Eur. J. Biochem. *151*, 333–335 (1985)

The amino acid sequences of the copper/zinc superoxide dismutases from swordfish and *Photobacter leiognathi* confirm the predictions made from the compositions

Athel CORNISH-BOWDEN

Department of Biochemistry, University of Birmingham

(Received March 22, 1985) – EJB 85 0299

Recent suggestions that the amino acid sequence of the copper/zinc superoxide dismutases of swordfish and *Photobacter leiognathi* do not support the theory that the bacterium obtained the gene for the enzyme by transfer from its eucaryotic symbiont [Rocha, H. A., Bannister, W. H. and Bannister, J. V. (1984) *Eur. J. Biochem. 145*, 477–484] are examined. The amount of difference between the sequences is in good agreement with expectation from the amino acid compositions. Moreover, the gene-transfer hypothesis cannot be discarded without postulating an enormous increase in the rate at which the superoxide dismutase gene has accumulated amino acid substitutions since the divergence of the swordfish and cattle lineages.

Amino acid compositions provide a useful guide to protein relatedness in cases where the sequences are not known, provided that the data are interpreted with care and in relation to a model [1–4], and the investigation of the copper/zinc superoxide dismutase from *Photobacter leiognathi* by Martin and Fridovich [5] showed that they could lead to a discovery of major biological importance. The composition and other properties of the copper/zinc superoxide dismutase from this organism, a luminous bacterium found in the light organ of the ponyfish (*Leiognathus splendens*), indicated that it is remarkably similar to the corresponding copper/zinc superoxide dismutase of the host fish. The similarity was too great to be easily explained except by supposing that there had been a gene transfer from fish to bacterium during the estimated 3×10^7 years of co-evolution of the two species.

The recent determination of the amino acid sequences of the copper/zinc superoxide dismutases from *P. leiognathi* [6] and swordfish [7] has led to claims that the evidence for gene transfer is weaker than the compositions had suggested [6–8]. The two sequences differ at about 72% of loci, a rather greater degree of difference than had been expected. The sequence data for these two proteins are, however, in excellent agreement with the expectation from their compositions, well within the expected statistical uncertainty.

Amino acid compositions are most conveniently compared by means of the index $S\Delta n$ [1], because this index, unlike others that have been used, provides a direct estimate of the amount of sequence differences, with a predicted [1] and measured [2] coefficient of variation of about 38%. It is defined for proteins of equal length by the following equation:

$$S\Delta n = {}^1/_2\Sigma(n_{iA} - n_{iB})^2$$

where n_{iA} and n_{iB} are the numbers of residues of the ith type in two proteins A and B respectively, and the summation is carried out over all of the types of amino acid distinguished by composition measurements.

The composition data of Martin and Fridovich [5] were based on a polypeptide length of 156 residues for both *P. leiognathi* and swordfish enzymes, and give a value of $S\Delta n = 61$, i.e. they suggest that there are 61 ± 23 differences between the two sequences, excluding differences between glutamate and glutamine or between aspartate and asparagine, or about $39 \pm 15\%$ differences. The sequence determinations showed lengths of 151 for *P. leiognathi* [6] and 158 for swordfish [7], and, more important, one expects experimental error in composition determinations to exaggerate the amount of difference between them [2]. Recalculation of $S\Delta n$ from the sequence-derived compositions gives an uncorrected value of 93.5, or 95.5 after correction for the 7-residue difference in length [2]. This is considerably larger than expected, mainly because of a large discrepancy between numbers of threonine residues in the swordfish enzyme estimated originally [9] and found from the sequence [7]. This discrepancy was, unfortunately, aggravated by a smaller discrepancy in the opposite direction for the number of threonine residues in the *P. leiognathi* enzyme.

The corrected value of $S\Delta n = 95.5$ implies that 96 ± 36 residues of the longer sequence will be unmatched when it is aligned with the shorter sequence. This value represents about $60 \pm 23\%$ difference, and is in excellent agreement with the actual amount of difference, which is $158 - 42 - 2 = 114$, or 72%. (In obtaining this value two aspartate/asparagine pairs have been added to the 42 identities counted by Rocha et al. [7].)

It is hardly justifiable to conclude from this good agreement between prediction and observation that 'the very low sequence homology between swordfish liver copper/zinc superoxide dismutase does falsify the proposed gene transfer theory' [7]. The swordfish enzyme was, perhaps, an unfortunate choice of a fish enzyme for sequence determination and

Correspondence to A. Cornish-Bowden, Department of Biochemistry, University of Birmingham, P. O. Box 363, Birmingham, England B15 2TT

Enzyme. Superoxide dismutase, superoxide:superoxide oxidoreductase (EC 1.15.1.1).

Table 1. *SΔn values for copper/zinc superoxide dismutases from* Photobacter leiognathi *and various fish*
The values were calculated from the data in Table VII of [7] which are based on a polypeptide length of 156 for each enzyme (i.e. 312 residues per dimer). The data for swordfish were quoted in the original paper from data of Bannister et al. [9]

	P. leiognathi	Ponyfish	Red snapper	Atlantic croaker	Speckled trout	Black sea bass	White marlin	Swordfish
Photobacter leiognathi	0							
Ponyfish	16	0						
Red snapper	29	15	0					
Atlantic croaker	39	33	36	0				
Speckled trout	31	21	14	47	0			
Black sea bass	52	69	30	42	31	0		
White marlin	36	24	25	38	9	42	0	
Swordfish	61	42	34	47	46	76	34	0

```
                    5         10        15        20        25        30        35        40        45
P. leiognathi     Q D L T V K M T D L Q T G K P V G T I E L S Q N K Y G V V F T P E L A D - L T P G M H G
Cattle            A T K A V C V L K G D G P V Q G T I H F E A K G D T - - V V V T G S I T G - L T E G D H G
Swordfish         V L K A V C V L R G A G E T T G T V Y F E Q E G N A N A V G K G I I L K G - L T P G E H G
C. crescentus       /A V G V I K/                                                       /I A G G G P T A Z R/

                    50        55        60        65        70        75        80        85        90
P. leiognathi     E H I H Q N G - - - - S C A S S E K D G K V V L G G A A G G H Y D P E H T N L H G F P W T
Cattle            F H V G Q F G D N T Q G C T S - - - - - - - - - - - - - - A G P H F N P - L S K K H G G P K D
Swordfish         F H V H G F G D N T N G C I S - - - - - - - - - - - - - - A G P H F N P - A S K K H A G P H D
C. crescentus                                                                 /G H F H A - L S Z T H G A P K/

                    95        100       105       110       115       120       125       130       135
P. leiognathi     D D N H K G D L P A L F V S A N G L A T N P V L A P - - - L R T L K E L K G H A I M I H A
Cattle            E E R H V G D L G N V T A D K N G V A I V D I V D P L I S L S G E Y S I I G R T M V V H E
Swordfish         E D R H V G D L G N V T A D A N G V A K I D I T D - K I S L T G P Y S I I G R T M V I H E
C. crescentus   /P A S H A G B L G B V R T A Z T S V A K/       /B S L Z T L I A P Y S V A G K T V V L H A

                    140       145       150       155       160       165
P. leiognathi     G G D N H - - - - S D M P K A L G G G G A R V A C G V I Q
Cattle            K P D D L G R G G N E E S T K T G N A G S R L A C G V I G I A K
Swordfish         K A D D L G R G G N E E S L K T G N A G S R L A C G V I G T E
C. crescentus     V P B A L G Z G A T K Z L L K G G B A K/   /K A C V V I G/
```

Fig. 1. *Sequences of copper/zinc superoxide dismutases from four sources.* Sequences for *Photobacter leiognathi* [6], cattle [12], swordfish [7] and *Caulobacter crescentus* [11] are aligned as published by previous workers [7, 11]. The numbering does not refer to any one sequence but is arranged so that every locus in the alignment is numbered consecutively. Other alignments would be reasonable, the most obvious example being one in which the gap at locus 38 was removed by beginning the *C. crescentus* fragment at locus 34. These alternatives would alter the quantitative aspects of the discussion in the text somewhat but would have no effect on the qualitative conclusions. In counting identities, amide uncertainties in the *C. crescentus* fragments have been interpreted as identities wherever possible: thus the B (Asx) at locus 97 is treated as D (Asp), but the B at locus 100 as N (Asn)

for testing the gene-transfer hypothesis, given that the composition data indicated that it was the least similar to the *P. leiognathi* enzyme of all the seven fish enzymes considered. Values of *SΔn* for all pairs of copper/zinc enzymes included in the original study are shown in Table 1, where it may be seen that the composition of the swordfish enzyme shows large differences not only from the bacterial enzyme but also from the other fish enzymes. This may partly reflect the fact that the data for the swordfish enzyme were obtained in a different laboratory [9], as any biases in the composition measurements tend to cancel in within-laboratory comparison but to be amplified in between-laboratory comparisons. Nonetheless, to avoid any suggestion of wisdom with hindsight it seemed more appropriate to use the actual data quoted by Martin and Fridovich [5] without 'improving' them by the use of compositions derived from sequences.

It would appear that the gene-transfer hypothesis proposed by Martin and Fridovich [5] has not been disproved by any sequence data currently available, but that it would be highly desirable to have sequence data for the ponyfish enzyme. Meanwhile, it is significant that detailed study of the sequence of the *P. leiognathi* enzyme has recently shown [10] that it agrees closely with the eucaryotic enzymes in relation to active-site residues and predicted secondary structure, and that a satisfactory three-dimensional model of it can be constructed by analogy with the known *X*-ray structure for the cattle enzyme.

Additional support for the gene-transfer theory can be derived from the sequence data for fragments of the copper/zinc superoxide dismutase of a free-living bacterium, *Caulobacter crescentus* [11]. A comparison between the sequences of the enzyme from the two bacteria, *P. leiognathi* and *C. crescentus*, and two vertebrates, swordfish and cattle, is shown in Fig. 1. It is striking that the *C. crescentus* fragments agree much better with the eucaryotic sequences than with the *P. leiognathi* sequence, and much better than the *P. leiognathi* sequence agrees with the two eucaryotic sequences. In general, if one were interpreting the sequence data in the absence of any other phylogenetic information, it would be quite reasonable to regard the swordfish sequence as intermediate

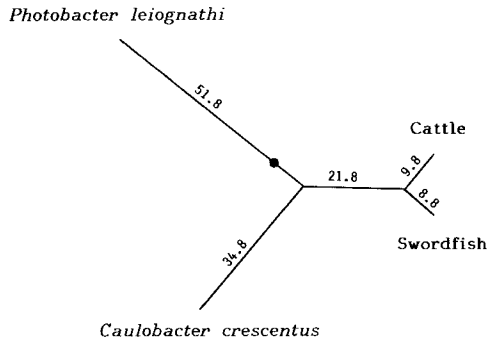

Fig. 2. *Minimum-length tree for the sequence data of Fig. 1.* The tree is derived from the data of Fig. 1 at loci for which fragments of the *C. crescentus* enzyme exist. It has a total length of 127 amino-acid substitutions and is the shortest tree that can be constructed from the data. Trees for the two other possible topologies have lengths of 128 and 129. Although the total length is integral, the lengths of the branches are not because in cases where substitutions could be placed in alternative locations they have been divided among the possibilities; for example, locus 6 requires two substitutions, which have been credited as 0.67 each for the central branch and the branches leading to the two bacteria. The circle represents the mid-point along the route between the two bacteria

between the two bacterial sequences, which differ more from one another than either does from the eucaryotic sequences.

Determination of the most parsimonious tree linking the four sequences in Fig. 1 (using only loci 4–9, 35–44, 75–89, 91–110, 115–155 and 158–164, i.e. the parts for which data are available for *C. crescentus*) yields the tree shown in Fig. 2, which has a total length of 127 amino acid substitutions. However, although the topology of this tree, in which the two bacteria are paired together, as are the two vertebrates, is the one most in agreement with normal ideas of phylogeny, it is only barely preferred over the alternatives: the two trees with *P. leiognathi* paired with swordfish and cattle have total lengths of 128 and 129. This very poor discrimination given by the maximum-parsimony approach is a consequence of the fact that there are only three loci (35, 75 and 135) for which the three topologies require different numbers of substitutions, as discussed elsewhere [13]. Either of these other trees would positively require gene transfer from eucaryote to *P. leiognathi* or to *C. crescentus*, or to both bacteria in two separate events, as otherwise one would have to make the absurd postulate that the swordfish and cattle lineages separated from one another before they separated from the bacteria. Even the tree shown in Fig. 2 presents serious problems, however, if one rejects the idea of gene transfer. If one places the root at the mid-point between the two bacteria, i.e. at the point indicated by a circle in Fig. 2, the tree suggests that the cattle enzyme has accumulated a minimum of 40 substitutions since divergence from the *P. leiognathi* lineage, presumably around 10^9 years ago, of which about 10 have occurred since divergence from the swordfish. Even allowing for the fact that the most parsimonious tree is unlikely to be the true tree, because of back substitutions and repetitive substitutions at the same locus, it is inescapable that rejection of gene transfer implies an enormous increase in the rate of evolution, seen in both swordfish and cattle lineages, in the last 3% of the time since the eucaryote-procaryote divergence.

The most parsimonious interpretation of the sequence data must surely be that, although Fig. 2 may give a reasonable account of the evolution of the four proteins, it cannot be regarded as a meaningful record of the evolution of their host species, i.e. it is clear that gene transfer has occurred, at least once and perhaps twice, so that the proteins have diverged much more recently than the species. The much larger numbers of substitutions in the branches leading to the bacterial proteins suggests a much faster rate of change in these lineages, which would be entirely reasonable, given that after transfer of a gene from a eucaryote to a procaryote a gene would find itself in a very different environment with quite different selective pressures.

REFERENCES

1. Cornish-Bowden, A. (1977) *J. Theor. Biol. 65*, 735–742.
2. Cornish-Bowden, A. (1979) *J. Theor. Biol. 76*, 369–386.
3. Cornish-Bowden, A. (1981) *Trends Biochem. Sci. 6*, 217–219.
4. Cornish-Bowden, A. (1983) *Methods Enzymol. 91*, 60–75.
5. Martin, J. P., Jr & Fridovich, I. (1981) *J. Biol. Chem. 256*, 6080–6089.
6. Steffens, G.-J., Bannister, J. V., Bannister, W. H., Flohé, L., Günzler W. A., Kim, S.-M. A. & Ötting, F. (1983) *Hoppe-Seyler's Z. Physiol. Chem. 364*, 675–690.
7. Rocha, H. A., Bannister, W. H. & Bannister, J. V. (1984) *Eur. J. Biochem. 145*, 477–484.
8. Flohé, L., Günzler, W. A., Kim, S.-M. A., Ötting, F., Steffens, G.-J., Bannister, J. V. & Bannister, W. H. (1984) in *Oxygen radicals in chemistry and biology* (Bors, W., Saran, M. & Tait, D., eds) pp. 793–801, Walter de Gruyter, Berlin.
9. Bannister, J. V., Anastasi, A. & Bannister, W. H. (1977) *Comp. Biochem. Physiol. 56B*, 235–238.
10. Bannister, J. V. & Parker, M. W. (1985) *Proc. Natl Acad. Sci. USA 82*, 149–152.
11. Steinman, H. M. (1982) *J. Biol. Chem. 257*, 10283–10293.
12. Steinman, H., Naik, V. R., Abernethy, J. L. & Hill, R. L. (1974) *J. Biol. Chem. 249*, 7326–7338.
13. Cornish-Bowden, A. (1983) *J. Theor. Biol. 101*, 317–319.