

Do genes-in-pieces imply proteins-in-pieces?

from C. C. F. Blake

THERE has been some discussion in these columns recently of the significance and implications of the discovery that eukaryotic structural genes are interrupted by stretches of non-informational DNA. Gilbert (*News and Views* 271, 501; 1978) has suggested that this organisation could serve to speed evolution by providing mechanisms for the generation of novel proteins from parts of old ones. Subsequently Doolittle (*News and Views* 272, 581; 1978) has argued that the 'genes-in-pieces' structure is in fact a primitive form that was present in the genomes of the common ancestors of prokaryotes and eukaryotes.

These hypotheses are of great interest to those investigating protein structure because the structures of most larger proteins (molecular weight >20,000) also seem to be made up from 'pieces'. On the larger scale the pieces correspond to the independently-folded globular units called domains, although there are indications that domains may themselves be composed of smaller units called supersecondary structures (Rao & Rossmann *J. molec. Biol.* 76, 241; 1973). Although not enough is known about protein folding to rule out the possibility that domains are formed by some folding mechanisms particular to long polypeptide chains, so far no clear physicochemical explanation of their presence has been proposed. This leaves open the possibility of a specifically genetic origin, and in particular the question of whether they represent relics of the process by which the protein was

originally formed. Suggestions along these lines have been made (Rossmann *et al. Nature* 250, 194; 1974; Levine *et al. Nature* 271, 626; 1978) to account for the presence of domains of apparently similar structure and function in different enzymes.

In the light of these observations I wish to consider here the consequences of the rather obvious assumption that exonic regions of DNA (Gilbert has proposed that we use 'exon' for the stretches of structural DNA, and 'intron' for the interrupting stretches of non-informational DNA) correspond to integrally folded protein units—domains or supersecondary structures. If a new combination of exonic regions is to produce a novel protein molecule with new activity that confers advantage, then the new polypeptide chain must meet two conditions simultaneously. First, the chain must be able to fold into a stable globular form, as a necessary but not sufficient requirement for the second condition, namely an active site with new properties. If the new protein is made from random sections of old ones, the chances of meeting both conditions at the same time would undoubtedly be very small, perhaps vanishingly so. However, if the DNA was so organised that an exonic region corresponded to a folded protein fragment, then combinations would be much more likely to be stable through being the 'sum of parts', thus markedly improving the chances of meeting the dual requirements. If a protein built from two or more folded units maintains, as is most probable, the same folding pathways as in the original separated, or differently combined units, and, as also seems probable, subsequent point mutational events could not erase the resultant domain struc-

ture, it would remain as a permanent feature of the protein to be easily recognised when its structure was examined.

Furthermore in a new enzyme formed by the combination of two or more domains, the only region where the new constellations of amino-acid side-chains required to produce an active site with new properties could exist is specifically at the interfaces of the domains. All other regions of the new protein would contain the same patterns of amino acids as in the protein units from which they originated. In addition the domain interface could readily provide the surface depression or cleft that characterises active sites, and is an integral part of their activity. In every domain enzyme that I am aware of, the active site is in fact located precisely in the domain interface with each domain contributing to the cluster of amino acid side-chains that makes up the active site.

It can be seen that the evidence from protein structure is at least consistent with Gilbert's proposal that new protein molecules can be constructed from parts of pre-existing ones, and suggests the possibility that the parts, the transcripts of the exonic regions, are folded protein units. With Doolittle's extension of the time scale, necessary because many domain enzymes from prokaryotes and eukaryotes have homologous structures, there is now the possibility of a general mechanism for the formation of domain proteins that seem to be so common. Finally, it can be noted that the proposed correlation of exonic regions with folded protein units is open to test as soon as the eukaryotic DNA that codes for one of the known domain enzymes is sequenced. □

C. C. F. Blake is a lecturer in Molecular Biophysics at the University of Oxford.

also turned out to be a glucan and was active in inducing the formation of rishitin, a terpenoid phytoalexin, in the potato tissues.

The most recent development in this field is published by M. Stekoll and C. A. West of the University of California at Los Angeles in the January issue of *Plant Physiology* (61, 38; 1978) where they report the purification and characterisation of an elicitor from the culture filtrate of the fungus *Rhizopus stolonifer*. Although *R. stolonifer* could hardly be said to be a pathogen, in the strict sense, nevertheless the active substance induces the formation of a macrocyclic diterpene phytoalexin, known as casbene, in castor bean (*Ricinus communis*). Cas-

ebene was originally identified by West's group as being synthesised at relatively high rates in cell free extracts of castor bean seedlings which had previously been challenged by any one of a number of several fungi, including *R. stolonifer* and *Aspergillus niger* (Robinson & West *Biochemistry* 9, 70, 80; 1970; Sitton & West *Phytochemistry* 14, 1921; 1975). In the present very exhaustive study, Stekoll and West have developed a cell-free assay for casbene elicitor activity, in which split seedlings of castor bean are exposed to the test substances for a few hours, after which a cell-free extract is prepared and assayed for casbene synthetase; that is, the capacity to incorporate ¹⁴C-mevalonate into casbene. Chemical

characterisation and fractionation indicated that the elicitor is, in this case, a protein. On the other hand although substantial loss of elicitor activity occurred after Pronase treatment, periodate oxidation completely destroyed the capacity of purified fractions to stimulate casbene synthetase. The molecular weight of the active fraction on gel filtration was about 30,000 ± 5,000 and the authors conclude that the elicitor is a protein, and probably a glycoprotein.

Thus, two classes of elicitor seem to exist, one based on specifically branched glucan residues, which apparently do not contain any necessary protein, and one which is probably proteinaceous, but which requires car-