ments are aligned in all possible ways. For each alignment, a score is obtained by counting +1 for a match, $-\mu$ for a mismatch, and $-\delta$ for a letter inserted or deleted (a gap).

For example, AGCACT and AGGT can be aligned as

AGCACT
AG–G–T

to receive score $S = 3 - \mu - 2\delta$. They can also be aligned as

AGCACT
AGGT– –

to receive score $S = 2 - 2\mu - 2\delta$. The score $M_{m,n}(x,y)$, which is defined to be the maximum of all the scores obtained for all possible pairs of segments I and J, is then calculated by an algorithm whose computing time is proportional to the product of m and n.

To calculate the probability of those large values of $M_{m,n}(x,y)$ for which the similarity is significant, one has to know, at least approximately, the distribution of $M_{m,n}(x,y)$ under the assumption that x and y are unrelated. That is, the letters $x_1,\ldots, x_m$, $y_1,\ldots, y_n$ are independently chosen with the same distribution from the alphabet {A,C,G,T}. Karlin and Altschul (*7*) obtained approximations for the probabilities of large values of $M_{m,n}(x,y)$ for the case $\delta = \infty$ (that is, without gaps), assuming the expected score of two letters to be negative. Arratia, Gordon, and Waterman (*8*) considered the score $M_{m,n}(t)$, which is the maximum of the scores obtained by considering only those pairs of segments I and J of a given length t, for the case $\mu = 0$. They established approximations for $M_{m,n}(t)$ under certain mild conditions by the method of Poisson approximation.

The set of all values of the parameters (μ,δ) can be divided into two regions, S_1 and S_2, such that for $m = n$, the growth of $M_{n,n}(x,y)$ is proportional to n in S_1 and the growth of $M_{n,n}(x,y)$ is proportional to $\log n$ in S_2. The cases considered by Karlin and Altschul (*7*) and Arratia, Gordon, and Waterman (*8*) are in the logarithmic region. The work of the latter (*8*) has provided a basis for Waterman and Vingron (*9*) to use the Poisson clumping heuristic of Aldous (*10*) to calculate the probabilities of large values of $M_{m,n}(x,y)$ in the entire logarithmic region.

Let us see how the method of Poisson approximation is applied in the problem of Arratia, Gordon, and Waterman (*8*). Because $\mu = 0$, the score for each pair of the segments I and J of a given length t is just the number of matches. Let s be a given positive integer. Associate an event with each pair of I and J. If the score is at least s for a particular pair of I and J, we say that the associated event occurs. The number of events that occur, say, U is the number of those scores which are at least s. Therefore, $P[M_{m,n}(t) < s] = P(U = 0)$.

We would have been done if the distribution of U was approximately Poisson with mean, say, λ^*. For then, we would have had $P[M_{m,n}(t) \geq s] = 1 - P(U = 0) \simeq 1 - e^{-\lambda^*}$. However, this is not the case. The events associated with the I's and the J's occur in clumps. By the Poisson clumping heuristic of Aldous (*10*), it is the number of clumps that is expected to have approximately the Poisson distribution. So we declump and modify the events so as to obtain events which are associated with the clumps. Let W denote the number of clumps that occur. The method of Poisson approximation is then applied with the local approach. The result is that $P[M_{m,n}(t) \geq s] \simeq 1 - P(W = 0) \simeq 1 - e^{-\lambda}$, where λ is the mean of the approximating Poisson distribution.

There are many situations in which occurrences of events happen in clumps. The book by Aldous (*10*) provides many such examples. In these situations, the appropriate approximating distribution is the compound Poisson distribution. One of the new developments relating to the method of Poisson approximation is the extension of the method to compound Poisson approximation by Barbour, Chen, and Loh (*11*). This work extended the associated difference equation to an integral equation. Although Arratia, Goldstein, and Gordon (*5*) also considered compound Poisson approximation, the approach of Barbour, Chen, and Loh (*11*) is different and holds promise for producing better results. Much work is also being done on multivariate or process approximation, which was initiated by Barbour (*12*) and Arratia, Goldstein, and Gordon (*4*) using different approaches. Finally, for approximation for relative errors, which is very useful when the probabilities are small, a new approach was introduced in Chen and Choi (*13*).

References and Notes

1. C. M. Stein, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics, and Probability* (Univ. of California Press, Berkeley, 1972), vol. 2, pp. 583–602.
2. L. H. Y. Chen, *Ann. Probab.* **3**, 534 (1975).
3. A. D. Barbour, L. Holst, S. Janson, *Poisson Approximation* (Clarendon, Oxford, 1992).
4. R. Arratia, L. Goldstein, L. Gordon, *Ann. Probab.* **17**, 9 (1989).
5. ———, *Stat. Sci.* **3**, 403 (1990).
6. T. F. Smith and M. S. Waterman, *J. Mol. Biol.* **147**, 195 (1981).
7. S. Karlin and S. F. Altschul, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2264 (1990).
8. R. Arratia, L. Gordon, M. S. Waterman, *Ann. Stat.* **18**, 539 (1990).
9. M. S. Waterman and M. Vingron, in preparation.
10. D. Aldous, *Probability Approximations via the Poisson Clumping Heuristic*, vol. 77 of *Applied Mathematical Sciences* (Springer, New York, 1989)
11. A D. Barbour, L. H. Y. Chen, W.-L. Loh, *Ann. Probab.* **20**, 1843 (1992).
12. A. D. Barbour, *J. Appl. Probab.* **25(a)**, 175 (1988).
13. L. H. Y. Chen and K. P. Choi, *Ann. Probab.* **20**, 1867 (1992).

14. I thank M. Waterman for helpful discussions on the molecular biology application and Z. Chen, C. T. Chong, and Y. K. Leong for their helpful comments on the preliminary drafts of this Perspective.

Conformational Flexibility of Enzyme Active Sites

Chen-Lu Tsou

The activity of enzymes is strongly dependent on their conformational integrity. Our laboratory has been interested in the precise relationship between enzyme activity changes and protein unfolding. The observation that, under denaturing conditions, loss of enzyme activity can precede marked changes in protein conformation led us to hypothesize that enzyme active sites may display more conformational flexibility than the enzyme molecules as a whole (*1*, *2*). Here I discuss recent results that support this concept.

The author is at the National Laboratory of Biomacromolecules, Institute of Biophysics, Academia Sinica, 15 Datun Road, Beijing, 100101, China.

Exposure of the enzyme creatine kinase to denaturants such as guanidine hydrochloride (GuHCl) and urea results in an initial phase of rapid inactivation; this inactivation can be conveniently measured by following the substrate reaction with a stopped-flow apparatus (*3*). In parallel, conformational changes induced by the denaturants can be monitored by conventional methods that detect changes in intrinsic fluorescence, absorbance in the ultraviolet, circular dichroism, or exposure of buried thiol groups. Comparison of conformation and activity changes of creatine kinase during denaturation indicates that enzyme inactivation occurs at a much lower concentration of denaturant than is required to

bring about unfolding of the enzyme molecule. At a given GuHCl concentration, the inactivation rate is over three orders of magnitude higher than the rate of conformational change (see table) (*1*). Similar results have been obtained for D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ribonuclease A (RNase A), and several other enzymes (*2*).

Comparison of rate constants for enzyme unfolding and inactivation

Enzyme	Temperature (°C)	Unfolding		Inactivation	
		GuHCl (M)	Rate (s^{-1})	Rate (s^{-1})	Residual activity (%)
Creatine kinase	25	0.5	0.0043	3.6	15
	25	1	0.053	4.3	0
GAPDH	25	0.5	0.0018	>50	0
RNase A	10	1	1	>50	27
	10	3	0.01		3
Papain	25	5	0.0078	0.58	0

For pancreatic RNase A, enzyme inactivation occurs at <2 M GuHCl, whereas conformational changes are first detectable at >3 M GuHCl (*4*). In contrast to the rate of conformational change, which is in the range of several seconds, the rate of inactivation is too fast to be measured even by following the substrate reaction in a stopped-flow apparatus (see table). To exclude the possibility that the rapid decrease in enzyme activity is due to inhibition by the chemical denaturants rather than to a conformational change (*5*), we have also studied the changes that accompany thermal denaturation. For tetrameric GAPDH, thermal inactivation occurs at a lower temperature than does the conformational change, and at the same temperature, the inactivation rate is markedly higher than both the rates of conformational change and dissociation of the tetrameric enzyme (*6*). Similar results have also been obtained for adenylate kinase (*7*). Since no chemical inhibitors are present during thermal denaturation, inactivation must be due to perturbation of enzyme conformation, presumably at the active site.

For GAPDH and creatine kinase, there is now direct evidence that conformational changes at the active site are involved in the initial phase of enzyme unfolding. We introduced a fluorescent probe into the active site of creatine kinase and compared the changes in its intrinsic fluorescence with changes in the emission intensity and anisotropy of the probe in increasing concentrations of GuHCl (*8*). Decreases in emission intensity and anisotropy of the probe occur at a much lower concentration of GuHCl than is required to induce changes in the intrinsic fluorescence. The extent of enzyme inactivation coincides precisely with the decrease in emission intensity and anisotropy of the fluorescent probe, indicating partial exposure and increased mobility of the probe at the enzyme active site. Similar results have been obtained in studies of GAPDH labeled with a fluorescent NAD^+ (nicotinamide adenine dinucleotide) derivative at the active site (*9*). Kinetic measurements likewise indicate that both enzyme inactivation and exposure of the active site fluorophore occur at much higher rates than the unfolding of the enzyme molecule as a whole.

Although several laboratories have also observed that denaturant-induced enzyme inactivation occurs before measurable protein conformational changes (*10*), other workers have reported that the changes occur simultaneously (*11*). One possible explanation for this discrepancy relates to the fact that inactivation is, by necessity, monitored under saturating concentrations of substrate. Conceivably, under these conditions the substrate may protect or partially reactivate the enzyme during activity determinations, which would lead to an underestimate of the extent or rate of inactivation (*12*). In the case of papain (*13*), the apparent rate constants for inactivation and for conformational changes during GuHCl denaturation are not substantially different; however, the microscopic rate constants for the inactivation of the free enzyme are two orders of magnitude greater than those representing the conformational changes (see table).

In sum, these observations suggest that enzyme active sites are formed by relatively weak molecular interactions and hence may be conformationally more flexible than the intact enzymes. Other recent data are consistent with this hypothesis. The enzyme lactate dehydrogenase is stabilized in ammonium sulfate solutions or by cross-linking with glutaraldehyde; such treatment results in about a 33% loss of enzyme activity, possibly because these agents introduce rigidity into the active site. The stabilized enzyme regains its original activity after treatment with low concentrations (0.2 to 0.5 M) of GuHCl. Such treatment may reactivate the enzyme by restoring flexibility at the active site (*14*). Some enzymes, such as glucan 1,4-α-maltotetraohydrolase and dihydrofolate reductase, are activated by low concentrations of GuHCl. At low temperatures the extents of activation are greater but higher concentrations of GuHCl are required for maximal activation. The loss of enzyme activity at low temperature could be due, at least in part, to a rigid conformation at the active site. GuHCl may activate the enzyme by restoring active site flexibility.

The "induced-fit" model of enzyme action (*15*) postulates that enzymes exist in many different conformational states at equilibrium with one another, and that this equilibrium can be perturbed by the presence of substrates or other ligands. Full expression of catalytic activity may require rapid shifts in conformational states at the enzyme active site; if so, this would help to explain the requirement for enhanced flexibility in this region of the protein.

References

1. Q. Z. Yao, M. Tian, C. L. Tsou, *Biochemistry* **23**, 2740 (1984).
2. C. L. Tsou, *Trends Biochem. Sci.* **111**, 427 (1986).
3. ———, *Adv. Enzymol.* **61**, 388 (1988).
4. W. Liu and C. L. Tsou, *Biochim. Biophys. Acta* **916**, 455 (1987).
5. T. E. Creighton, *Biochem. J.* **270**, 1 (1990).
6. Y.-Z. Lin *et al.*, *Biochim. Biophys. Acta* **1038**, 247 (1990).
7. Y.-L. Zhang, J.-M. Zhou, C. L. Tsou, *ibid.* **1164**, 61 (1993).
8. H. M. Zhou, X. H. Zhang, Y. Yin, C. L. Tsou, *Biochem. J.* **291**, 103 (1993).
9. Y. S. Ho and C. L. Tsou, *Nature* **277**, 245 (1979).
10. J. F. Miller and D. W. Bolen, *Biochem. Biophys. Res. Comm.* **81**, 610 (1978); S. M. West and N. C. Price, *Biochem. J.* **265**, 45 (1990); R. A. Edwards, A. L. Jacobson, R. E. Huber, *Biochemistry* **29**, 11001 (1990); M. Nozais, J. J. Bechet, M. Houadjeto, *ibid.* **31**, 1210 (1992).
11. M. Engelhard, R. Rudolph, R. Jaenicke, *Eur. J. Biochem.* **67**, 447 (1976); H.-J. Lee, Y.-H. Chen, G.-G. Chang, *Biochim. Biophys. Acta* **955**, 119 (1988); C. D. Lindsay and R. H. Pain, *Eur. J. Biochem.* **192**, 133 (1990).
12. C. L. Tsou, *Chin. Sci. Bull.* **34**, 393 (1989).
13. J. Xiao, S. J. Liang, C. L. Tsou, *Biochim. Biophys. Acta* **1164**, 54 (1993).
14. Y. Z. Ma and C. L. Tsou, *Biochem. J.* **274**, 207 (1991).
15. D. E. Koshland Jr., *FEBS Lett.* **62**, E47 (1976).