spectrometry. However this instrumentation would likely not be available to an enzyme manufacturer. The ideal way to identify the active site is the determination of the structure of the enzyme, containing its substrate by crystallography. Academic researchers and the pharmaceutical industry are increasingly using this. Another way to identify the active site is to use affinity labeled or "suicide" substrates, if they are available. The labeled active-site amino acids can then be identified after peptide hydrolysis. However, many manufacturers of enzyme products do not sell pure proteins and they are unlikely to have ready access to the equipment used to locate the active site or other specific regions of an enzyme. Determining the active site of an enzyme activity contained within an impure protein mixture typically is not feasible.

7. EPA wants to assess the efficacy of existing sequencing technologies.

a. How accurate and reproducible are readily available amino acid sequencing techniques and instrumentation?

Both protein and nucleotide sequencing are reliable ways of determining the amino acid sequence of an enzyme. Most protein sequences are deduced from nucleotide sequences as it is fast, cheap and very reliable. The error rate is about 1 in 1000 bases or less, and it is common to obtain >700 bases in a single sequencing run. Nucleotide sequencing is used routinely to obtain enzyme sequences after any sort of manipulation, such as moving to a new vector, mutagenesis, etc.

The expense in determining the amino sequence of an enzyme through amino acid sequencing is illustrated by the following fee structure for protein sequencing at the University of California, San Diego (<http://proteinsequencer.ucsd.edu/>) for users outside the UC System working at a 'For-Profit' organization:

- a. \$46/amino acid residue for amino acids 1-10 (minimum charge is 5 residues)
- b. After 10 amino acids, residues 11 and beyond are charged half-price.

The National Center for Biotechnology Information (NCBI) also provides information describing a typical sequencing process, a complex and highly specialized process.

b. How accurate and reproducible are readily available nucleotide sequencing techniques and instrumentation?

The accuracy of results and reproducibility (actually, the precision) of nucleotide sequencing techniques and instrumentation is largely dependent on the repetition rate of determinations. The base error rate is in the range of 1% for a single determination and decreases by repetition and reading the sequence from alternate (both) ends.

c. Does the accuracy of the result depend on the choice of method?

The accuracy depends on a variety of factors: sample preparation and storage, method, skill of operator, etc. However, there is no absolute measure of accuracy since some small residual of uncertainty remains.

d. How rapidly are sequencing techniques improving or new techniques being developed?

History suggests that improvements and new techniques develop at an ever increasing rate.

e. How reliably can one predict the amino acid sequence of the final gene product based on the nucleotide sequence?

Since nucleotide sequencing is considered highly reliable, one can accurately predict the amino acid sequence of the final gene product based on the nucleotide sequence. The reliability of the amino acid sequence is thus dependent on the nucleotide sequence. Modern laboratory techniques now allow the determination of protein sequences to be performed almost exclusively by decoding of nucleotide sequences.

8. What additional information would be gained, if any, by a more detailed structural description that included in addition to amino acid sequence:

While both sequence and function are the most useful properties for identifying an enzyme, additional information can be gained by recognizing other properties or characteristics of enzymes. Information such as glycosylation sites, coenzymes and cofactors would be helpful. A description of such factors are described below.

a. glycosylation sites (and the composition of these carbohydrate moieties),

Glycosylation is important in some cases, such as in some eukaryotic extracellular proteins. Glycosylation usually has little effect on enzyme activity, although it can affect stability. Expression in bacteria or yeast will usually not do this correctly, so eukaryotic expression systems such as Baculovirus in insect cells are employed. A more important consequence for the health of workers is that glycosylation can change the immunogenicity and allergic reactions due to a given enzyme. Background on glycosylation would refine the description of the enzyme if it

included information on the sites of glycosylation and the glycosyl groups involved.

b. coenzymes (prosthetic groups),

In response to both parts b and c (see below), it should be noted that the terminology of "cofactors" and "coenzymes" is confusing, and it is better to use more precise terms. There are "cosubstrates" which participate in the reaction by binding to the enzyme and then dissociate from it as a product and can be recycled by other enzymes. Prosthetic groups, which are bound to the enzyme, remain unchanged at the end of the reaction cycle. It is possible that modified prosthetic groups may be incorporated, which would change the specificity of the enzyme. This might be deliberate, or an accidental consequence of the host organism used for expression.

c. cofactors

Cofactors remain bound to the enzyme and are required for catalysis. In heterologous expression systems, there is always the possibility that cofactors may not be properly or completely inserted, yielding an enzyme of lower activity. Another possibility is that the host organism attempts to insert an inappropriate cofactor or metal ion. An example is the insertion by *E. coli* of iron-sulfur clusters containing zinc instead of iron (Archer et al. 1994) .

d. other post-translational modifications to residues of the amino acid chain?

In a few instances post-translational modifications of a protein may be required for its activity. For example the conversion of a seryl to a pyruvoyl residue is required in some enzymes that use the pyruvoyl residue in their active site. Regarding other post-translational modifications to residues of the polypeptide chain, there will be variations, depending on whether the protein was designed in the original organism to be exported or secreted from the cell. Certain cellular systems such as the twin-arginine export system in bacteria can either transport the protein through the membrane, or embed the protein in it, depending on the protein sequence. Such amino acid modifications can, in some cases, significantly influence the activity and/or stability of an enzyme preparation.

Source

The **SOURCE** of an enzyme refers to (1) the organism from which the gene encoding the enzyme was derived, i.e., the *original* source and (2) the organism or manufacturing platform (e.g., tissue culture) in which the enzyme is produced, i.e., the *production* source. In the questions below, please consider what scientific support there is for using source information to differentiate among enzymes and what level of detail would be scientifically appropriate for this purpose.

9. What information about an enzyme's structure could be gained by knowing

a. the original source of the enzyme?

Knowing the original source of the enzyme is particularly important where the enzyme preparation is not a purified enzyme, but a crude extract, such as pancreatic juice, which contain

a mixture of enzymes, is often of undefined sequence.

For individual enzymes, the relevance of original source depends on the uniqueness of the enzyme to that organism. A greater understanding of enzyme structure via knowledge of the original source of the enzyme is dependent upon knowledge of the structure of the enzyme produced by a given source organism and the uniqueness of the enzyme to that organism. In most cases, an enzyme structure will be largely conserved among similar organisms in spite of peripheral mutations. Thus, source will have limited value in differentiating enzymes unless the structure of the enzyme from each source is known, and it can be established that the enzymes from different sources are structurally related or different.

b. the production source of the enzyme?

In theory, producing an enzyme with a biological system other than the original organism is done to manufacture the product as an exact copy of the original. If the manufacturing organism changes the enzyme in some way, the change may or may not be significant. In general, comparing enzymes from different sources requires some criteria to define the decision plane of what and when is a difference significant for regulatory purposes.

10. If *original* source information were used as an identification element to discriminate among enzymes, what level of taxonomic specificity (e.g., family, genus, species, subspecies, population, biovar, culture line) would be most scientifically appropriate to use for each of the following categories? What if *production* source information were used? (Note: EPA recognizes that taxonomic revisions may change the names of particular organisms and can utilize mechanisms for normalizing organism nomenclature, but that consideration does not need to be addressed by the panel.)

- a. plants**
- b. animals**
- c. fungi**
- d. bacteria**
- e. other micro-organisms**

If original source information were used as an identification element to discriminate among enzymes, using the lowest taxonomic level available would be the most desirable. Differences among enzymes from different parts of a source organism may occur in some cases.

However, such differences are unlikely to be a general phenomenon of a specific organ or tissue. One cannot assume that all or even a preponderance of the enzymes from one tissue or organ differ from that of another tissue or organ, although they may. Furthermore, two enzymes having the same catalytic action may well exist in the same tissue or organ. For example enzymes may differ at different stages of development (juvenile vs. adult beta-galactosidase). Hence, this issue must be addressed in terms of the formal and informal nomenclature structures.

Not only do organism names change, so do their taxonomic ranks due to these name changes. Nomenclature codes as regulated by the scientific community only extend down to the

subspecies level. Even in the official codes, there is no central adjudication of the proper use of a taxonomic level or its application to a specific biological entity (exception: virology). There is no universally accepted structure below subspecies. In addition, there are common practices that vary among biological groupings. The current status of Codes of Nomenclature and some of their attributes is contained in Table 1.

11. How could source be described if taxonomic names were inappropriate because either the original or production source were artificial? Examples of such new technologies could include enzymes produced/developed through gene splicing or ex vivo chemical synthesis.

As noted previously by the Panel, the original source of the enzyme is an important descriptor. When the production source is artificial, i.e. *in vitro* translation, the system used to produce the enzyme will be a defined one and thus can be identified. To date the synthesis of a truly artificial enzyme by ex vivo chemical synthesis is rare, and is not likely to be used commercially. Enzymes produced/developed through gene splicing are derived from defined sources and usually contain the backbone from a specific enzyme that is then modified. Such enzymes may be composed of parts of more than a single original source enzyme, but these can be defined.

Generally the structure of an engineered protein will be related to a naturally-occurring protein, even after extensive modification. The components of a chimeric enzyme (i.e. one that is engineered from components of other enzymes) can be described in terms of its various sources, and the modifications made to it. In addition, information on the vector to which the protein is introduced is helpful. In some cases it could be important to know the source, location and population of the enzyme. However as noted previously by the Panel, modifications to the sequence can have drastic effects on the activity and specificity of the enzyme.

12. What information about an enzyme's structure could be gained by additional details about source including:

a. the particular tissue or organ of a given source organism from which they were derived (e.g., swine pancreatic tissue vs. swine salivary glands)?

Differences among enzymes from different parts of a source organism may occur and should be taken into account. They are termed iso-enzymes (or just isozymes), and arise from a variety of different mechanisms, including duplicated genes, introns, and alternative gene splicing. The distribution of isoenzymes is a feature of the specific organ or tissue. Although one cannot assume that all or even a preponderance of the enzymes from one tissue or organ differ from that of another, there are known differences among enzymes from different parts of a source organism. Those cases need to be defined.

b. the chemical, geographic, and/or environmental conditions from which source organisms were isolated (e.g., soil, water, feces, etc.)?

For unknown enzymes, only characterized by an activity, the chemical, geographic, and/or environmental conditions may be a useful descriptor. Some structural information may well be associated with place of isolation. High temperature environments may select for thermostable enzymes. Hypersaline environments are sources of halophilic bacteria, which accommodate their environment by using high concentrations of ions to maintain internal pressures. Their enzymes are in some cases highly stable, though they may require high salt concentrations for stability outside of the cell.

c. manipulations of the enzyme's *original* source prior to gene transfer (e.g., through rDNA technology, radiation treatment, altered rearing conditions, etc.)?

It is unlikely that manipulation of the enzyme's original source would produce a stable transmissible change. Even if such manipulation had an effect on enzyme structure, the specific nature of the change would be unpredictable in the general sense and would have to be determined empirically.

d. manipulations of an enzyme's *production* source prior to and/or following gene transfer?

The conditions noted in response to part c above are applicable to part d.

e. other relevant aspects of source that are not mentioned (please specify what would be of value).

While the conditions of growth of the production organism will affect the characteristics of the product, such as restriction or supplementation of nutrients, trace elements or vitamins, there are no other specific aspects of source outside those mentioned in sections 12a through 12b that would add additional information about the structure of an enzyme.

Processing

The PROCESSING of an enzyme refers to procedures used to isolate the enzyme from the production organism or manufacturing platform, procedures used to purify the enzyme, and/or any chemical reactions to which the enzyme is subjected to produce the final enzyme product. In the questions below, please consider what scientific support there is for using certain processing information to differentiate among enzymes and identify the level of detail that would be scientifically appropriate for this purpose.

13. What information about an enzyme's structure could be gained by knowing which of certain processing techniques were used in its production?

Processing techniques used in the production of an enzyme will generally not affect its structure. The methods used for purifying enzymes depend on various general properties of the protein. As an example, ion exchange chromatography separates on the basis of charge and gel

filtration, size and possibly shape. These are the commonly used techniques that are not specifically applied to a particular enzyme. Affinity chromatography, which is now the preferred approach, can be directed at catalytic site organization but one usually needs to have some information ahead of time to make this effective. In most cases, since there is no significant purification for most commercial industrial enzymes, it seems unlikely that any useful information about the enzyme's structure will be gained from a consideration of its processing. Factors in production that have a bearing on the structure of the enzyme could include treatments that affect the protein's oligomeric state, such as cross-linking, and immobilization to a solid support. Knowledge of processing techniques could be useful in an exclusionary sense. Thus, if the process does not yield a pure, single protein, then it will be immediately obvious that many informational details delineated by the Panel will be unattainable.

14. EPA anticipates that certain processing techniques may be so routine and/or chemically inconsequential that their reporting would be unnecessary, while other processing techniques would have significant effects on the chemical structure and/or properties of an enzyme. The Agency is trying to assess how practical it would be to create a list of processing techniques that need not be included as part of enzyme identity.

a. What processing techniques are used in the isolation and purification of enzymes?

There are a host of methods for obtaining a purified protein (enzyme) but few if any of these seemed to be used in making TCSA inventory enzymes. These utilize various properties of proteins such as charge and size. There are also various forms of affinity interactions (substrate mimics, antibodies, tags etc). These are indeed relatively routine in their application and a list could be created. However, new affinity purification methods are changing regularly and since they are not thought to be generally employed for the commercial enzyme preparations to be included in the TCSA inventory, there seems to be little advantage in doing so.

b. Which processing techniques could change the chemical structure of the enzyme? Which could change chemical properties that would indicate an underlying structural change?

By and large, most methods for isolation of enzymes are designed to NOT alter proteins and ones that do so are usually not used. However, there are some well-established methods that are known to cause changes. Exposure to heat, pH extremes, and oxygen that can routinely change structural features (such as the oxidation of methionine and the deamidation of side chain amides). The recent advances in mass spectrometry analysis (MALDI, ESI) have revealed that this occurs to a much larger extent than was previously realized, but it is unlikely this analysis would be applied to manufactured enzymes. Many of these changes are also time-dependent and occur simply as a function of time. Thus, they cannot be strictly tied to the processing step, although there are certainly conditions (such as noted above) that will accelerate these changes. Importantly, in considering any structural alterations from the 'native' state, one should only consider covalent modifications and not transient changes such as proton association/disassociation.

A commonly-used step particularly for crude enzyme preparations is autolysis; the cells containing the enzyme are allowed to digest for some time. This process depends on proteases and lipases; sometimes these are derived from the cell, sometimes added to the preparation. They help to release insoluble proteins such as those that were anchored to the membrane, but the product is often cleaved in several places, and may have sections removed. Another method that is used to isolate membrane-bound proteins is the use of detergents.

It is also important to consider, particularly in mixtures of enzymes that characterize commercial preparations, that purification steps can separate cofactors or other modulators, which would in turn affect activity.

c. Describe the chemical or structural changes expected to occur from the use of the processing techniques identified in 14(b).

The structure of the protein is determined by means such as crystallography, on purified preparations. Although processing may lead to modifications of amino acids, the central structure should remain the same. There is a difference between the applications to purified enzymes as opposed to unpurified enzymes. The types of covalent modifications that might be encountered were described above.

d. Which processing techniques would not be expected to cause any structural changes to the enzyme? Which would not be expected to cause any chemical property changes?

Most methods involving chromatography or electrophoresis should not affect enzyme structure.

15. EPA is trying to anticipate whether inclusion of processing in enzyme identity will increase in importance as a result of future advances in enzyme production.

a. What new processing techniques are being developed?

Since industrial enzymes are relatively inexpensive products, new techniques are slow to be implemented. Established processing techniques that have been used for decades have been refined to the point that their efficiencies are very high. There is little room in the production cost of an industrial enzyme to absorb the cost of implementing a new processing technique in order to improve the profit margin in the final product. New processing techniques that are being developed are aimed at creating new products that will support a premium price in the market place. One example of these new technologies would be to change the “format” of an enzyme from a water-soluble enzyme to an immobilized solid state enzyme. Another example would be techniques to selectively remove an undesired enzyme activity that prevents the product from being used in a particular industrial application. One method of purification of enzymes is via the use of affinity chromatography. This process is expensive and often requires modification of the gene coding for the enzyme. When affinity tags are added to an enzyme

structure, they may have to be removed in another processing step in which case, because of the cost involved, only those enzyme products that can support premium pricing will be candidates for purity enhancement using affinity techniques.

b. How might these techniques change an enzyme's chemical structure or properties?

Most of the processing techniques do not significantly change an enzyme's chemical structure or its properties. There are immobilization techniques that change the solubility of an enzyme but not its catalytic function. Except for solubility, immobilization techniques do not change the chemical structure or properties of an enzyme. Affinity purification techniques that do not require enzyme modification will not alter the enzyme's chemical structure or function. At this time, modification in a gene coding for an enzyme for the purpose of purification is far too expensive for use in the production of industrial enzymes, though it may be used in the future.

c. How frequently are new processing techniques for enzymes adopted?

The adaptation of new processing techniques cost a lot of money and can only be cost effective for new products that can support a premium price. The majority of new industrial enzymes are processed via the well-established techniques that have been in place for decades.

Other/General Questions:

16. Aside from function, sequence, source, and processing, are any other data elements crucial for enzyme identification?

The Panel considered assay conditions as a possible data element for enzyme identification, but was unable to determine whether assay conditions itself should be considered as a part of the function data element. Other conditions that could be considered have been referred to by the Panel in response to previous questions.

The minimum structural requirement for a substrate of an enzyme reflects the binding site requirements. This information would allow one to predict other materials that would serve as a substrate besides the one used to describe the function in the nomenclature.

The mechanism of catalysis used by an enzyme adds additional information about an enzyme. There are several enzymes that act on the same substrate and produce the same products but do so using different mechanisms.

A number of enzymes contain allosteric sites that bind modulators to enhance or control activity. The presence or absence of these sites could be used to identify different enzymes catalyzing the same reaction. Salts (both cations and anions) serve as modulators for a number of enzymes. There are many examples of enzymes having identical catalytic function,

but differ as to the ion that modulates the activity.

Many hydrolytic enzymes (proteases, lipases, amylases, etc) display different levels of activity based on the substrate used in their assay. Therefore when reporting enzyme activity level, it is important to identify the substrate used in the assay as well as the enzyme assay conditions.

Industrial enzymes can be very impure products and often contain many enzyme activities along with the marketed enzyme. An identification element could be constructed around the degree of purification from the mixture produced during the fermentation or isolation process. Several industrial enzymes are often produced from a common fermentation or isolation. They often differ by the removal of selected enzymes that prevent their use in an industrial application.

17. Are there any special considerations that should be taken into account when identifying enzymes with multiple, non-identical subunits? For example,
a. when only one subunit is modified?

It is important to specify the nature of the whole enzyme complex as well as the individual subunits. When a subunit of a multienzyme complex is removed, it may show much lower activity. In addition, the reaction specificity may be altered or relaxed. In addition, if a subunit fails to be inserted, the remaining protein often fails to assemble correctly. This can lead to loss or modification of structure. Activity may be decreased and specificity altered. Therefore, it is important to specify the nature of the whole complex as well as the individual subunits.

b. when a modified enzyme is a component of an enzyme complex?

The answer for part 17a above applies, whether the catalytic entity comprises one subunit or has multiple subunits. The way in which enzymes containing non-identical subunits are currently described in the EC list is that there is an EC number for the whole enzyme and EC numbers for the individual catalytic entities. This is not always done however, because of the problem of defining the primary EC class when different classes of reaction are involved. The catalytic entity in some complexes is not even an individual subunit; it may be a domain of an extended polypeptide. For example many of the polyketide synthases have a single polypeptide chain comprising multiple catalytic sites that work sequentially to assemble a molecule such as an antibiotic. These catalytic domains behave essentially like subunits, but since they are a single polypeptide they are more stable.

c. when a multi-functional, multi-component enzyme performs a sequence of reactions?

In a case such as the polyketide synthases noted above, the catalytic sites can be identified separately.

d. when an enzyme has another non-catalytic function, e.g., a binding site?

If relevant, this could be described as another function.

e. under any other circumstances?

The ratio of subunits may be relevant, such as when a complex contains too many or (more commonly) too few of a subunit. If a subunit fails to be inserted, the consequent failure to assemble correctly can lead to loss or modification of structure; activity may be decreased and specificity altered.

As presented previously, enzymes have been shown to contain more than one catalytic activity. It is often the case in multifunctional enzymes that the various functions are carried out by different subunits of the molecule. Taking into account the differences in subunit structure in enzymes made up of non-identical subunits would add little if anything to their identity. It is sufficient to only identify the catalytic activities present within the enzyme.

Any of the four conditions above (a-d) potentially can differentiate among multiple subunit enzymes and indeed between multiple subunit enzymes and single unit enzymes having similar activity properties. Thus, the knowledge of the existence of such conditions enhances the ability to detect “new” enzyme products.

18. Although EPA believes that all four identification elements are critical for enzyme identification for TSCA purposes, the Agency is trying to judge their relative importance.

a. Do any data elements warrant greater emphasis than others because differences in those data element(s) reflect more significant differences in an enzyme’s physical and/or chemical properties than the others do?

The identification of enzyme function is of prime importance, and sets the context for all else. However, function, by itself may be too broad for definitive consideration if novelty is the objective. The existing registries have to include other information. Simple examples include differentiation of isoenzymes, names of source organisms in some cases, etc. For reasons given previously, in some cases the same enzyme activity commands different identifiers in the lists. The sequence, if available, may be helpful if compared sequences are the “same” (as determined by similarity overall and/or of reactive site), then novelty is excluded. The situation is more problematic as similarity decreases but the activity remains the same. Source is less helpful as it is not consistent in its applicability. Still, source should not be ignored as it can give a general context for an enzyme, particularly for crude enzyme containing extracts. However, the practicality of obtaining reliable data from the applicant may inhibit utility. For example, a statement that the enzyme was obtained from mouse spleen raises the question as the common name “mouse” is defined variously as covering one genus of rodent or two depending on the dictionary consulted.

Processing can be of importance. The first processing consideration is the purity of the

preparation. A pure enzyme protein can be evaluated in ways that are denied for a mixture of substances.

b. If data for sequence, source, and processing were the same for two enzymes (at the level of detail you have determined to be appropriate in the questions above), what additional information about chemical structure and/or properties would be provided by distinguishing the enzymes based on function?

The implication of the sequence being the same for two enzymes is based on the assumption that the protein preparations being pure to enable determination of the two enzymes' sequences. If the sequence were truly identical and without modification, then one would expect everything else to be the same. There may be cases in which multiple conformations exist, but knowing the source and processing for identical sequences and the identical activities of the proteins is unlikely to define one or the other enzyme as new. The question becomes moot in the common situation where enzyme preparations are impure, rendering sequence determination impractical. In such situations, as much information as practical from all the categories proposed by the Agency will aid in determination of relative sameness or difference. Without a determination of sequence identity, only relative similarity can be determined. Thus, the more kinds of information gathered, the greater the possibility to determine whether two enzymes with the same activity are the same or different.

c. If data for function, sequence, and processing were the same for two enzymes (at the level of detail you have determined to be appropriate in the questions above), what additional information about chemical structure and/or properties would be provided by distinguishing the enzymes based on (1) original source and (2) production source?

If the sequence were truly identical and without modification, then one would not gain much new information based on the original source or the production source. If on the other hand the sequence was not totally defined, information on the production source could be useful.

Limited useful information would be derived by deducing the characteristics of the protein from (1) original source and (2) production source. The same enzyme may differ in source and still be the same enzyme, or the same source may yield two enzymes with the same activity.

d. If data for function, sequence, and source were the same for two enzymes (at the level of detail you have determined to be appropriate in the questions above), what additional information about chemical structure and/or properties would be provided by distinguishing the enzymes based on processing?

The conditions noted in response to parts 18.b and 18.c above are applicable to part 18.d.

REFERENCES

- Archer VEV, Breton J, Sanchezgarcia I, Osada H, Forster A, Thomson AJ and Rabbitts TH. 1994. Cysteine-Rich Lim Domains Of Lim-Homeodomain And Lim-Only Proteins Contain Zinc But Not Iron. *Proceedings of the National Academy of Sciences of the United States of America*. 91: 316-320.
- Carter P and Ells JA. 1988. Dissecting The Catalytic Triad Of A Serine Protease. *Nature*, 332: 564-568.
- Hicks MG and Kettner C. 2003. Experimental Standard Conditions of Enzyme Characterizations. pp. 203-213, Beilstein Institute, Frankfurt, ISBN 3-8325-0727-2 <http://www.beilstein-institut.de/escec2003/proceedings/>
- Kane JF, Violand BN, Curran DF, Staten NR, Duffin KL, and Bogosian G. 1992. Novel In-Frame Two Codon Translational Hop During Synthesis Of Bovine Placental Lactogen In A Recombinant Strain Of *Escherichia coli*. *Nucleic Acids Res.* Dec 25;20(24):6707-12.
- Rosenberger, RF 1994a. Translation Errors During Recombinant Protein Synthesis In Genetic Stability And Recombinant Protein Consistency. Brown F. and Lubiniecki As (eds) *Dev. Biol. Stand.* Basel, Karger, 1994, vol 83, pp 21-26.
- Rosenberger RF. 1994b. Translational Errors During Recombinant Protein Synthesis. *Dev. Biol. Stand.* 83, 21-26.
- Scorer CA, Carrier MJ, and RF Rosenberger. 1991. *Nucleic Acids Res.* 19: 3511–3516.
- Weickert MJ and Apostol I. 1998. High-Fidelity Translation of Recombinant Human Hemoglobin in *Escherichia coli* *Appl Environ Microbiol.* 64, 1589–1593.

Table 1. Status of Nomenclature Codes

Org. acronyms	Orgs. Overseeing/ man. Code	Code	Contact information	Nname for mat.	Physical nature of deposit	Requirement
ICSU	International Council for Science		www.icsu.org			
IUBS	International Union of Biological Sciences		www.iubs.org			
ICZN	International Commission on Zoological Nomenclature	International Code of Zoological Nomenclature	www.iczn.org	Type specimen	Nature varies with organism - deposition in a recognized institution not required by the Code, most authors do, making specimens available.	Now must be designated and clearly identified for any species described after 2000
IAPT	International Association for Plant Taxonomy	International Code of Botanical Nomenclature	www.botanik.univie.ac.at/iapt/	Type	single plant, parts of one or several plants, or of multiple small plants - usually mounted on a single herbarium sheet or in an equivalent preparation, such as a box, packet, jar or microscope slide	either a single specimen conserved in one herbarium or other collection or institution, or an illustration for any species described after 1990
ICNCP	International Commission for the Nomenclature of Cultivated Plants	International Code of Nomenclature for Cultivated Plants	www.ishs.org/sci/icracpe.o.htm			
IUMS	International Union of Microbiological Societies		www.iums.org			
ICSP	International Committee on Systematics of Prokaryotes	International Code of Nomenclature of Bacteria	www.the-icsp.org	Type culture	living culture	Publication in the International Journal of Systematic and Evolutionary Microbiology and deposited in two recognized culture collections in two different countries
ICTV	International Committee on the Taxonomy of Viruses	International Code of Virus Classification and Nomenclature	www.ncbi.nlm.nih.gov/ICTV/rules.html	Physical type not used	Nomenclatural type description serves this function	Accepted description maintained by ICTV

The Internet contact information links to the most current versions of the Codes, decisions on nomenclature, information on the history and administration of the Codes and other relevant information.

The official printed versions of the Codes are necessarily out of date as of the publication in that they obviously cannot contain changes post printing.

