COMMENTARY

Hazards of Deducing Enzyme Structure–Activity Relationships on the Basis of Chemical Applications of Molecular Biology

No subject in biology has attracted more interest from all branches of chemistry than the molecular basis for enzyme action, where nature demonstrates a diversity of applications of chemical principles to catalysis. The field has a long history and has drawn heavily on the technologies and intellectual frameworks of analytical, inorganic, organic, and physical chemistry. Most recently, research on enzyme catalysis has introduced molecular biology and recombinant DNA technology to isolate, produce, and determine the primary structures of enzymes and to manipulate structure and activity. The integration of new technologies from the biological sciences realized long-held wishes of those in the field: the facile substitution of amino acids at specific locations in an enzyme so that structure–function relationships and theories of catalysis can be tested directly, and so that new specificities and catalytic efficiencies can be designed and constructed.

However, the enthusiasm for these experiments should be tempered with the recognition that the introduction of powerful technologies from the biological sciences also brings a minefield of hazards. I give here one example of a complication which I suspect some people who do structure–activity experiments are not aware of, and of how to deal with this complication in a straightforward way which has not generally been used. This example deals with the interpretation of the low activity of an enzyme variant which has a specific amino acid substitution in its active site.

Assume that the gene for an enzyme has been cloned, that a specific substitution has been introduced by site-directed mutagenesis, and that the sequence of the altered gene has been confirmed. The resulting protein has a low activity which can be determined quantitatively. From this observation calculations are done to estimate the contribution of a specific group to the lowering of the free energy of activation and to the catalytic rate constant k_{cat}. On the surface there is nothing ambiguous about interpreting a low activity that is the consequence of a specific amino acid replacement. However, the design and execution of one more experiment might show that the mutant protein is completely inactive and that the explanation for the low activity lies elsewhere.

A specific example is provided by experiments on the enzyme β-lactamase, which has been extensively investigated.[1-3] The active site serine 68 (Ser68) reacts with the carbonyl carbon of the β-lactam ring and thereby promotes antibiotic ring cleavage. The Ser68 hydroxyl group nucleophile can be replaced by a sulfhydryl to give an enzyme with a k_{cat} which is 2% of that of the wild-type species.[3] The significant activity of the Ser68 → cysteine (Cys) mutant is consistent with the requirement for a nucleophile that adds to the carbonyl carbon of the β-lactam ring.

It is possible to measure accurately activities that are even smaller than 2% of that of the wild-type enzyme. Substitution of the Ser68 codon with a codon for glycine (Gly) yields a β-lactamase which has a low but reproducible activity that is about 0.1% of that of the wild-type enzyme or about 5% of that of the Cys68 enzyme.[4] This activity is well above the background observed in extracts of a strain which lacks the plasmid-encoded mutant gene. If the reduced activity can be ascribed entirely to a lower k_{cat}, then a 1000-fold reduction corresponds to an increase in the free energy of activation of approximately 4 kcal mol^{-1}. The activity could be interpreted as a consequence of an enzyme-induced strain on the β-lactam ring which accompanies binding of the antibiotic to the Gly68 mutant protein, with water (or a hydroxide ion) replacing the Ser68 hydroxyl as the attacking nucleophile. This interpretation invites analytical calculations of the degree of enzyme-induced strain of the β-lactam ring, the effective pK of water in the vicinity of the active site, and so on. However, even though the experiment is conceptually straightforward and was properly executed, and even though there is an opportunity to compute induced strain and other factors in catalysis, it is wrong to attribute the activity to the Gly68 mutant enzyme itself.

The underlying assumption in all experiments of the sort described above is that knowledge of the nucleotide sequence of the gene for a protein is tantamount to knowledge of the amino acid sequence. There is ample evidence to support this assumption. However, there is also much evidence that the biochemical machinery which converts DNA sequences into proteins makes occasional mistakes.[5-8] These can be errors in transcription and in translation, although the latter are probably more common. These mistakes can lead to microscopic heterogeneity of the proteins which are produced from a fixed DNA sequence, with the protein that is specified by the gene mixed with minute amounts of sequence variants. A variant produced by an error of translation, for example, can have an intrinsic activity that is different from that of the predominant species. This is the likely explanation for the putative activity of a β-lactamase Gly68 mutant which turns out to have no measurable activity whatsoever.

The definitive experiment to clarify this issue is to separate active from inactive chains and to determine directly the amino acid at position 68 in the active species. However, separation of a scarce enzyme species, which differs from the major species by a single amino acid, can be difficult. In a recent study, isoelectric focusing separated species of triosephosphate isomerase which differed only by replacement of a single charged amino acid for one that is neutral.[9] Neither protein was scarce in this case, although the method might still be used to detect and

(1) Fisher, J.; Balasco, J. G.; Khosla, S.; Knowles, J. R. *Biochemistry* **1980**, *19*, 2895–2901.

(2) Dalbadie-McFarland, G.; Cohen, L. W.; Riggs, A. D.; Morin, C.; Richards, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 6409–6413.

(3) Sigal, I. S.; DeGrado, W. F.; Thomas, B. J.; Petteway, S. R., Jr. *J. Biol. Chem.* **1984**, *259*, 5327–5332.

(4) Toth, M. J.; Murgola, E. J.; Schimmel, P. *J. Mol. Biol.* **1988**, *201*, 451–454.

(5) Edelmann, P.; Gallant, J. *Cell* **1977**, *10*, 131–137.

(6) Yarus, M. In *Transfer RNA: Structure, Properties and Recognition*; Schimmel, P., Söll, D., Abelson, J., Eds.; Cold Spring Harbor Laboratory: New York, 1979; pp 501–515.

(7) Parker, J.; Friesen, J. D. *MGG, Mol. Gen. Genet.* **1980**, *177*, 439–445.

(8) Bouadaloun, F.; Donner, D.; Kurland, C. G. *EMBO J.* **1983**, *2*, 1351–1356.

(9) Nickbarg, E. B.; Davenport, R. C.; Petsko, G. A.; Knowles, J. R. *Biochemistry* **1988**, *27*, 5948–5960.

0001-4842/89/0122-0232$01.50/0 © 1989 American Chemical Society

study a minor species when it has a charge difference. Additionally or alternatively, active-site titrations and accurate measurements of *pre-steady-state*, as opposed to *steady-state*, kinetic parameters can rule out artifacts from contaminating species.[10]

Without isolation of a minor species or execution of careful kinetic measurements, however, there is a straightforward test for at least one type of error which can arise from infrequent mistranslation of a specific codon in the messenger RNA. Replacement of the codon in question with another that codes for the same amino acid should give the same result, unless there is a codon-specific error in translation. I would suggest that this replacement be done routinely as one of the procedures used to analyze and interpret a low activity.

Because of the degeneracy of the genetic code, most amino acids have more than one codon. The glycine codons are GCX, where X is A, G, C, or U. When the Ser68 AGC codon is changed to GGC, a low β-lactamase activity is observed, as described above. When AGC is replaced instead by GGA, however, the mutant β-lactamase is made in the same amount, but it has no detectable activity. Thus, the Gly68 mutant protein per se has no inherent activity. Instead, activity is correlated with the presence in the gene of a specific codon for glycine.

The probable explanation is that serine is occasionally inserted instead of glycine at the GGC codon (at position 68) of the messenger RNA that is transcribed from the mutant gene. When a bacterial strain is used which increases the frequency of misreading of mRNA codons, the activity associated with the GGC mutant is enhanced even further.[4]. Mistranslation of the GGC codon could occur through a unique mismatch of a specific serine transfer RNA (tRNA) at the first position of the codon–anticodon interaction. At least four serine tRNAs are required to decode the six serine codons, and one of the serine tRNAs has a GCU anticodon. the antiparallel pairing of the GGC codon with the GCU anticodon results in two Watson–Crick base pairs and a first-position G-U mismatch. (The G–U wobble interaction is common at the third position of the codon–anticodon interaction.) There is no serine tRNA which can read the GGA codon by a "two base reading" mechanism. This could explain why the GGA codon at position 68 specifies a completely inactive mutant enzyme.

The reading of a GGC glycine codon at position 68 by a specific serine tRNA may require the context of the particular codons which flank that region of the messenger RNA, in addition to other factors. Likewise, with other systems, there may be possibilities for misinsertions that are specific to the sequence of the particular messenger RNA. There also are other mechanisms for generating microscopic sequence variants, including the occasional misacylation of tRNAs with incorrect amino acids and subsequent escape of the misacylated tRNA species from the editing system.[6,11,12] No assumptions should be made about the accuracy of translation, because our understanding is limited. Overall error rates of 10^{-3} and higher have been measured in different systems.[5-8]

Those who use tools of molecular biology for structure–activity manipulations need to recognize that their ease of use can obscure a minefield of hazards. With the use of its tools, there is an obligation to learn more about the discipline itself. And there is the further obligation to acknowledge that the uncharacterized variables inherent to a biological technical system can ambush even the best informed. While detailed kinetic measurements and rigorous enzyme purifications can uncover or rule out spurious activities, a simple test for translational errors by replacement of the critical codon with a synonym should first be done in situations where much depends on the interpretation of a low activity.

(10) Leatherbarrow, R. J.; Fresht, A. R.; Winter, G. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 7840–7844.

(11) Schimmel, P. R.; Söll, D. *Ann. Rev. Biochem.* **1979**, *48*, 601–648.

(12) Fersht, A. R. *Enzyme Structure and Mechanism*, 2nd ed.; W. H. Freeman: New York, 1985; pp 347–368.

Paul Schimmel

Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139