contain unique, but unpredictable, regularities[5]. Are sections 114–163 and 269–318 therefore the parts of the sequence which Engel's theory would predict to be homologous? There is no

Table 1 Frequencies of Different Numbers of Identical Positions

No. of identical matches	Observed frequencies	Random sequences	
	Bovine GDH	Rearranged	Drawn from pool
0	5,155	5,270±321	4,721±406
1	15,828	15,987±360	14,881±952
2	23,604	23,983±359	23,077±277
3	23,307	23,284±568	23,516±482
4	17,026	16,755±329	17,433±727
5	9,878	9,402±235*	10,149±836
6	4,504	4,357±231	4,829±553
7	1,508	1,712±146	1,918±208
8	464	546±69	663±102
9	164	135±25	123±59
10	34	32±23	56±29
11	2	5.7±7.8	11.3±8.2
12	1	1.4±2.9	2.1±3.6
13	—	0.1±0.3	1.6±4.8
14	—	—	0.1±0.3

* Observed frequency more than two standard deviations from mean of frequencies for random sequences.

current evidence that section 269–318 constitutes the regulatory site and for the active centre the only definite fact is that lysine 126 is essential for enzyme activity. Since there are fifty sequences of length fifty residues which include residue 126, Engel's calculated probability for the random occurrence of twelve identical positions should be $50 \times 1.5 \times 10^{-3}$ ($= 7.5 \times 10^{-2}$) and on this basis the observation quoted is not significant. It is not however to be seriously supposed that either the active centre or the regulatory site would correspond to fifty contiguous residues.

Although no valid evidence has so far been produced in its favour it is possible that Engel's proposal as to the origin of the regulatory site of enzymes is correct and the evidence from X-ray studies may prove important in this problem.

J. WILLIAMS
A. G. WILKINS

Molecular Enzymology Laboratory,
Department of Biochemistry,
University of Bristol,
Bristol BS8 1TD

Received April 18; revised October 29, 1973.

1 Moon, K., Piszkiewicz, D., and Smith, E. L., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 1380 (1972).
2 Engel, P. C., *Nature*, **241**, 118 (1973).
3 Haber, J. E., and Koshland, D. E., *J. molec. Biol.*, **50**, 617 (1970).
4 Dickerson, R. E., *J. molec. Biol.*, **57**, 1 (1971).
5 McLachlan, A. D., *J. molec. Biol.*, **64**, 417 (1972).

Dr ENGEL replies:

Williams and Wilkins have done the evidence less than justice in two respects: (1) Our disagreement hinges upon the choice of a basis for statistical assessment of the observed sequence match. Absolute validity can neither be attained nor defined; one can only attempt to select reasonable criteria. My original probability calculation may have been too arbitrarily exclusive but, equally, Williams's and Wilkins's computer-search is too arbitrarily all-inclusive. To ignore relevant information must bias the calculation against discovery of the truth. The evidence for the importance of lysine-126 comes not only from studies of kinetics and binding (see ref. 1), but also from sequence comparisons which show that lysine-126 is in a conserved region[2–4]. This being so, even fifty may be an overestimate of the necessary number of sequences containing lysine-126 to be compared.

(2) Even if one concedes a probability as high as 7.5% for random occurrence of the observed degree of matching for a sequence including lysine-126, this takes no account of the additional evidence from the comparison with GPDH[5]. Here there can be no question of screening the whole molecule: the alignment is already fixed on the basis of the observed homology of sequences surrounding the lysine residues in GDH and GPDH that react with PLP[6,7]. The two GDH sequences are, of course, aligned on the basis of the internal homology. The probability of the observed matching of nine positions out of sixty-four in GDH 2 and GPDH may be calculated as follows.

The match[6,7] between GDH 1 and GPDH is assumed to reflect a genuine evolutionary relationship (six identities out of sixty-four). The match[5] between GDH 1 and GDH 2 (thirteen identities out of sixty-four) is taken, for the purposes of calculation, to be a fortuitous occurrence judiciously selected. The probability, higher than for twelve out of fifty, is 0.21, based on 64×310 comparisons. (For such large numbers of comparisons the approximation of multiplying the probability calculated for a single comparison by the number of comparisons breaks down and the second term in the binomial expansion must be included.)

Thirteen random matches with the sixty-four residues of GDH 1 could coincide with 0, 1, 2, 3, 4, 5 or 6 of the matching positions in the GDH 1/GPDH comparison. If no positions match in all three sequences, this defines $13+6=19$ positions that differ in GDH 2 and GPDH. Thus nine matches would have to be found among the remaining forty-five pairs of residues. Similarly, one congruence in all three sequences defines eighteen positions that differ in GDH 2 and GPDH, so that eight identities must be found in forty-six comparisons, and so on. The probability of finding n positions identical in all three sequences is given by

$$P_1 = \frac{^{58}C_{(13-n)} \times {}^{6}C_n}{^{64}C_{13}}$$

The probability of finding $9\text{-}n$ further positions matching in GDH 2 and GPDH is given by

$$P_2 = {}^{(45+n)}C_{(9-n)} \times \frac{19^{(36+2n)}}{20^{(45+n)}}$$

The overall probability of the total of nine matches that is in fact found is

$$\sum_{n=0}^{6} P_1 P_2 = 4.2 \times 10^{-3}$$

The cumulative probability, therefore, of the two observed matches with GDH 2 is $4.2 \times 10^{-3} \times 2.1 \times 10^{-1} = 8.8 \times 10^{-4}$.

To say the least this signifies a noteworthy coincidence. The availability of more sequence data should soon provide further evidence for or against the hypothesis of partial gene duplication in GDH.

P. C. ENGEL

Department of Biochemistry,
University of Sheffield, S10 2TN

Received May 7; revised November 16, 1973.

1 Chen, S. S., and Engel, P. C., *Eur. J. Biochem.* (in the press).
2 Moon, K., Piszkiewicz, D., and Smith, E. L., *Proc. natn. Acad. Sci. U.S.A.*, **69**, 1380 (1972).
3 Wootton, J. C., Chambers, G. K., Taylor, J. G., and Fincham, J. R. S., *Nature*, **241**, 42 (1973).
4 Holbrook, J. J., Roberts, P. A., and Wallis, R. B., *Biochem. J.*, **133**, 165 (1973).
5 Engel, P. C., *Nature*, **241**, 118 (1973).
6 Smith, E. L., Landon, M., Piszkiewicz, D., Brattin, W. J., Langley, T. J., and Melamed, M. D., *Proc. natn. Acad. Sci. U.S.A.*, **67**, 724 (1970).
7 Forcina, B. G., Ferri, G., Zapponi, M. C., and Ronchi, S., *Eur. J. Biochem.*, **20**, 535 (1971).