MEMORANDUM

SUBJECT: Document to Support the FIFRA Scientific Advisory Panel Meeting on TSCA Inventory Nomenclature for Enzymes and Proteins

FROM: Neil Patel, Associate Director
Economics, Exposure and Technology Division
Office of Pollution Prevention and Toxics

TO: Clifford Gabriel, Director
Office of Science Coordination and Policy

Attached is the supporting document for the FIFRA Scientific Advisory Panel Meeting on TSCA Inventory Nomenclature for Enzymes and Proteins
scheduled for May 3-4, 2005 at the Holiday Inn Rosslyn at Key Bridge.

The attached document is: Treatment of Enzymes on the TSCA Chemical Inventory.
U.S. Environmental Protection Agency (2005).

If you have any questions, please contact Dr. Henry Lau at (202) 564-8572 or myself at (202) 564-8804.

Treatment of Enzymes on the TSCA Chemical Inventory

I. ISSUE

EPA's Office of Pollution Prevention and Toxics (OPPT) is responsible for implementing the Toxic Substances Control Act (TSCA). It requires EPA to ensure that

chemicals sold and used in the United States pose no unreasonable risks to human health or the environment. TSCA also requires EPA to maintain a list of all chemicals existing in U.S. commerce as a way to identify new substances that should be reviewed for potential risks.

In 1977, EPA promulgated the Inventory Reporting Regulations that provided the basis for the first compilation of the TSCA Chemical Substance Inventory ("the TSCA Inventory" or "the Inventory"). It listed more than 62,000 substances, including about 150 enzymes. The 1977 regulations did not specifically address how enzymes should be identified when reported to EPA. Consequently, most enzymes were only broadly defined, with some individual listings actually describing a whole category of commercial enzymes. EPA is therefore unable to identify when an enzyme is "new" and should be subject to Premanufacture Notification (PMN) reporting requirements that enable the Agency to review new chemicals before introduction into U.S. commerce. Specific reporting guidance for enzymes is needed so that enzymes will be accurately, uniquely, and unambiguously identified on the TSCA Inventory.

EPA is currently working to develop an improved and standardized method of identifying proteinaceous enzymes (hereafter "enzymes") on the TSCA Inventory so that each Inventory listing will cover only one enzyme. EPA issued an Advanced Notice of Proposed Rulemaking (ANPR) in November 2004 that outlined its proposed approach to enzyme identification, and solicited comment on all aspects of the proposal. In the ANPR, EPA proposed that enzymes be identified using information on function, source, processing, and sequence. EPA is convening a group of technical experts under the aegis of the Scientific Advisory Panel to evaluate several scientific and technical issues associated with this proposal.

II. THE TSCA INVENTORY

Congress recognized during its development of TSCA that very little was known about chemicals in commercial use. When TSCA was enacted in 1976, it was not even known how many chemicals there were, in what quantities or where they were produced, what their byproducts or impurities were, or who was exposed to them under what conditions. Therefore, Congress gave EPA the authority to compile an Inventory of existing chemical substances and to develop additional information on these basic questions.

The first inventory was published in 1979, based on information reported to EPA by chemical manufacturers, importers, and processors. The Inventory - to which new chemicals are added when they go into production - shows now that over 82,000 commercial chemical substances are, or have been, manufactured or imported into the

United States since January 1, 1975. (Most of the more than twenty-two million known chemical substances in existence are excluded from TSCA reporting, such as those substances used only for research and development[1].)

It is important to note that *the Inventory is not a list of toxic or hazardous chemicals*. Rather, it lists existing chemicals, generally by their specific chemical name, giving an overall picture of the chemicals used for commercial purposes in the United States. Chemicals not on the Inventory and not specifically excluded from TSCA's regulatory authorities or exempt from PMN reporting must be reviewed by EPA under the PMN program before they are allowed into U.S. commerce.

Premanufacture notices submitted for new chemicals or significant new uses of existing chemicals are to include: the specific chemical identity (the chemical name) of the substance; its molecular structure and molecular formula; proposed categories of use; an estimate of the amount to be manufactured, imported, or processed; its impurities; the byproducts resulting from the manufacture, processing, use, and disposal of the chemical; estimates of the exposure; and any available test data related to the health and environmental effects of the chemical. Inventory names are developed from this information according to a rigorous, comprehensive set of nomenclature rules to ensure that a single, unique, and unambiguous name can be constructed for each chemical substance. In most cases, the preferred Chemical Abstracts (CA) Index Name for a substance contains sufficient information to reflect a specific, corresponding chemical structure.

Some substances listed on the Inventory are tagged to indicate that their preferred CA Index Names are not sufficient or complete enough to permit unambiguous identification of the substance. For such substances, the Inventory lists a supplemental definition. Definitions may include a genus/species or identity of other types of source material, the nature of the process by which the substance was manufactured or processed , the predominant components, and characterizing physical data such as approximate boiling point range. The information EPA is proposing to require for identification of enzymes on the TSCA Inventory is thus similar to information that is already provided for other substances when it is needed to create a unique, unambiguous description. More detailed information about current Inventory practices and examples of current Inventory listings are available in Appendix I.

[1] Eight product categories are exempt from TSCA's regulatory authorities: pesticides, tobacco and tobacco products, certain nuclear materials, firearms and ammunition, food, food additives, drugs, and cosmetics. Many of these product categories are regulated under other Federal laws. Chemicals produced in small quantities solely for research and development purposes are also exempt. In addition, any person may apply for an exemption from PMN reporting for chemicals used solely for test marketing purposes or for those determined by EPA not to present an unreasonable risk of injury to human health or the environment.

III. CURRENT TSCA INVENTORY LISTING CONVENTIONS FOR ENZYMES

Enzymes currently on the TSCA Inventory are identified by a Chemical Abstracts Service (CAS) Registry Number[2] and Chemical Abstracts (CA) Index Name[3] with supplemental definitions of varying detail. These listings are based on information that submitters provided to EPA without having any agency guidance as to the type or specificity of information that should be included in enzyme descriptions.

In general, the type of information included in current enzyme listings consists solely of the category of the enzyme's catalytic activity (although in a few cases supplemental definitions include additional information such as source, identity of co-factors, or processing techniques used in the enzyme's production). Enzyme catalytic activity has historically been the basis for CAS and other enzyme nomenclature systems. Often a new catalytic activity is discovered before anything is known about the structure of the catalyst. In addition, a satisfactory, systematic nomenclature for protein structures that is useful for EPA's regulatory purposes does not yet exist.

Using function to identify enzymes on the TSCA Inventory is an approach that differs substantially from that used for other listed chemicals. CA Index Names generally allow derivation of the chemical structure (or as much of the chemical structure as is known). Structure provides a way to unambiguously identify chemicals so that they can be differentiated and thus new chemicals can be recognized. Function allows no such differentiation because often several structurally-diverse enzymes have the same function. Thus, even under ideal circumstances, a function-based enzyme nomenclature system would inappropriately subsume multiple enzymes under a single listing on the TSCA Inventory. This problem is compounded on the current TSCA Inventory because many descriptions of enzyme function that were submitted to EPA defined only the most general catalytic activity of the enzyme (e.g., proteinase). Both the type and specificity of information included in current enzyme names on the Inventory therefore create listings that often encompass several enzymes that have structural differences and are recognized to be distinct products. In fact, most new commercial enzymes that are created, regardless of similarity or lack thereof, could be subsumed under one of the current listings.

---

[2] A CAS Registry Number is included for each nonconfidential substance on the Inventory. CAS Registry Numbers are assigned in sequential order as new substances are entered into the CAS Registry File. Each Registry Number in the TSCA inventory designates only one chemical substance in terms of numbers and types of atoms, the bonding between atoms, and stereochemistry, insofar as that substance has been elucidated or defined. The fact that a substance has been assigned a CAS Registry Number does not define the substance as being on the TSCA Inventory. The non-confidential portion of the TSCA Inventory includes only approximately 80,000 substances in the CAS Registry File, which represents less than one percent of the total number of CAS Registry Numbers.

[3] Preferred CA Index Names are derived according to a rigorous, comprehensive set of nomenclature rules to ensure that a single, preferred name can be constructed for each chemical substance. For substances that have definite chemical structures, the preferred CA Index Name contains sufficient information to permit derivation of those structures.

IV. REASONS TO UPDATE CURRENT TSCA INVENTORY LISTING CONVENTIONS FOR ENZYMES

EPA believes that current reporting conventions do not allow the Agency to fulfill its obligation to identify and review new enzymes before they are manufactured or used in the United States. Enzyme listings must allow the Agency to determine that an enzyme is new and is therefore subject to reporting requirements under TSCA so that the Agency could review the chemical for potential risks to human health or the environment. Enzymes are used in a wide variety of products and industries including detergents; cleansing and degreasing agents; leather, textiles, and paper manufacturing industries; waste degradation; oil recovery; and biosensing. Many of these uses expose large populations to products containing enzymes, and the potential exposure is constantly increasing as new enzymes and uses are discovered. Nevertheless, the risk implications of many enzymes produced since the initial Inventory compilation have not been evaluated by EPA because the current enzyme listings do not allow the Agency to differentiate among many widely disparate enzymes currently in commerce. Broad Inventory listings that may encompass hundreds or even thousands of distinct substances prevent the Agency from fulfilling the mandate of TSCA by undermining EPA review of all new chemicals.

Broad Inventory listings also lead to uncertainty for EPA and the regulated community. The lack of clarity surrounding enzyme listings has led to an increasing number of inquiries from chemical manufacturers who need assistance in determining whether an enzyme they plan to manufacture is already listed on the Inventory or whether they will have to submit a PMN notification to the Agency. Resolving this uncertainty creates an unnecessary burden for both EPA and the regulated community, as the Agency must make a case-by-case determination for each enzyme, many of which are ultimately found to be subsumed under a current Inventory listing.

V. CHALLENGES ASSOCIATED WITH LISTING ENZYMES ON THE TSCA INVENTORY

The current function-based approach for listing enzymes on the Inventory has become increasingly inadequate for TSCA Inventory listings as the number and diversity of enzymes has grown. A disparate set of enzymes may share a single enzymatic function, even one that is narrowly defined. However, there is no other standard, scientifically-recognized nomenclature system for enzymes with which to replace the current function-based system. For non-enzymatic chemicals, EPA relies heavily on a structural approach to identification. However, adopting a strictly structure-based approach for enzymes on the Inventory would be impractical.

Proteinaceous enzymes consist of polypeptide chains. The amino acid sequence ultimately determines how the chains fold to form three-dimensional (3-D) structures with active sites for binding to particular substrates. In some enzymes, proteins are the only structural components. However, most enzymes also contain covalently-bonded, non-protein moieties such as carbohydrates that can have a significant effect on an enzyme's 3-D structure and function. Other important factors include metal ions (called cofactors) and low molecular weight organic molecules (called coenzymes).

A complete structural description of an enzyme would need to describe, at a minimum, the complete amino acid sequence and the location and identity of all non-protein moieties. Determination of such a precise description would be technically difficult. However, even if this information were readily available, such a detailed description would so finely differentiate enzymes, that even enzyme preparations that are considered the same for all practical purposes would have unique identities. As is the case for most biological substances, a certain degree of genetic variation can be accommodated without altering the function or three-dimensional structure of an enzyme.

Genetic variation in what is consistently considered to be a single enzyme can arise at least three different ways: 1) an enzyme sample may contain differences due to inherent genetic variability within the population of an organism serving as production source at any point in time, 2) an enzyme sample may change over time as the genetics of the predominant organism serving as production source change, and 3) fluctuations in environmental conditions during enzyme production may produce variation.

V. A. Inherent genetic variability of the source organism

Most commercial enzymes are produced by microbial fermentation of bacteria or fungi. A large population of microorganisms must be cultured to enable production of a sufficient quantity of enzyme product. A bacterial population may have a spontaneous mutation frequency of 10^{-6} or 10^{-7} mutations per gene per generation. With a production volume of up to 100,000 liters at a density of 10^{12} colony-forming units per milliliter, the fermenting population is likely to contain genetic variation at any point in time, even when accounting for DNA repair mechanisms. This variation may result in some members of the population producing slightly different enzyme structures from the same genetic locus. The result is an enzyme preparation generally dominated by one structure but containing inherent variability.

Enzymes may also be produced by extraction and isolation from plant or animal sources, for example, from swine pancreatic tissue or from papaya. Natural genetic

variation will exist within a populations of plants and animals that may lead to inherent variation in these types of enzyme preparations as well.

V. B. Variability of the source organism over time

Even when initial variation within a cell culture is minimized, continued growth of the microorganism population throughout enzyme production provides opportunity for the introduction of genetic variation through mutation. Just as with any population, genetic drift (changes in allele frequency due to chance alone) and selection (changes in allele frequency due to differential survival of organisms that are better able to use environmental resources) may cause the precise genetic makeup of the population to change over an extended production period. Changes in the composition of the source population over time may yield slightly different enzyme structures throughout the production period unless regular quality control adjustments are made. The significance of the specific genomic location of these genetic changes will vary. Some point mutations will produce no change in the amino acid sequence of the enzyme, while others may produce a conformational change. The latter may not necessarily affect specific enzymatic activity and could thus evade detection if only enzymatic function were evaluated.

V. C. Variability produced by differences in production conditions

Enzyme structure may also be affected by environmental conditions during enzyme production. The goal of a given production facility will generally be to maintain constant conditions to the extent possible. However, variation in enzyme product may be caused by changes in pressure, density of the population, substrate availability, osmotic pressure, aeration, and oxygen availability among other factors. Quality control may involve ensuring that the final enzyme product has the desired functionality but is unlikely to involve verification of the enzyme structure in spite of probable non-sequence structural variations.

V. D. Apparent variability

In addition to the *actual* variation in an enzyme sample, variation will *appear* to occur due to limitations in our ability to accurately determine and describe an enzyme's structure, limitations that are exacerbated by the relative impurity of many enzyme preparations. Product impurity and other sample preparation problems can lead to incomplete or reduced quality sequence information.

Some enzyme products may be highly-purified materials for analytical or medical use. Other applications may require only very crude preparations such as ground and/or extracted plant material or animal organs. Such crude products are likely to contain only small amounts of the active enzyme with a large proportion of by-products. Purification is only attempted if a more concentrated or purer product is desired because each purification step reduces the total yield. To retain the activity in commercial preparations, stabilizing agents and/or preservatives are typically added to ensure the maintenance of the native conformation by preventing unfolding of the protein and/or its degradation by other enzymes. Some commercial enzyme preparations may consist of enzyme mixtures, either by intent or through a lack of purification. Impurities and byproducts in enzyme preparations may impede the accurate description of the enzyme structure.

VI. EPA'S PROPOSED APPROACH FOR ENZYME IDENTIFICATION ON THE TSCA INVENTORY

EPA wants to establish TSCA nomenclature guidance for enzymes that enables the Agency to unambiguously differentiate among enzymes while taking into account the challenges associated with fitting enzymes into a structure-based system. EPA has developed an approach it believes balances these concerns. EPA intends to use a combination of four elements to create an enzyme's name for listing on the TSCA Inventory: function, sequence, source, and processing (described below). A supplemental description would be part of an enzyme's identity on the Inventory that would incorporate information about each of these characteristics at a level of detail that must still be determined.

None of the four identification elements alone would be adequate to unambiguously describe an enzyme. Historically, function has been used but has proved to be inadequate for EPA to be able to determine clearly and unambiguously whether a potential new enzyme is already listed on the Inventory. Adoption of a strictly structure-based approach for enzyme identification would complicate the differentiation among enzymes by often requiring unnecessary detail. EPA's approach for enzyme identification would retain function as part of enzyme identity to avoid confusion that could result from complete departure from conventional enzyme nomenclature and because function provides useful information. In addition, using one aspect of enzyme structure (amino acid sequence) along with certain source and processing information would provide an appropriate amount of both direct and indirect information about enzyme structure to differentiate enzymes and allow review of unique products.

In convening this expert panel, EPA is soliciting scientific input on how and to what degree changes in each of these elements affects the chemical identity of enzymes. This

information is critical to the Agency's efforts to develop a nomenclature system that will enable EPA to discriminate among distinct chemical substances.

VI. A. 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 differentiate enzymes based on this catalytic activity[4]. The NC-IUBMB assigns enzymes a code number designated by the prefix EC. The code number is based on the reaction catalyzed by the enzyme, the nature of the bond involved, and the substrate acted upon. Enzymes with different amino acid sequences may be assigned the same EC code. EPA intends to incorporate function into enzyme identity on the TSCA Inventory by using these EC codes and the associated function descriptor(s).

Using function follows historically important enzyme nomenclature conventions of the NC-IUBMB that are scientifically accepted, readily interpretable, and widely used for enzyme identification. Although EPA has determined that function alone is not sufficient for describing an enzyme's identity for regulatory purposes, the Agency believes it would be a necessary component of enzyme identity. Information about the type of reaction catalyzed by an enzyme indicates potential and likely uses for the enzyme. A description of function is the only way that multiple catalytic activities in a single enzyme may be identified. Finally, information about function is readily available and would require little additional effort to provide it to EPA. The Agency is seeking guidance as to how specifically function should be described to uniquely identify enzymes if function were to be a part of the enzyme definition.

VI. B. Amino Acid Sequence

The AMINO ACID SEQUENCE of an enzyme polypeptide chain 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. Sequence information is a precise way to characterize the primary structure of an enzyme, and it provides information that is analogous to the structural information that EPA uses to differentiate among most standard chemicals. In addition, changes in sequence may lead to changes in important chemical properties that should be reviewed by the Agency. The Agency is seeking guidance as to how much variation in amino acid sequence would be

[4] For more information about classification and nomenclature of enzyme-catalyzed reactions by the IUBMB, see http://www.chem.qmul.ac.uk/iubmb/enzyme/rules.html

appropriate to uniquely identify enzymes if amino acid sequence were to be a part of the enzyme definition.

VI. C. Source

The SOURCE of an enzyme would refer 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. Given that post-translational processes (including but not limited to methylation and glycosylation) may vary with the source and can affect the chemical properties of an enzyme, EPA anticipates that information about an enzyme's source will be useful in precisely and unambiguously identifying and distinguishing between enzymes.

Using production source information for identification of enzymes is consistent with current practices of the scientific community, enzyme manufacturers, and protein database repositories. Source may provide information about differences in chemical structure or chemical properties, including post-translational changes associated with different production sources that would not be reflected in the amino acid sequence. Source information may also allow identification of potential impurities or byproducts associated with particular organisms or manufacturing platforms. Such indirect information about chemical structure or properties would be particularly important when product purity or technical difficulties make obtaining accurate sequence information difficult or even impossible. Finally, as with function, source information is readily available to submitters. The Agency is seeking guidance as to how specifically original and production source should be described to uniquely identify enzymes, if original and/or production source were to be a part of the enzyme definition.

VI. D. 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. Certain processing information would enable EPA to identify chemical differences among enzymes that resulted from the way in which they were isolated and purified. Certain processes may change chemical properties such as structure, product stability, efficacy, and use. The use of certain processing information would allow identification and review of modified enzymes and corresponding reaction byproducts that result from the chemical treatment and would enable EPA to review new techniques for isolating and modifying enzymes. The Agency is seeking guidance

as to what processing techniques would be appropriate to include and at what level of detail they should be described to uniquely identify enzymes if processing were to be included as part of the enzyme definition.

Questions for the SAP Panel

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.

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?

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)?
   b. non-catalytic enzyme functions that are not represented by EC codes (e.g., binding properties)?
   c. other additional information about function that could be used to differentiate enzymes (please specify what would be of value)?

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?
b. How frequently are new catalytic functions for existing enzymes discovered?
c. How good are existing models to assess the likelihood that an enzyme may have several catalytic functions?
d. What information is required to utilize such models?

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?

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.
   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.
   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.
   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?

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)?

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?

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? that differ in amino acid composition by 0.5%? 1%? 10%? etc.?
   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?
   c. How important are deletions and/or excisions in determining differences between enzymes?
   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?

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?
   b. How accurate and reproducible are readily available nucleotide sequencing techniques and instrumentation?
   c. Does the accuracy of the result depend on the choice of method?
   d. How rapidly are sequencing techniques improving or new techniques being developed?
   e. How reliably can one predict the amino acid sequence of the final gene product based on the nucleotide sequence?

8. What additional information would be gained, if any, by a more detailed structural description that included in addition to amino acid sequence:
   a. glycosylation sites (and the composition of these carbohydrate moieties),

b. coenzymes (prosthetic groups),
c. cofactors, and/or
d. other post-translational modifications to residues of the amino acid chain?

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?
   b. the production source of the enzyme?

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

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.

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)?
b. the chemical, geographic, and/or environmental conditions from which source organisms were isolated (e.g., soil, water, feces, etc.)?
c. manipulations of the enzyme's *original* source prior to gene transfer (e.g., through rDNA technology, radiation treatment, altered rearing conditions, etc.)?
d. manipulations of an enzyme's *production* source prior to and/or following gene transfer?
e. other relevant aspects of source that are not mentioned (please specify what would be of value).

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?

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?
    b. Which processing techniques could change the chemical structure of the enzyme? Which could change chemical properties that would indicate an underlying structural change?
    c. Describe the chemical or structural changes expected to occur from the use of the processing techniques identified in 14(b).

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?

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?
    b. How might these techniques change an enzyme's chemical structure or properties?
    c. How frequently are new processing techniques for enzymes adopted?

Other/General Questions:

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

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?
    b. when a modified enzyme is a component of an enzyme complex?
    c. when a multi-functional, multi-component enzyme performs a sequence of reactions?
    d. when an enzyme has another non-catalytic function, e.g., a binding site?
    e. under any other circumstances?

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?
    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?
    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 original source? production source?

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?

Appendix I: Definitions and Examples of Substances on the TSCA Inventory

I. Class 1, Class 2, and UVCB Substances

The TSCA Inventory lists chemical substances with diverse characteristics, and Inventory nomenclature conventions vary accordingly. In terms of composition, some chemical substances are single compounds composed of molecules with particular atoms arranged in a definite, known structure. EPA designates such substances as Class 1 substances. Examples of Class 1 substances include: acetone, iron, benzene, and dimethylmercury.

Some commercial chemical substances have unknown or variable structures or compositions, or they are composed of a complex combination of different molecules. EPA designates these as Class 2 substances. Many Class 2 substances (including enzymes) are called UVCB substances, for "chemical substances of **u**nknown or **v**ariable composition, **c**omplex reaction products and **b**iological materials." Each name for a UVCB substance represents more than one molecular entity; as such, each UVCB can be considered to be a narrowly-defined category of molecules, often closely related. Examples of Class 2 UVCB substances are listed below.

II. Inventory Definitions for UVCB Chemical Substances

Entries on the Inventory identify the commercial chemical substance as precisely as possible using the information reported by the submitter. Listings depend on the degree of knowledge that the submitters possess and report about such substances as well as on how submitters intend to represent the chemical identities to the Agency and to customers. Therefore, sometimes substances that are chemically indistinguishable or even identical have different listings on the Inventory. On the other hand, sometimes a lack of detail in submitted information results in UVCB substances with Chemical Abstracts (CA) Names that are not specific or complete enough to permit unambiguous identification of the substance. The CA names for such substances may contain:

- Process terms that are not chemically descriptive (e.g., distillation residues, distillation overheads, by-products, low-boiling, catalytic reformed);
- Trade jargon (e.g., slack wax, spelter, winterized, deodorized distillates, steep liquor, foots oil);
- Unqualified or very broadly qualified substance class terms (e.g., pyridine bases,

petroleum resins, phenols (petroleum)); or
- Physical rather than chemical terms (e.g., microcrystalline, pulp, agglomerates, sinter, viscous).

The substances with inadequate CA Names are further described with supplemental "definitions" that are considered an integral part of the name for TSCA purposes. In general, the definitions serve to narrow the scope of the CA Names. Thus, any substance that matches a CA Name on the TSCA Inventory but does not match the Inventory substance definition is not considered to be covered by that Inventory name.

UVCB chemical substance definitions typically begin by stating that the substance is a combination of substances of a certain class and indicating the nature or the process by which it was derived. The next sentence(s) usually identify the predominant components. UVCB chemical substance definitions also may include information such as the typical or allowed carbon number, physical properties ranges, the types of atoms or substances that may be included, and/or the raw material sources or processes of manufacture.

III. EXAMPLES OF UVCB NOMENCLATURE

The following examples of Inventory names are provided to illustrate the variability of Inventory listings. These examples are not an exhaustive representation of all possible terms used in TSCA Inventory names. They are presented only to demonstrate that Inventory UVCB names vary according to the type of chemical substance being described.

III. A. Substances of Unknown Composition

A wide variety of substances of unknown composition are listed on the TSCA Inventory. Examples include distillation residues, spent cooking or neutralizing liquors, and residual oils. Each Inventory substance, however, has a commercial purpose under TSCA; if it were an impurity, a byproduct, or a waste with no commercial use, it would not be eligible for TSCA Inventory listing.

Example 1. Dust, iron-ore, sinter (CASRN[5] 69012-53-9)

[5] CAS Registry Number.

The UVCB definition for this substance is: "Dust generated during the making, breaking and handling of sinter which is recovered through the use of pollution abatement equipment."

Example 2. Fuel oil, no. 2 (CASRN 68476-30-2)

The UVCB definition for this substance is: "A distillate oil having a minimum viscosity of 32.6 SUS at 100°F to a maximum of 37.9 SUS at 100°F"

III. B. Substances of Variable Composition

Many substances with variable composition are listed on the Inventory. The examples below demonstrate how source terms may be used in Inventory names when the submitter reports the use of a single source. Process terms are used if they are provided by the submitter, are well-known, and are well-described.

Example 3. Fatty acids, soya (CASRN 68308-53-2)

This substance has no UVCB definition. It is a UVCB substance because it includes variable carbon chain lengths: soya fatty acids are a mixture of saturated and unsaturated C14-C18 fatty acids.

Example 4. Fatty acids, castor-oil (CASRN 61789-44-4)

This substance has no UVCB definition. It is a UVCB substance because this substance includes variable carbon chain lengths: castor-oil fatty acids are a mixture of saturated and unsaturated C14-C18 fatty acids. This substance is listed separately in the Inventory from Example 3 because the fatty acids are derived from a different source.

Example 5. Fatty acids, castor-oil, hydrogenated (CASRN 61790-39-4)

This substance has no UVCB definition. It is a UVCB substance because this substance includes variable carbon chain lengths. This substance is listed separately in the Inventory from Example 4 because additional chemical processing was carried out on the fatty acids.

Example 6. Natural gas (petroleum), raw liq. mix (CASRN 64741-48-6)

The UVCB definition for this substance is: "A complex combination of hydrocarbons separated as a liquid from natural gas in a gas recycling plant by processes such as refrigeration or absorption. It consists mainly of saturated aliphatic hydrocarbons

having carbon numbers in the range of C2 through C8." Thus, this substance includes variable carbon chain lengths.

III. C. Substances that are Biological Materials or are Made from Biological Materials

Some chemicals derived from biological materials are completely defined in terms of chemical structure and thus are listed as Class 1 substances, e.g., sucrose. However, numerous biological materials and chemicals derived from biological materials are listed as UVCB substances on the TSCA Inventory, either by themselves or as components of further reaction products; including enzymes, organisms, and products of the biotechnology industry. Names of these more complex structures often define the chemical identity with certain elements that characterize the chemical in terms of source, composition, and/or manufacturing processes.

Example 7. Beeswax (CASRN 8012-89-3)

The UVCB definition for this substance is: "The wax obtained from the honeycomb of the bee. It consists primarily of myricyl palmitate, cerotic acid and esters and some high-carbon paraffins."

Example 8. Soybean, flour (CASRN 68513-95-1)

The UVCB definition for this substance is: "A fine-ground powder made by steaming soybeans, followed by removal of hulls and mechanical grinding." If another type of soybean flour were made without steaming the soybeans or without removal of the hulls, the resultant flour would not fit within this definition and would require a different listing on the Inventory.

Example 9. Collagens (CASRN 9007-34-5)

The UVCB definition for this substance is: "A fibrous protein comprising one third of the total protein in mammalian organisms. It is a polypeptide containing three peptide chains and rich in proline and hydroxyproline." In this case source is defined to indicate that the collagen must be from a mammalian species, although no tissue or organ source is specified. Structure is defined in that all single chain proteins are excluded, and proline and hydroxyproline must be abundant.

Example 10. Gelatins (CASRN 9000-70-8)

The UVCB definition for this substance is: "A complex combination of proteins obtained by hydrolysis of collagen by boiling skin, tendons, ligaments, bones, etc."

In this case no specific animal (or tissue/organ) source is given, although the source of the collagen used to produce gelatin must meet the specifications given in the definition of collagen (example 9). Any acid/base catalytic reaction conditions may be used in the hydrolysis, although other processing reactions performed on collagen would not fall under this definition.

Example 11. Gelatins, hydrolyzates (CASRN 68410-45-7)

The UVCB definition for this substance is: "Enzymatic digest produced by hydrolysis of gelatin." If the gelatin were hydrolyzed chemically rather than enzymatically, the resultant product would not fit the definition and would require a different Inventory listing.

Example 12. Oils, lavender (CASRN 8000-28-0)

The UVCB definition for this substance is: "Extractives and their physically modified derivatives. Lavandula officianalis, Labiatae." Note that use of another source even if the same composition of oil were obtained would require a separate Inventory listing since it would represent a different substance.

Example 13. Glutens, corn (CASRN 66071-96-3)

The UVCB definition for this substance is: “The dried residue from corn after the removal of the larger part of the starch and germ and the separation of the bran in the wet-milling manufacture of corn starch or syrup, or by enzymatic treatment of the endosperm.”