Standards for Reporting Enzyme Data: The STRENDA Consortium: What it aims to do and why it should be helpful

TIPTON, Keith F., et al.

Abstract

Data on enzyme activities and kinetics have often been reported with insufficient experimental detail to allow their repetition. This paper discusses the objectives and recommendations of the Standards for Reporting Enzyme Data (STRENDA) project to define minimal experimental standards for the reporting enzyme functional data.

Reference


DOI: 10.1016/j.pisc.2014.02.012
REVIEW

Standards for Reporting Enzyme Data: The STREND A Consortium: What it aims to do and why it should be helpful

Keith F. Tipton, Richard N. Armstrong, Barbara M. Bakker, Amos Bairoch, Athel Cornish-Bowden, Peter J. Halling, Jan-Hendrik Hofmeyer, Thomas S. Leyh, Carsten Kettner, Frank M. Raushel, Johann Rohwer, Dietmar Schomburg, Christoph Steinbeck

Trinity College Dublin, School of Biochemistry and Immunology, Dublin 2, Ireland
Vanderbilt University, Department of Chemistry, 7330 Stevenson Center, Nashville, TN 37235, United States of America
University Medical Centre Groningen, Department of Pediatrics, Hanzeplein 1, 8713 GZ Groningen, The Netherlands
Swiss Institute of Bioinformatics, CMU – 1, rue Michel Servet, 1211 Geneva 4, Switzerland
Centre National de Recherche Scientifique – BIP, 31 chemin Joseph-Aiguier, B.P. 71, 13402 Marseille Cedex 20, France
University of Strathclyde, Department of Chemistry, 16 Richmond Street, Glasgow, Scotland G1 1XQ, United Kingdom
University of Stellenbosch, Department of Biochemistry, Private Bag X1, Matieland, 7602 Stellenbosch, South Africa
The Albert-Einstein-College of Yeshiva University, Department of Microbiology and Immunology, 1300 Morris Park Avenue, Bronx, NY 10461, United States of America
Beilstein-Institut, Trakehner Str. 7-9, 60487 Frankfurt am Main, Germany
Texas A&M University, Department of Chemistry, PO Box 30012, College Station, TX 77843-3255, United States of America
Technical University Braunschweig, Department for Bioinformatics and Systems Biology, Langer Kamp 19b, 38106 Braunschweig, Germany
EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

Received 28 August 2013; accepted 4 November 2013; Available online 21 March 2014

This article is part of a special issue entitled “Reporting Enzymology Data - STRENDA Recommendations and Beyond”. Copyright by Beilstein-Institut.

Corresponding author.
E-mail address: ktipton@tcd.ie (K.F. Tipton).

http://dx.doi.org/10.1016/j.pisc.2014.02.012
2213-0209 © 2014 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).
Abstract

Data on enzyme activities and kinetics have often been reported with insufficient experimental detail to allow their repetition. This paper discusses the objectives and recommendations of the Standards for Reporting Enzyme Data (STRENDA) project to define minimal experimental standards for the reporting enzyme functional data.

© 2014 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

Introduction

In any form of communication it important to understand what others are talking about and in science it is essential for data to be reported in a form that allows others to repeat, verify and apply the determinations. Unfortunately, that has not always been the case with enzyme activity and kinetic data, because insufficient experimental details have been provided. An idea of the nature of the difficulties can be obtained from enzyme properties and kinetics databases, such as BRENDA (http://www.brenda-enzymes.org) and SABIO-RK (http://sabio.villa-bosch.de) (Schomburg et al., 2014; Wittig et al., 2014). It is not uncommon to find that older values for activity were determined at ‘room temperature’ or in phosphate buffer, pH 7.2, with no indication of the buffer concentration or the counter ion used. Since enzyme activities and kinetic properties are dependent on the assay conditions (e.g., temperature, pH, ionic strength and other system components) under which they are determined, as well as on the nature of the system being studied, it is essential that these data are fully documented in any reports. Furthermore, the expression of enzyme activities in ill-defined or arbitrary units is not uncommon and it is relatively rare to find any meaningful statistical estimation of the errors of all reported enzyme parameters.

The STRENDA commission

The Standards for Reporting Enzyme Data (STRENDA) commission (http://www.beilstein-institut.de/en/projects/strenda) was set up in 2003, as a result of an initiative of the Beilstein-Institut (http://www.beilstein-institut.de; Kettner and Hicks, 2005; Apweiler et al., 2005), in order to address these problems. A series of meetings on ‘Experimental Standard Conditions of Enzyme Characterizations’ (ESCEC) has been held at which experts discussed possibilities for improvement of reporting enzyme data. Their conclusions emphasised the urgent need for recommendations for the standardisation of data reporting in this area, and that such standards should be independent of the organism being studied and intended application of the data. The task of the STRENDA commission was to investigate how this could be achieved. The present composition of the commission is listed on its website (http://www.beilstein-institut.de/en/projects/strenda), where the proceedings of the previous ESCEC meetings can also be found. Membership is open for additional scientists willing to help in the work and input from the scientific community is welcomed.

The objective of the STRENDA Commission is to provide a framework for ensuring that enzyme functional data are recorded with adequate detail of the assay conditions and reliability. This aim is not to tell people how to assay enzymes or what conditions they must use but simply to ensure that they provide sufficient information.

Basic information

It is relatively easy to think about what one might need to know from any paper reporting enzyme activities. Some of the obvious questions are listed below:

1. About the enzyme
   (a) What was the enzyme assayed?
   (b) What was the source? (species, tissue/organelle).
(c) Was it pure or a crude tissue extract?
(d) Had it been modified in any way? (e.g. fusion protein, His-tagged, proteolyzed, de-glycosylated etc.).

2. About the determination
(a) What was the reaction followed?
(b) What was the assay temperature?
(c) What buffer pH and concentration?
(d) What was the substrate concentration?
(e) What were the concentrations of any other substrates necessary?
(f) How much enzyme was used in each assay?
(g) Were any activators/inhibitors/stabilisers present?
(h) What assay procedure was used and were initial (linear) rates determined?

3. The results
(a) What was the enzyme activity obtained (with error estimates)?
(b) Was the activity proportional to the enzyme concentration?
(c) What are the units of the values reported?

Most of these are self-evident and should not require further explanation. It might not be thought of as asking too much of those reporting enzyme activities to provide such data, but it is quite common to find some of this essential information missing from publications. For example, the literature contains several examples of statements of the type ‘the enzyme was assayed by a modification of the method of xy et al.’ without detailing what the modifications were. The full composition and pH of the assay mixture is required. For identifying the enzyme studied, the EC number and accepted name, which can be found through the ExplorEnz website (http://www.enzyme-explorer.org), together with its source should be adequate but, since EC classification is functional system that is based on the reaction catalysed rather than the structure or location of the enzyme, it may also be necessary to identify a specific isoenzyme. Several alternative names, which are sometimes ambiguous or misleading, have been used for the same enzyme in many cases, but these may generally be related to the EC number and accepted name by searching ExplorEnz.

There is no recommendation as to which substrate(s) should be used for assays, but it is important that they are identified and their concentrations specified. Confusion can arise in, the names used for substrates, with different names being used for the same compound. IUPAC names (Panico et al., 1994) should be unambiguous but these are sometimes too cumbersome for routine use. The CAS (Chemical Abstracts Service) number of the compound should allow its identification through the free Common Chemistry utility (http://www.commonchemistry.org). Alternatively several databases provide alternative names that have been used for individual compounds together with their IUPAC names (http://pubchem.ncbi.nlm.nih.gov; http://www.chemspider.com; http://www.ebi.ac.uk/chebi/; http://www.genome.jp/kegg/). A common problem with compounds that exist in more than one isomeric form is the failure to indicate which form was used.

The question of whether the enzyme under study has been modified in any way is important since such modifications may affect its behaviour. It is common to find that proteolysed preparations are used, either by design or accident, with the assumption that if the enzyme preparation has activity, it must be satisfactory. However, there is a considerable amount of evidence that this may not be a valid assumption. Proteolytic cleavage can occur quite easily during extraction and purification of enzymes and this is, for example, known to affect the pH optimum of fructose bisphosphatase (EC 3.1.3.11) as well as the allosteric properties of that enzyme (Nimmo and Tipton, 1982) and of glutamate dehydrogenase [NAD(P)\(^{+}\)] (EC 1.4.1.3) (McCarthy and Tipton, 1985). Despite this, an increasing number of studies have been conducted with preparations that are truncated, fused with another protein, contain tags, such as poly-His, lack native glycosylation or are suspended in some unusual detergent without any investigation as to whether these have altered the behaviour of the enzyme.

The units in which enzyme activities are given should be specified, but their form has not been standardized. Activities are generally expressed as the amount product formed in unit time per amount enzyme protein present. This is often known as the International unit (IU) when 1 IU is the amount of enzyme that produces 1 \(\mu\)mol of product per min. The SI equivalent of the IU is the katal (mol/s) and this may alternatively be used as a unit of activity (conversion factors 1 IU = 16.67 nkat; 1 katal = 6 \times 10^7 IU). This is the recommended unit of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) and the International Union of Pure and Applied Chemistry (IUPAC) (Dybkaer, 2001). However, many biochemists find this an inconveniently small unit of activity and continue to use the IU (see also Bisswanger, 2014). It is also common to find enzyme activities expressed in non-standard units, such as the amount of enzyme catalysing a specified change in absorbance within a specific time (s, min or h). Since these are often referred to as units, there is scope for confusion with the IU.

The stoichiometry of the reaction assayed is also of importance in this context. For example, the enzyme carbamoyl-phosphate synthase (ammonia) (EC 6.3.4.16) catalyses the reaction

\[
2\text{ATP} + \text{NH}_3 + \text{CO}_2 + \text{H}_2\text{O} = 2\text{ADP} + \text{phosphate} + \text{carbamoyl phosphate}
\]

and so the activity expressed in terms of disappearance of ATP or formation of ADP would be twice that obtained if any of the other substrates or products were measured. Similarly, the NAD(P)\(^{+}\)-dependent nitric-oxide synthase (EC 1.14.13.165) catalyses

\[
2\text{L-arginine} + 3\text{NAD(P)H} + 3\text{H}^+ + 4\text{O}_2 = 2\text{L-citrulline} + 2\text{nitric oxide} + 3\text{NAD(P)\(^{+}\)} + 4\text{H}_2\text{O}
\]

Thus, it is important to specify reaction studied and the substrate or product measured when expressing activity of an enzyme.

Expression of enzyme activity in this way assumes that the initial velocity is proportional to the enzyme concentration. Although that is usually the case, there are cases where it is not (Dixon et al., 1979; McDonald and Tipton, 2002) and so it is always important to check. Similarly it is important to measure the initial rate of the reaction catalysed. With the increased use of high-throughput assays, in which the amount of product formed (or substrate used) is determined after some fixed time, it is, of course, necessary to ensure that the values obtained do, indeed, represent initial velocities.
In the field of clinical biochemistry it is necessary to have closely controlled conditions for assaying specific enzymes of diagnostic relevance so that values can be related between laboratories and to ‘normal’ ranges. The IFCC has produced “several primary reference procedures” for the assay of such enzymes (see, e.g. Schumann et al., 2011), which provide complete assay details. Other researchers have a greater freedom to select assay conditions that they find convenient. Several manufacturers produce test kits for specific enzymes, although it is not always easy to find their exact composition.

**More-specific information**

The list discussed above might be sufficient if one’s only interest was to compare enzyme activities between laboratories, individuals, species or tissues. However, additional information may be necessary for other types of work. The $K_m$ value(s) could be important to help one chose the assay substrate concentrations and the maximum velocity ($V$) might help in deciding how much enzyme to use. The complete kinetic parameters might be needed for systems biology or mechanistic studies. In that case it would also be of value to know how the kinetic parameters were determined and the error estimates associated with each value.

So far the list of requirements has avoided telling researchers what to do. For example, the method that was used determining the $K_m$ and $V$ (or $k_{cat}$) values is requested, together with the range of substrate concentrations used, without any guidance on whether there is any preferred procedure. Double-reciprocal (Lineweaver-Burk) plots continue to be widely used, despite this being well-known to be the least accurate of the procedures used (Dowd and Riggs, 1965), but it is judged better to have the information available than to censor any that may be less reliable. Error estimates (standard errors, standard deviations or confidence limits) should be provided for each value reported.

In the case of enzymes that show apparently cooperative kinetics, the substrate concentration that gives half-maximum velocity ($S_{0.5}$) and some measure of the cooperativity is also required. Hill coefficient ($h$ or $n_H$) is the most widely used of these, although the ‘saturation ratio’: ($R_s$), defined as

$$R_s = \frac{[S] \text{ at } 90\% V}{[S] \text{ at } 10\% V}$$

which will be 81 for a system following simple Michaelis-Menten kinetics and approximately $81^{1/h}$ for a cooperative system, is an acceptable alternative. Note that although the symbol $n$ continues to be often used for the Hill coefficient it invites confusion with the number of binding sites.

**Inhibitors**

Much research is now concentrated on enzyme inhibition, because of its great importance for drug development. This necessitates the provision of additional information, which will depend on the type of inhibition. For all types of inhibition it is important to show whether the inhibition is reversible by removal of excess inhibitor, for example by dilution or dialysis of the enzyme-inhibitor mixture, and whether the inhibition increases with the time that the enzyme is incubated with the inhibitor.

For simple reversible inhibitors, the substrate and inhibitor concentration ranges used in the study should be provided in addition to the $K_i$ values and types of inhibition observed. The concentrations of any other required substrates are necessary since the $K_i$ value will be dependent on these for most reaction mechanisms. It is also possible to find cases of partial inhibition where an excess of inhibitor does not completely prevent the reaction from occurring. These are, fortunately, quite rare and their treatment has been discussed in detail elsewhere (Dixon et al., 1979; Tipton, 1996). Similar considerations apply, of course, to data for activators, with the important difference that there may be some activity in the absence of activator.

Some inhibitors have such high affinities for the enzyme that the concentrations required for inhibition are comparable to those of the enzyme. Such tight-binding inhibitors, where the $K_i$ is similar to the enzyme concentration, pose specific problems, because the binding of the inhibitor to the enzyme will significantly reduce the free inhibitor concentration and so the assumption that the total inhibitor concentration is equal to the free inhibitor concentration, which is implicit in the usual treatments of reversible inhibition, is no longer valid. The rates of development of inhibition and recovery of activity after removal of the excess inhibitor may also be relatively slow. Specific graphical and computer-based procedures are available for determining the kinetic parameters and the type of inhibition (Williams and Morrison, 1979; Szedlacek and Duggleby, 1995).

In the case of irreversible inhibitors it is important to know whether inhibition is time-dependent, and if so how long enzyme and inhibitor were incubated together before the activity was determined. The different types of irreversible inhibition and their kinetic behaviour have been discussed in detail elsewhere (McDonald and Tipton, 2012).

In many instances $IC_{50}$ (or $I_{90}$) values are reported. These are simply defined as the amount of inhibitor that gives a 50% decrease in activity. For reversible inhibitors these have little meaning unless one knows the type of inhibition and the substrate concentrations. The relationships between $IC_{50}$, $K_i$ and $K_m$ values and substrate concentrations for the different types of inhibition have been reported (Dixon et al., 1979; McDonald and Tipton, 2002). For irreversible, time-dependent inhibitors the value will depend on the time for which the enzyme was pre-incubated with inhibitor before assay. In the presence of excess inhibitor one would expect the $IC_{50}$ to approach a value of half the enzyme concentration as the pre-incubation time is increased.

Such considerations mean that the use of $IC_{50}$ values should be discouraged, indeed, many authors have been discouraging their use for over half a century, but the fact remains that tables of such values continue to appear in the literature (especially in the pharmacological literature) posing the dilemma as to whether to include them.

**Forms and checklists**

Few people enjoy filling out forms. In fact some would prefer a visit to the dentist to having to do so. Nevertheless, it is important to collect the data in tabular form if they are to be made easily accessible and also to provide checklists for authors, and journals to ensure that the necessary data have been provided. A problem is that although it is relatively easy
to list what data one would like to have, it becomes more convoluted and quasi-legalistic when put on a form in terms of information fields to be completed. The nastier and more complicated the form, the more the resistance one might expect from the user. The design of such a data deposition form has been a major preoccupation of the STRENDA Commission and it has undergone many revisions before the current on-line form that is that is planned to be released in the first half of 2014. Currently, on the STRENDA website a prototype of the productive version is provided for further comments and suggestions for improvement (http://www.beilstein-institut.de/en/projects/strenda; Apweiler et al., 2010).

Over 30 international journals (listed on the STRENDA website) have, so far, encouraged adherence to the STRENDA guidelines and it is hoped those working in the field will see the advantage of following them in reporting their own data. It is not the function of the STRENDA Commission to force scientists to use the form before their data can be published, rather it is to be hoped that they will come to appreciate the value of doing so.

Dissemination

As well as collecting information, it is important to make it readily and freely accessible to everyone who may want to use it. That involves creating a database. This can be a real service because it is often difficult and time-consuming to find such data from the literature. Naturally, it is reasonable to ask whether we need yet another database. There are many databases that duplicate each other, with each claiming to offer some advantage over those already extant, although the only apparent advantage often appears to be that of allowing the publication of yet another database paper. Enzyme activity and kinetic data can be found elsewhere (http://www.brenda-enzymes.org; http://sabio.villa-bosch.de/), but the uniqueness of this approach is that it intends to provide the data together with the conditions under which they were determined to allow others to duplicate or apply it. Furthermore, the data should be in a form that can be freely used by other databases and incorporated into them in whole or in part.

Extensions

Biochemists may have different reasons for determining enzyme data. Industrial enzymologists may be particularly interested in behaviour at elevated temperatures, whereas ease of assay may be a prime concern of others. This may involve using non-physiological substrates, working under conditions far removed from those occurring within the cell or adding ‘unnatural’ components to the assay mixture. Systems biologists would like the data to be collected under standard conditions that approximate to those pertaining in the tissue, cell or organelle they wish to model. However, even a brief survey of the literature will indicate that this has been far from the case. Even with what is apparently the same enzyme, different laboratories often assay under different conditions and the assay conditions used for different enzymes in the same metabolic pathway can differ markedly. Some attempts have been made to formulate recommendations about assay conditions (Dixon et al., 1979), but these are somewhat imprecise and of little relevance to physiological conditions.

Originally many studies were conducted at ‘room temperature’, which could, of course, vary widely between laboratories. It was then recommended that enzymes should be assayed at 25 °C, which was, at that time, regarded as a standard ‘room temperature’. However, not all laboratories were able to meet this requirement and the standard assay temperature was raised to 30 °C. Even this gradual thermal inflation does not satisfy those studying human enzymes, who would regard a temperature of 37 °C as being closer to that in most tissues and conditions. However, this definition of physiological temperature for a mammalian system would not be appropriate, for example, to thermophilic bacteria or poikilotherms.

The recommended that the assay pH should “where practicable, be optimal” (Dixon et al., 1979). Is also not very helpful, since the optimum pH may depend on the choice of substrate, the substrate concentrations, buffer, temperature and ionic strength and there are no strict recommendations for any of these. Furthermore the optimum pH may be far removed from the pH at which an enzyme is perceived to operate in vivo. For example, the optimum pH for arginase (EC 3.5.3.1) is reported to be about pH 10 in horse, pH 9.8 in rat and pH 11 in Bacillus brevis. Those working with mammalian systems might favour an assay pH of about 7.2 which is believed to be around the physiological pH within the cell, but clearly this would be unphysiological for gastrointestinal enzymes, such as pepsin and trypsin, or for lysosomal enzymes. Furthermore, the oxidation of ethanol by liver alcohol dehydrogenase (EC 1.1.1.1) is often followed at higher pH values because the equilibrium of the reaction greatly favours ethanol formation at neutral pH.

Naturally it would be appropriate to use physiological substrates for enzyme assays. However, many studies have used unphysiological substrates for ease of manipulation and assay. For example acetylthiocholine is frequently used to assay acetylcholinesterase (EC 3.1.1.7) because the thiocholine produced can be readily detected by reaction with sulphydryl reagent 5, 5’dithiobis-2-nitrobenzoate (Nbs2) releasing a yellow coloured compound whose formation can be followed spectrophotometrically at 412 nm (Ellman et al., 1961). Other examples include the use of 4-nitrophenyl phosphate to assay alkaline phosphatase (EC 3.1.3.1) (Schumann et al., 2011). The use of synthetic dyes as electron acceptors in oxidoreductase assays has been common and in some cases the physiological acceptor remains unknown. The demand for higher assay sensitivity and high-throughput procedures has resulted in the development of an increasing number of chromogenic and fluorogenic substrates (Goddard and Reymond, 2004; Reymond et al., 2009). Clearly, in such cases a considerable amount of work would be necessary to show whether the enzyme behaves identically towards such substrates as it does towards its physiological substrates.

It is often recommended that saturating substrate concentrations should be used (i.e. >10 Km, for all substrates), as discussed by Bisswanger (2014). This, of course, assumes that the Km values have already been determined, at least approximately. Furthermore, this might not always be practicable because of factors such as solubility, the occurrence of
high-substrate inhibition or a high absorbance of the assay mixture affecting the behaviour of optical assays (Dixon et al., 1979; McDonald and Tipton, 2002). It should also be remembered that any change in the assay conditions (e.g., pH, temperature, ionic strength) may affect the $K_m$ values.

The buffers and ionic strengths and used in enzyme assays vary widely and are often far from physiological. It might be unhappy in one or other of the common buffers (see e.g., Boyce et al., 2004). Furthermore, buffers that contain physiologically occurring compounds can be problematical in that they are likely to be substrates or inhibitors of for some other enzymes. Several enzymes are sensitive to inhibition by high ionic strengths and altering the concentrations of charged substrates if the full composition and pH of the assay mixture is given, it can be calculated under "in vivo-like" conditions (van Eunen et al., 2010; Goel et al., 2012). However, from the above examples, it should be clear that it is unlikely that a universal buffer medium, suitable for all enzymes in all tissues and organelles, will be found. Indeed different conditions should apply to the same enzymes from different sources. Individual standards will be required for each organism, organ and organelle to be studied, bearing in mind that these may not be constant under all metabolic conditions. Perhaps the answer will lie in more complex mixtures, including proteins as buffers, that more closely mimic the, crowded, in vivo environments of groups of enzymes. In its attempts at formulating more physiologically relevant assay conditions the STRENDA Commission needs advice from those working with specific systems.

Conflict of interest statement

None of the authors have any conflict of interest.

References


Web references


Sabio-RK: Biochemical Reaction Kinetics Database. (http://sabio.villa-bosch.de/).
Beilstein Institut. (http://www.beilstein-institut.de/).
ExplorEnz: The Enzyme Database. (http://www.enzyme-explorer.org/).
CAS Common Chemistry Utility. (http://www.commonchemistry.org/).
ChemSpider: The Free Chemical Structure Database. (http://www.chemspider.com/).
ChEBI: Chemical Entities of Biological Interest. (http://www.ebi.ac.uk/chebi/).