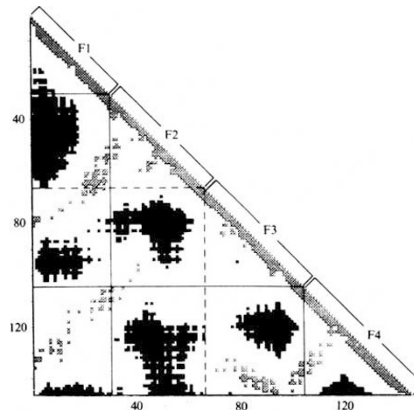
Exons and the structure, function and evolution of haemoglobin

from C.C.F. Blake

OVER the last few weeks there has been a flurry of papers in *Nature* that bear directly on the meaning of coding sequences or exons in the haemoglobin genes. In the mouse α- and β- globin genes there are three exons corresponding to amino acid residues 1–31, 32–99 and 100–141; and 1–30, 31–104 and 105–146, respectively. When the amino acid sequences of the α- and β-globins are aligned to maximize their structural homologies, their intron/exon junctions coincide. This immediately suggests that the 'meaning' of the exons is related in some way to the structure and function of the protein. The recent papers address themselves to the problem by analysing the exon-encoded regions in terms of the structure, function and evolution of the haemoglobin molecule.

Gō[1] has ingeniously used the diagonal plot (see the figure) to define regions of the polypeptide chain that are distant (> 27Å) from one another in the globin fold. He has found that there are four such regions, which are not domains, but which can be called sub-domains, or perhaps 'compact structures'. When included on the diagonal plot the exon boundaries neatly divide off these compact structures from one another, with the exception of the large central exonic region which is composed of two compact structures. Because of this Gō suggested that the central exonic region might consist of two 'fused' exonic regions with a division somewhere between residues 66 and 71. Quite remarkably this prediction has been rapidly verified by Marcker and colleagues[2] (see this issue of *Nature*, p.677) in their determination of the structure of the leghaemoglobin gene from soybean. This gene is composed of four exons, corresponding to amino acid residues 1–32, 32–68, 69–103, 104–C terminus. Marcker's results suggest that exons are generally very stable, but that they can undergo. fusion or separation. Gō's successful prediction of an exon on the basis of protein structure adds powerful additional support to the idea[3] that they correspond to compact protein structures.

As Gilbert's suggestion[4] of a relation between exons and protein functional units appears to have been validated in lysozyme[5], consideration of this aspect of haemoglobin is of particular interest. In haemoglobin the functional correlation began with the suggestion[6] that the central exonic region appeared to correspond to a haem-binding unit. This suggestion was



A plot of all the distances between α-carbon atoms in the haemoglobin β chain. Both ordinate and abscissa are the residue sequence number. Distances between pairs of atoms greater that 27 Å appear as dark regions. Solid lines are drawn at the joints between polypeptide segments (F1, F2 + F3, and F4 respectively) encoded by different exons — between residues 30 and 31, and 104 and 105. Note that these lines scarcely cross the dark regions.

analysed in detail by Eaton[7], who not only confirmed this property of the central exon, but also showed that the distribution of intersubunit contacts showed a strong correlation with the exonic regions. Nearly all the $\alpha_1\beta_2$ contacts are in the central exonic region, and nearly all the $\alpha_1\beta_1$ contacts are in the third exonic region. This seems to suggest that while the contacts responsible for the early cooperativity of $\alpha_1\beta_2$ dimers may have developed within the haem-binding structure itself, the later tetramer cooperativity required the presence of the third coding sequence, which Eaton speculates may have replaced its predecessor by recombination. At the same time Beychok and his colleagues[8] began an experimental examination of the properties of the central exonic region of human haemoglobin, which can be excised from the globin chain by clostripain treatment. Their studies first demonstrated that the separated central region was indeed capable of binding haem tightly and specifically. An extension[9] of this analysis has now shown that although the central exon peptide can bind haem, it cannot stabilize the haem–dioxygen complex, which requires both the presence of the side exon products and the complementary subunit. The regainment of activity is accompanied by an increase in helical content, suggesting that a structural conformational change may be involved. With the verification of the haem-binding function of the central exonic region, it is interesting to consider the significance of its division into two in leghaemoglobin. Despite Marcker's assertion that it has no obvious functional correlation, the division in fact neatly segregates the proximal and distal haem contacts. This suggests that the divided central exonic region may be a primitive form and that the two exons fused along the haemoglobin line to ensure a single protein product capable of specifically interacting with haem.

In a consideration of their results, Beychok and his colleagues note that the central exon product, isolated or in noncovalent association with the side peptide fragments, has spectral characteristics and other behaviour reminiscent of the b-type cytochromes. In 1975 Rossmann and Argos[10] demonstrated a structural similarity between the globins and cytochrome b_5, which was later extended to include cytochrome c_{551}[11]. In each case the closest agreement involved the haem-binding region and corresponds rather precisely to the central exon in the globin chain. When the protein folds were maximally superposed, the haem irons were found to be separated by 4.2Å in the globin – cytochrome b_5 comparison and by 5.3Å in the globin – cytochrome c_{551} comparison, and the haem normals by 13° and 9° respectively. The results of these chemical and structural comparisons suggest that the globins and the cytochromes may have evolved from a common haem-binding domain encoded by one or more exons, which by combining with other gene elements has generated a multiplicity of haem-binding proteins of diverse function. These conclusions are very similar to those obtained from the examination of the exon structure–function relationships in hen and T4 lysozymes[5], and lend strong support to Gilbert's proposal that in eukaryotes exons code for functional protein units which can serve in rapid protein evolution. The great puzzle now is why prokaryotic genes do not have this apparently advantageous structure. □

C.C.F. Blake is in the Laboratory of Molecular Biophysics, Department of Zoology, Oxford.

1. Gō, M. *Nature* **291**, 90 (1981).
2. Jensen, E.O., Paludan, K., Hyldig-Nielsen, J.J., Jorgensen, P. & Marcker, K.A. *Nature* **291**, 677 (1981).
3. Blake, C.C.F. *Nature* **273**, 267 (1978).
4. Gilbert, W. *Nature* **271**, 501 (1978).
5. Artymiuk, P.J., Blake, C.C.F. & Sippel, A.E. *Nature* **290**, 287 (1981)
6. Blake, C.C.F. *Nature* **277**, 598 (1979).
7. Eaton, W.A. *Nature* **284**, 183 (1980).
8. Craik, C.S., Buchman, S.R. & Beychok, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1384 (1980).
9. Craik, C.S., Buchman, S.R. & Beychok, S. *Nature* **291**, 87 (1981).
10. Rossmann, M.G. & Argos, P. *J. biol. Chem.* **250**, 7525 (1975).
11. Argos, P. & Rossmann, M.G. *Biochemistry* **18**, 4951 (1979).

0028-0836/81/260616-01$01.00 ©1981 Macmillan Journals Ltd