PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The Evolution of Plants and Animals Under Domestication: The Contribution of Studies at the Molecular Level [and Discussion]

E. M. Jope, C. M. Ann Baker, C. Manwell, D. Boulter, D. J. S. Hulmes and A. Miller

Phil. Trans. R. Soc. Lond. B 1976 **275**, 99-116
doi: 10.1098/rstb.1976.0073

Phil. Trans. R. Soc. Lond. B. **275**, 99–116 (1976) [99]
Printed in Great Britain

The evolution of plants and animals under domestication: the contribution of studies at the molecular level

By E. M. Jope, F.B.A.
Archaeology Department, The Queen's University of Belfast, Northern Ireland

Protein molecules are essential catalysts in life processes and also form much of the substance of living material. Their three dimensional structures determine their biological function. Their biosynthesis is primarily determined by arrays of nucleic acid macromolecules (DNA and RNA), and the amino acid sequences that constitute their long spatially organized peptide-chain molecules reflect at one remove this DNA coding system, and thus record a step-by-step history of some of the viable genetic events (natural or man-controlled) that have created the organism and the breed.

Amino acid sequences can be used to trace the progress of controlled breeding in two ways: by extrapolation back from living breeds, and by analysis of ancient protein material. Of the latter, bone or tendon or skin collagens and hair keratins are the most perfectly preserved as molecular structures through 20000 years and indeed much longer. Amino acid sequences are expensive to determine (collagen has 1052 amino acid residues), and the potential of this palaeobiological information has been as yet little exploited. The first approach has, however, been more explored, in both plants and animals.

Several protein systems must be studied in conjunction to reveal the phylogenetic threads in any one breed. As the three dimensional quaternary structure of protein molecules becomes more appreciated in relation to biological function, and as new techniques and procedures are developed, amino acid sequence data can become more informative in our ultimate understanding of early selective breeding.

Plant and animal breeders have from the earliest times been primarily interested in the results of their manipulations in terms of morphology, function, productivity, and behaviour. But the full achievement of their work has come about through a great multiplicity of individual molecular processes, with much interaction and feedback, which collectively carry into effect a genetic coding system packed in the genetic material (DNA and RNA) of the chromosomes. These processes and their products must therefore be analysed at molecular level to aid in retracing the steps in the multi-streamed cumulative progress of domestication and selective breeding that underlies food-production as we know it today.

The genetic instructions are coded as sequences of nucleotide base-pairs that make up the DNA of the chromosomes, and it is these base-pair sequences that define the character and ancestry of an individual. But these actual sequences are very difficult to determine, and the available methods give ambiguities of interpretation (Kohne 1968), because the information is so condensed. These instructions are, however, largely carried into effect by the proteins whose biosynthesis they dictate, as step-by-step additions of specific amino acids to build up the polypeptide chain; chain-termination is also coded. The amino acid sequences of the proteins thus reflect the composition of the genome, and the information is more openly spread out. The order of amino acids in these precisely replicated protein chains determines the spatial configuration of the protein molecule, and hence its chemical and biological function. Changes

in the DNA base-pair arrangement (by single substitution, deletions, duplication, etc.) will produce corresponding changes in amino acid sequence (mutations). Some of these mutations will be acceptable to the system, and become a viable replicable part of the organism and transmissible, with all the other changes consequent thereon, to its progeny; others will be unacceptable metabolically and biologically and disappear. There may also be occasional regular secondary changes after primary biosynthesis, but these can usually be detected (as for instance with the hydroxylation of proline in collagens, below, p. 105).

Amino acid sequences of proteins are, therefore, an important and revealing source of genetic information concerning the ancestry of a plant or animal individual and breed; they are also among their primary biological parameters. Though expensive to determine, many more amino acid sequence data are vitally needed in biological research today.

There is ultimately no substitute for full and exact sequencing of proteins; only from such data can changes and relationships be shown up in sufficient detail. There is, however, an accumulating body of data at a less specific level, grouping proteins by their behaviour in simple or gel electrophoresis. Acrylamide gel disc electrophoresis is particularly valuable, as with added detergent SDS it groups many molecules according to their molecular mass (Furthmayr & Timpl 1972). These rapid mass procedures can provide preliminary surveys of whole classes of material, showing up anomalies for more detailed investigation (see, for example, Manwell & Baker 1970; Johnson 1972); but with some classes of large protein molecule, the finer nuances may remain masked, such as single amino acid differences, which may nevertheless be crucial as phylogenetic evidence.

Every protein carries in its amino acid sequence and its spatial folding structure a record of its history and ancestry, of the viable genetic events that have created it. It is clear from the mass of data already available that different protein molecule types have had very different rates of evolution (Dickerson & Geis 1970; Dayhoff 1972; Chirpick 1975), The factors governing these rates are not yet at all clear, yet this aspect of the subject has been given little attention. Some puzzling examples are set out below (figure 1, sheep and goat haemoglobins). Many factors are involved, starting with the nucleotide base-pair changes in the archive DNA, which is after all a two-way probability equilibrium process and thus very sensitive. The microstructural chemistry and micro-energetics of the DNA and subsequent RNA events, especially in relation to chromosomal proteins, need intensive investigation (Kohne 1968).

Amino acid sequences may be used to trace the course of controlled breeding in two ways. First, by extrapolation back from living breeds. Secondly, by analysis of ancient protein material, for some classes of protein such as the collagens of bone and skin, or the keratins of hair, wool, horn or claw, are often well preserved as molecular structures through archaeological and even geological time (Wyckoff 1970; Jope 1969) (table 1). This latter approach has been very little explored, but as a potential source of information on domestication and selective breeding our archaeological bone and plant remains should be fully exploited in this respect and is more fully discussed below. For ancient plant material this would involve an extensive study of husk proteins.

The former approach has been more fully developed, and though its time scales can only be indirectly inferred, it has the great advantage that a whole range of different protein types (including the soluble as well as structural) can be comparatively investigated on one particular breed and as between related breeds. The complementary data thus give a cumulative picture, and the wide range of protein types strengthens the value of comparative studies.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Plant proteins are difficult to isolate in sufficient amounts and with the necessary molecular homogeneity, which is one of the reasons why there are still so few plant protein complete amino acid sequences available. (Boulter 1972, p. 228); for the food-crop plants that interest us most here there are still almost none. Yet amino acid sequence data are already showing their potential in plant research (Boulter 1972, 1974; pp. 114–5). They serve to refine the picture derived from electrolytic resolution on wheat and oat seed proteins (Johnson & Hall 1966; Johnson 1972; Ladizinsky & Johnson 1972) and resolve interpretation that conflict with those from cytogenetic studies (Zohary, Harlan & Vardi 1969) as to the extent of hybridization. The same promise is there for the development of maizes, where difficult problems still remain (Schwarz *et al.* 1965; Beadle 1972; Galinat 1971; Mangelsdorf 1974); but for the rices, amino acid sequencing of their proteins has not yet been started.

Work on plant proteins is so far largely on seed proteins, and these can survive well in ancient maize (Boulter, Derbyshire, Jope & Wheat 1976). If we are to exploit the full potential of ancient seed and cereal deposits from archaeological contexts in attempting to trace the early history of food production, it is likely that a fuller study of husk proteins would be needed; fortunately the new techniques of amino acid sequencing now becoming available make it possible to operate with the infinitesimal amounts of protein available from such husk material, and we may hope to see this work developed.

Animal proteins

Amino acid sequences have been more fully explored on proteins from the animal kingdom (Dayhoff 1972; Croft 1973), stimulated perhaps by the pressure of medical research, and can already contribute considerably to the complex history of animal breeding, which for pre-Columbian times means the Old World only. We shall concentrate here mainly on problems of sheep and goat (Ovicaprids) and their near relatives. The Ovicaprids have been important factors in agricultural economy since its beginning, and they provide us with a wide range of protein types: immuno- and other serum proteins, enzymes, haemoglobins and other carriers, and the milk proteins – as well as the insoluble fibrous proteins, the keratins of wool and horn and the collagens of bone, cartilage, tendon and skin.

A considerable range of protein polymorphisms has accumulated in the world sheep population through the intensive breeding efforts of some 10 000 years (see, for example, Agar, Evans & Roberts 1972; Dayhoff 1972; Rasmusen & Tucker 1973; Manwell & Baker 1970). These protein polymorphisms can, however, be grouped and systematized to some extent by considering the amino acid sequence data, first, of the haemoglobins, which themselves give a complex enough picture. In dealing with the various breeds of domestic sheep we must consider together the whole range of Ovicaprids; sheep, mouflon, Barbary sheep, and goat. This will be achieved mainly through the haemoglobins and subsequently other protein systems.

Haemoglobin (Hb), the universal mammalian oxygen carrier, has a long and complex molecular history. Blood haemoglobin in mammals is a large composite molecule (molec. mass 68 000) made up of four subunits, two of the α-chain with 142 amino acids, and two of the Hb β-chain with 146 amino acids (Perutz 1964; Dickerson & Geis 1969). It is only through this multiple folded structure that Hb acquires its unique but minutely modifiable properties. There is also muscle haemoglobin, myoglobin, which operates as a single unit of mol. mass 17 000 (152 or 153 amino acids) (Dayhoff 1972), on which there are no amino acid sequence data available for sheep. There is in mammals an immense production rate of blood Hb; in man,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

with a red blood cell life of 116 days (Shemin & Rittenberg 1945; Jope 1946) the entire blood Hb at any one time (about 0.2 kg) is therefore replaced more than 200 times in a normal lifespan of 70 years. So that a man in 70 years will produce about 50 kg of haemoglobin. There is thus a great range of haemoglobin production rate in the various genetically distinct systems. The red cell life span for sheep and cattle is not yet exactly known.

In the sheep-mouflon-Barbary sheep-goat group, the Hb β-chain can exist in at least seven molecular forms with different electrolytic mobilities. The differences as known in terms of amino acid sequences are summarized in figure 1, which at once stresses the serious gaps in our knowledge of the sequences for these Hbs; particularly serious are the lack of mouflon Hbα and goat HbβC. Figure 1 does, however, show up some remarkable interrelations, not always easy to interpret clearly. The exact, quantized, differences registered in the amino acid sequence changes are, moreover, a refinement beyond what can be shown up by electrophoretic analyses, and are thus potentially a more sensitive and precise index of phylogenetic relations.

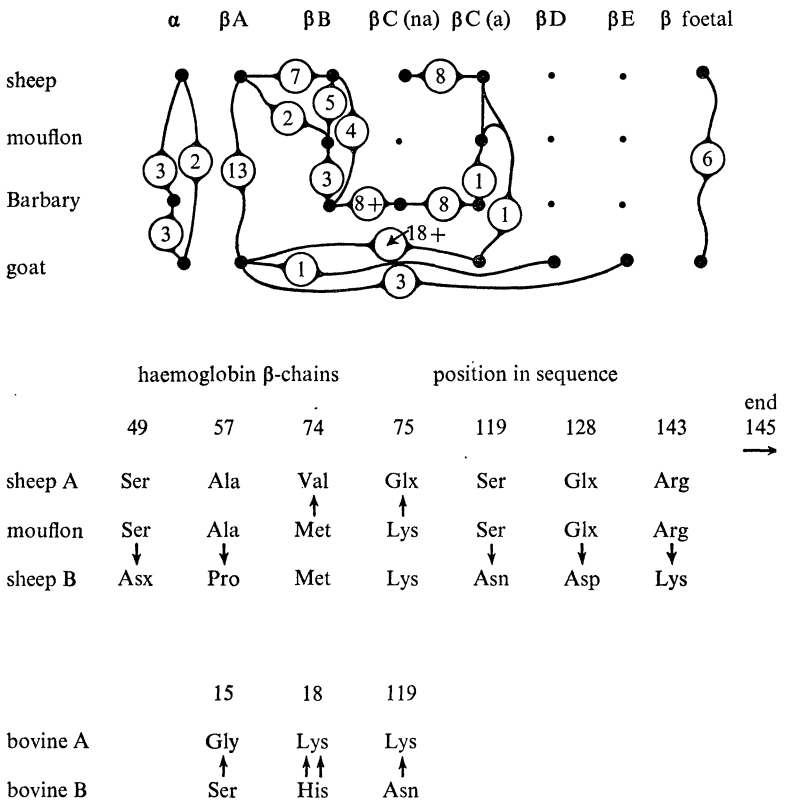

FIGURE 1. Diagrams summarizing the number of single amino acid differences in the various Hb chains of Ovicaprids. HbβC(a) is the Hb system coming into play in response to severe induced anaemia; HbβC(na) is the corresponding Hb under non-anaemic conditions. The lower diagrams show the chain positions of some of these differences, and the arrows indicate the number of nucleotide base-pair changes.

Figure 1 is constructed, mainly from the amino acid sequence data of Huisman and his colleagues at Augusta, Georgia (Huisman *et al.* 1965, 1967, 1970, 1972, 1973; Adams *et al.* 1968, 1969; Wilson, Miller & Huisman 1970; Wrightstone, Wilson, Miller & Huisman 1970). It reveals the close relations between the sheep and goat Hb systems, and is intended to show up the possibility of any kind of intermediates, or whether the complex system is due to

hybridizations (below, and p. 109). Only a few salient aspects can be discussed here. Mouflon is very close to goat in its HbβB chain, but identical with sheep for the HbβC(a′) chain. The latter, however, is a peculiar phenomenon; it comes into play only when the animal is artificially made severely anaemic; it then rapidly provides replacement for the HbβB chain. Hbα chain production presumably itself rapidly increasing to provide the necessary counterpart to make up the $(\text{Hb}\alpha)_2$ $(\text{Hb}\beta)_2$ molecule. This HbβC(a) system therefore lies dormant in the normal way almost all the time. Barbary sheep (*Ammotragus lervia* Pallas, 1777) is taxonomically enigmatic; its Hbα chain shows it to be further from either sheep or goat than they are from each other. Yet in its HbβB chain it is closer to mouflon than to sheep, and in its HbβC(a) chain it is very close to mouflon which is here identical with sheep.

Sheep and goat themselves may appear far apart (HbβA) or very close (Hbα). It may be therefore that, as Manwell & Baker suggest (1975; below, pp. 109–10), this complex pattern of interrelations can only be systematized by invoking hybridizations, and the full explanation must await the filling in of all the necessary sequence data to complete figure 1. This stresses the urgent need for increased intensity of sequencing research.

In a recent paper dealing very cogently with hybridization under domestication, Manwell & Baker (1975) make one suggestion of great interest to prehistorians and historians of agriculture. They argue that sheep HbβA and HbβB represent two ancient sheep populations, with mouflon HbβB independently representing a common ancestral line, from which it has diverged less, and that with no known intermediates, the present *Oves aries* stock was probably the result of hybridization between populations with HbβA and HbβB. They then note that HbβA has a higher oxygen affinity than HbβB, and is linked in a supergene with HbβC(a), which is only brought into operation under conditions of severe anaemia; the HbβA stock would therefore be better adapted to life at higher altitudes, while the HbβB stock would be more suited to lowland life. They then suggest that the practice of transhumance might well underlie such hybridizations, in which we should accordingly search for some variety. In the light of this a detailed molecular study of Hb and other proteins of the wild sheep of the varied terrains of northern Iran, of which so much is already known cytogenetically (Nadler, Lay & Harrison 1972) would be of great value. All too rarely can the varied methodological approaches be brought to bear on a problem, using the same relevant material.

Sheep have been selectively bred for more varied purposes than most animals, for their meat, their milk products, and above all for their wool. Consequently they show great variety in their multiple molecular species, as well as morphologically. We are fortunate in the British Isles in still preserving groups of some ancient breeds, such as the ancient stocks of Soay, Cotswold, and White-faced Woodland breeds at the Rare Breeds Survival Centre at Guiting Power in Gloucestershire (Henson 1974). We should like to know more about the ancestry of the Soay sheep, which could well be clarified by a concerted amino acid sequencing programme on their proteins. The same applies to other breeds of which we are fortunate enough to have surviving flocks available at Guiting Power, such as the mediaeval stock of Cotswold, or the White-faced Woodland of Derbyshire, whose good felting is so beloved of hatters, and into which 'Farmer George' introduced Spanish Merino strains in the 18th century (Ryder 1964, 1968).

Some of these ancient breeds of sheep may no longer be of any economic importance, but they do provide reserves of variety for the gene pool, the importance of preservation of which at least plant breeders have become increasingly aware (Harlan 1975).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Fibrous proteins: keratins

With the keratins of wool, hair, horn and claw we are dealing with the class of structural proteins whose function requires that they be protected from hydrolytic attack, achieved by organizing a non-polar envelope surface. They give material whose molecular structures can be preserved intact through archaeological and geological time.

Few keratin amino acid sequences are yet available (Dayhoff 1972, D303) and those are mostly on the high-sulphur fraction of Merino wool. The molecules have a high cysteine content, and the cysteines tend to conserve their position in the chain, the structure being held by –S–S– interchain links. Considerable variation seems possible in many other portions, but wool fibres are complex structures, formed by a complex process involving keratinization and death of whole cells (Dickerson & Geis 1969, pp. 37–50), and keratins in wool fibres are molecularly heterogeneous, involving at least two groups of proteins, as shown by giving several different end-groups. At least parts of the molecule have a regular ordered α-helix structure (Astbury 1932; Dickerson & Geis 1969, 38–40). It would need laborious separation treatment on columns, but amino acid sequences of keratins should certainly be explored as potential material for phylogenetic information concerning the progress of selective breeding in material preserved from antiquity.

TABLE 1. AMINO ACID ANALYSES OF BONE COLLAGENS $[a1(I)]_2\alpha 2$

	modern						
	horse femur	ox	ox tibia	RB ox tibia	Neolithic ox tibia	Pleistocene horse femur	cartilage; modern human $\alpha 1$(III) dentine
Gly	322.3	316.9	329.1	320.9	331.4	325.5	350
Pro	118.5	118.1	120.1	118.7	127.8	117.2	107
Hyp	88.7	102.5	104.9	85.0	87.8	95.3	125
Ala	112.2	107.9	108.1	124.1	118.4	127.4	96
Val	27.4	24.4	29.4	24.9	22.6	25.0	14
Leu	28.4	30.1	22.8	25.5	25.1	25.0	22
Ile	12.0	13.4	12.0	13.6	10.8	12.2	13
Ser	36.0	34.8	32.4	30.4	26.3	33.5	39
Thr	19.2	18.6	18.3	14.8	12.6	18.6	13
Asp	45.3	42.5	47.2	45.6	44.8	45.7	42
Glu	73.2	63.8	70.5	80.1	79.5	72.8	71
Met	4.9	5.0	4.8	4.5	2.2	1.9	8
Phe	14.9	19.3	14.5	11.3	13.4	12.2	8
Tyr →	4.4	7.2	5.0	1.2	1.9	0.6	3
Lys	28.7	30.3	25.5	33.1	36.5	24.7	30
Hylys	4.6	8.2	5.0	5.3	5.1	5.9	5
Arg	53.4	50.3	49.9	53.4	48.8	54.4	46
His	6.5	6.1	5.5	3.3	5.2	2.4	6
Cysteine							2
analysis	Tucson	Tucson	Belfast	Belfast	Belfast	Tucson	Alabama

Tucson data: Wyckoff 1972; Belfast data: Jope, H.M., unpublished.

Fibrous proteins: collagens

Potentially more important for our purpose of acquiring palaeophylogenetic information could be the *collagens* of bone, tendon and skin. Acrylamide gel electrophoresis shows that the full length of their long polypeptide chains can survive unbroken through archaeological and even geological time (Jope 1969, bone of 430×10^6 years; Wyckoff 1970, 1972). Table 1 shows

sample amino acid analyses on bone collagens from archaeological contexts; in many cases agreement with modern overall analyses is good, and such preliminary analyses should clearly govern the choice of bone samples for further investigations. Table 1 shows, moreover,

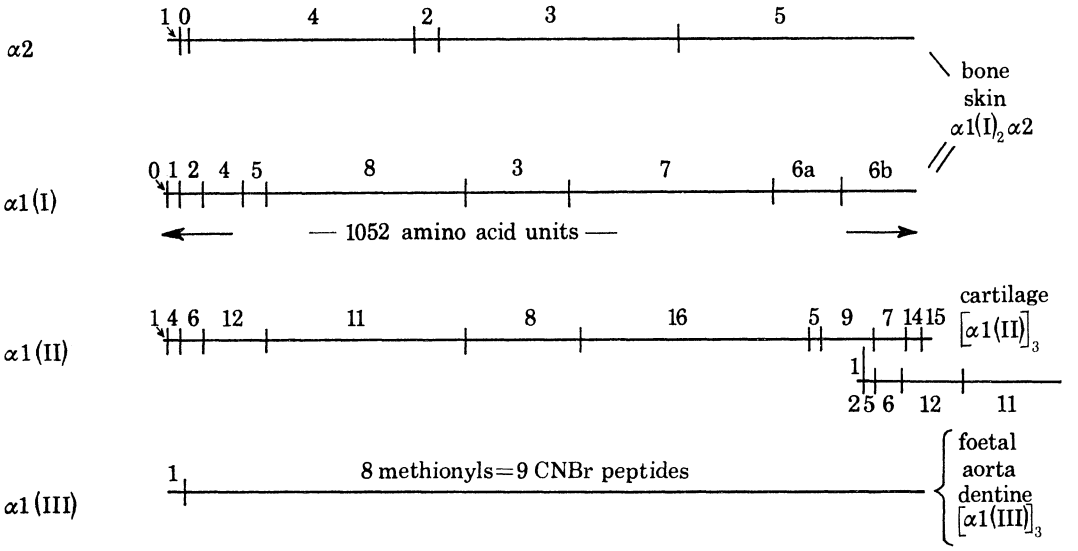

FIGURE 2. Collagen chains broken by CNBr at the methionyl positions into peptides, Cb 1, 2, 3, etc.

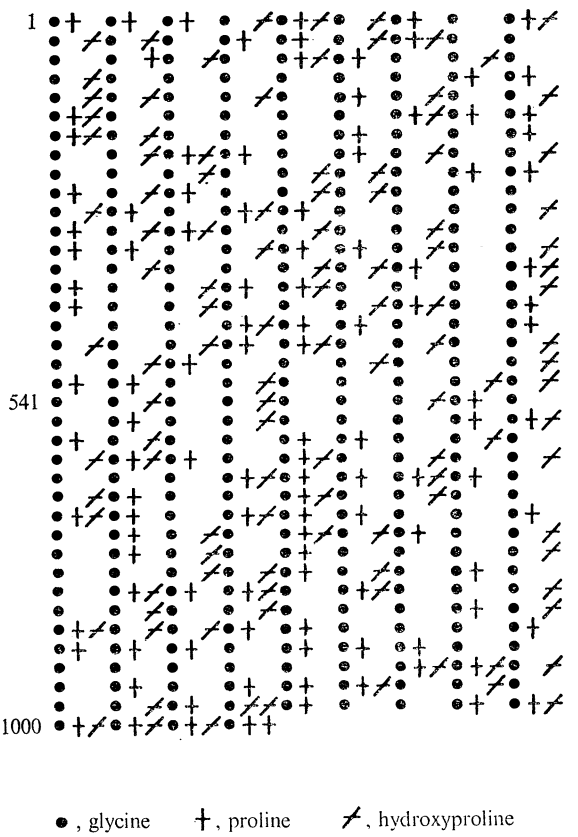

FIGURE 3. Amino acid sequence of collagen α – 1 molecule (1–1011) summarized diagramatically to show the regular repeating structure, with the small glycine ● in every third position (necessary for the close packing of the triple helix), and the deposition of prolines + and of hydroxyprolines ≠ providing corner-turning.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

that even the more labile amino acids (methionine, with its –SH, is probably the most evanescent, and a useful criterion for preservation) can survive in bone collagens from archaeological contexts. Amino acid analyses should clearly govern the choice of bone samples for further investigations.

Actual amino acid sequencing of collagens is more formidable. Recent work has shown that there are several distinct collagen chains. Bone collagen is a triple-helical structure, with two helices of the α-1(I) chain closely wound with one α-2, to give $[\alpha 1(I)]_2\alpha 2$. Cartilage collagen would have been simpler to work, for it consists of three chains the same, $[\alpha 1(II)]_3$, and no laborious preliminary chain separation would have been necessary. Each $\alpha 1(I)$ is 1052 amino acids long, a very specialized structure, with glycine imperative at every third position (figure 3) as the smallest amino acid, to allow the close winding of the triple helix that gives the rigid rod properties. Each of these triple helix rods associates with four other rods *en échelon*, thus overlapping to give lengthwise tensile strength (figure 2) (Miller 1973; Piez & Miller 1974).

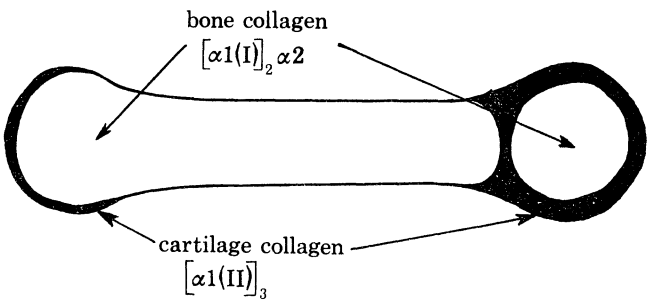

FIGURE 4. A developing long bone showing differential distribution of cartilage collagen $[\alpha 1(II)]_3$ and bone collagen $[\alpha 1(I)]_2\alpha 2$. The bone develops initially from a cartilage model (after Miller 1973).

After a decade of work by teams the world over, we now have a complete but composite amino acid sequence for the 1052 amino acids of the collagen $\alpha 1$ chain (Hulmes *et al.* 1974). It is certainly a daunting task to amass the homologous data that would reveal any mutational differences between breeds: a single amino acid different here and there in the 1052; yet it can be done, and it is our main hope for extracting genetic information on early breeding from ancient material. The molecular structural dictates are very stringent: glycine in every third position through 1011 of the 1052, and a requisite number of prolines or hydroxyprolines for corner-turning to give the helix. But there are more lax positions where the requirements of the helix are less vital, especially in the non-helical tails (table 2; Bornstein & Nesse 1970). Amino acid sequencing alone can reveal this with genetic exactness; Electron microscopy can help by revealing any lengths of the chain where anomalies occur, so that only specific peptide lengths need to be sequenced in detail (Rauterberg & Kuhn 1968; Miller & Parry 1972). Collagen sequencing is only possible by breaking the long chain into smaller peptides; figure 2 shows for instance how cyanogen bromide breaks the collagen chain at every methionyl.

Collagen molecules are biosynthesized initially as procollagen monomers with non-helical tails that prevent them from self-assembly into the triple-helical collagen trimer structure until the appropriate time and location for bone, skin or tendon formation, when enzymatic removal of the tails allows this self-assembly, leading to further association into microfibrils (Bornstein 1974*a*, *b*). Bone itself has a complex structural history, summarized for a long bone in figure 4 (Miller 1973). The bone is first developed as a model in cartilage collagen, $[\alpha 1(II)]_3$, which is

then gradually replaced through selective enzymic action by bone collagen, $[\alpha 1(I)]_2\alpha 2$. The process of calcification in progress during these collagen stages is still, however, not fully understood.

Conclusion

This summary survey of amino acid sequence data on proteins suggests some ways in which such work might lead to some very profitable palaeogenetic conclusions if systematically pursued on a large enough scale. The amino acid sequence, which dictates its spatial configuration and hence its function, is one of the basic biological parameters of a protein, and hence of the organism of which if forms an essential working part. This alone should be good enough reason for assembling such data on a meaningful scale. But the phylogenetic content of the data gives a strong added reason for pursuing such work, now with automated sequencers less laborious than it was. It is true that we should like to see also an accumulating body of DNA base-pair sequence data, but the information packed therein is so condensed and complex, and the procedures so difficult of interpretation, that for a long time to come we shall need the amino acid sequence data to help with this interpretation. In the meantime the amino acid sequence data can serve a valuable purpose in studying the history of domestication and selective breeding.

Table 2. Amino acid compositions of CNBr peptides from collagens; non-helical region near NH_2 terminus

	α1(I)-CB1				α2-CB1			
	chick		rat		chick		rat	
	bone	skin	bone	skin tendon	bone	skin	bone	skin
Asp	1	1	1	1	2	2	1	1
Ser	2	2	3	3	1	1	2	2
Glu	1	1	1	1	1	1	1	1
Pro	2	2	2	2	3	3	2	2
Gly	3	3	3	3	2	2	3	3
Ala	2	2	1	1	2	2	1	1
Val	2	2	2	2	0	0	1	1
Tyr	2	2	2	2	1	1	1	1
Hylys	0.5	0	1	0	0.5	0	1	0
Lys	0.5	1	0	1	0.5	1	0	1
HomoSer	1	1	1	1	1	1	1	1
total	17	17	17	17	15	15	14	14

sequences α2 CB1 of skin collagens

chick PCA-Tyr-Asp-Pro-Ser-Lys-Ala-Ala-Asp-Phe-Gly-Pro-Gly-Pro-Met —

rat PCA-Tyr — Ser-Asp-Lys-Gly-Val-Ser-Ala-Gly-Pro-Gly-Pro-Met —

References (Jope)

Adams, H. R., Boyd, E. M., Wilson, J. B., Miller, A. & Huisman, T. H. J. 1968 The structure of goat hemoglobins. III. *Archs Biochem. Biophys.* **127**, 398–405.

Adams, H. R., Wrightstone, R. N., Miller, A. & Huisman, T. H. J. 1969 *Archs Biochem. Biophys.* **132**, 223–236.

Agar, N. S., Evans, J. V. & Roberts, J. 1972 Red cell potassium and hemoglobin polymorphism in sheep. *Anim. Breed. Abstr.* **40**, 407–436.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Astbury, W. T. 1939 X-ray studies of compounds of biological interest. *Ann. Rev. Biochem.* 113–132.
Astbury, W. T. & Woods, H. J. 1933 X-ray studies of the structure of hair, wool, and related fibres. II. The molecular structure and elastic properties of hair keratin. *Phil. Trans. R. Soc. Lond.* A **232**, 333–394.
Bailey, A. J., Robbins, S. P. & Balian, G. 1974 Biological significance of the intermolecular crosslinks of collagen. *Nature, Lond.* **251**, 105–109.
Bornstein, P. 1974*a* The biosynthesis of collagen. *A. Rev. Biochem.* **43**, 567–603.
Bornstein, P. 1974*b* The structure and assembly of procollagen – a review. *J. Supramolecular Structure* **2**, 108–120.
Bornstein, P. & Nesse, R. 1970 The comparative biochemistry of collagen. *Archs Biochem. Biophys.* **138**, 443–450.
Boulter, D. 1972 Protein structure in relation to the evolution of higher plants. *Sci. Prog. Oxf.* **60**, 217–229.
Boulter, D. 1974 The evolution of plant proteins with special reference to higher plant cytochromes C. *Current Adv. Plant Sci.* 1–16.
Boulter, D., Derbyshire, E., Jope, E. M. & Wheat, J. B. 1976 In the press.
Croft, L. R. 1973 *Handbook of protein sequences.* Oxfordshire: Eynshem
Chirpich, T. P. 1975 Rates of protein evolution: a function of amino acid composition. *Science, N.Y.* **188**, 1022–1023.
Chung, E. & Miller, E. J. 1974 Collagen polymorphisms: characterisation of molecules with the chain composition $[\alpha I(III)]_3$ in human tissues. *Science, N.Y.* **183**, 1200–1201.
Dayhoff, M. O. 1972 *Atlas of protein sequence and structure*, 5 (National Biomedical Research Foundation, Silver Spring, Maryland 20901).
Dickerson, R. E. 1971 The structure of cytochrome *c* and the rates of molecular evolution. *J. molec. Evolution* **1**, 26–45.
Dickerson, R. E. & Geis, I. 1969 *The structure and action of proteins.* Menlo Park, California: W. A. Benjamin Inc.
Edman, P. & Begg, G. 1967 A protein sequenator. *Eur. J. Biochem.* **1**, 80–91.
Fietzek, P. P., Furthmayr, H. & Kühn, K. 1974 Comparative sequence studies on α2-CB2 from calf, human, rabbit and pig-skin collagen. *Eur. J. Biochem.* **47**, 257–261.
Fraser, R. D. B. 19 Keratins. *Sci. Am.* **221**, 86–96.
Furthmayr, H. & Timpl, R. 1971 Characterizations of collagen peptides by sodium dodecylsulfate-polyacrylamide electrophoresis. *Anal. Biochem.* **41**, 510–516.
Galinat, W. C. 1971 The origin of maize. *Ann. Rev. Genetics* **5**, 447–478.
Harlan, J. R. 1975 Our vanishing genetic resources. *Science, N.Y.* **188**, 618–621.
Henson, J. L. 1974 *Cotswold Farm Park Rare Breeds Survival Centre* (127 Abbots Road, Abbots Langley, Herts, England).
Huisman, T. H. J., Reynolds, C. A., Dozy, A. M. & Wilson, J. R. 1965 The structure of sheep hemoglobins. *J. biol. Chem.* **240**, 2455–2460.
Huisman, T. H. J., Adams, H. R., Dimmock, M. O., Edwards, W. C. C. & Wilson, J. B. 1967 The structure of goat hemoglobins. I. *J. biol. Chem.* **242**, 2534–2541.
Huisman, T. H. J., Dasher, G. A., Moretz, W. H., Jun, Dozy, A. M., Wilson, J. B. & van Vliet, G. 1968 Studies of haemoglobin types in Barbary Sheep (*Ammotragus lervia*). *Biochem. J.* **107**, 745–751.
Huisman, T. H. J. & Miller, A. 1972 Hemoglobin types in Barbary sheep. *Proc. Soc. exp. Med. Biol.* **140**, 815–819.
Hulmes, D. J. S., Miller, A., Parry, D .A. D., Piez, K. A. & Woodhead-Calloway, J. 1973 Analysis of the primary structure of collagen for the origins of molecular packing. *J. molec. Biol.* **79**, 137–148.
Johnson, B. L. 1972 Seed protein profiles and the origin of the hexaploid wheats. *Am. J. Bot.* **59**, 952–960.
Johnson, B. L. & Hall, O. 1965 Analysis of phylogenetic affinities in the *Tuticinae* by protein electrophoresis. *Am. J. Bot.* **52**, 506–513.
Jope, E. M. 1946 The disappearance of sulphaemoglobin from circulating blood in relation to red cell destruction (Discussion on life and death of the red blood corpuscle). *Proc. R. Soc. Med.* **39**, 755–762.
Jope, E. M. 1965 Man's use of natural resources. *Adv. Sci.* **22**, 1–9.
Jope, H. M. 1969 The protein of brachiopod shell: IV. Shell protein from fossil inarticulates; amino acid composition and disc electrophoresis of fossil articulate shell protein. *Comp. Biochem. Physiol.* **30**, 225–232.
Kohne, D. E. 1968 Taxonomic application of DNA hybridization techniques. In *Chemotaxonomy and serotaxonomy* (ed. J. G. Hawkes), vol. 2, pp. 117–130. Systematics Association. London and New York: Academic Press.
Ladizinsky, G. & Johnson, B. L. 1972 Seed protein homologies and the evolution of polyploidy in *Avena*. *Can. J. Genet. Cytol.* **14**, 875–888.
McGovern, P. T. 1969 Goat and sheep hybrids. *Anim. Breed. Abstr.* **37**, 1–12.
Mangelsdorf, P. C. 1974 *Corn: its origin, evolution and improvement.* Cambridge, Massachussetts: Harvard University Press.
Manwell, C. & Baker, C. M. A. 1970 *Molecular biology and the origin of species.* London: Sidgwick & Jackson.
Manwell, C. & Baker, C. M. A. 1975 Protein polymorphisms in domesticated species: evidence for hybrid origin? In *Proc. Conference Workshop in Population Genetics, Israel*, 1975 (ed. S. Karlin & E. Nevo). London and New York: Academic Press.
Miller, A. & Parry, D. 1973 Structure and packing of microfibrils in collagen. *J. molec. Biol.* **75**, 441–447.
Miller, E. J. 1973 A review of biochemical studies on the genetically distinct collagens of the skeletal system. *Clinic. Orthop.* **92**, 260–280.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Nadler, C. F., Lay, D. M. & Hasslinger, J. D. 1971 Cytogenetic analysis of wild sheep populations in Northern Iran. *Cytogenetics* **10**, 137–152.
Perutz, M. F. 1964 The hemoglobin molecule. *Scient. Am.* **211**, 64–79.
Piez, K. A. & Miller, A. 1974 The structure of collagen fibrils. *J. Supramolec. Struct.* **2**, 121–137.
Rasmusen, B. A. & Tucker, E. M. 1973 Transferrin types and reproduction in sheep. *Animal Blood Groups Biochem. Genet.* **4**, 207–220.
Rauterberg, J. & Kühn, K. 1968 Electron microscopic method of ordering the cyanogen bromide peptides of the α-1 chain from acid soluble calf-skin collagen. *F.E.B.S. Lett.* **1**, 230–232.
Ryder, M. L. 1964 The history of sheep breeds in Britain *Agric. Hist. Rev.* **12**, 1–12, 65–82.
Ryder, M. L. 1968 *Z. Tiersucht. u-Zuchtungsbiol.* **85**, 143–170.
Schroeder, W. A., Shelton, J. R., Shelton, J. B., Robberson, B. & Babin, D. R. 1967 A comparison of amino acid sequences in the β-chains of adult bovine hemoglobins A and B. *Archs Biochem. Biophys.* **120**, 124–135.
Schlusselberg, J. & Debouverie, D. 1975 Study of double and triple simultaneous substitutions of amino acids in cytochromes C. *J. Molec. Evol.* **5**, 187–197.
Scawen, M. D., Ramshaw, J. A. M., Brown, R. H. & Boulter, D. 1974 The amino acid sequence from *Sambucus nigra* L. (Elder). *Eur. J. Biochem.* **44**, 299–303.
Scott, J. E. 1975 Composition and structure of the pericellular environment. *Phil. Trans. R. Soc. Lond.* B **271**, 235–242.
Shemin, D. & Rittenberg, D. 1945 The life-span of red blood cells in man, by N-15 marked glycine. *J. biol. Chem.* **159**, 567–578.
Watson, J. D. 1970 *Molecular biology of the gene*, 2nd edn. New York: W. A. Benjamin Inc.
Watts, R. L. 1970 Proteins and plant phylogeny. In *Phytochemical Phylogeny* (ed. J. B. Harborne), pp. 145–178.
Wilson, J. B., Miller, A. & Huisman, T. H. J. 1970 Production of hemoglobin C in Moufflon (*Ovis musimon* Pallas, 1811) and in Barbary Sheep (*Ammotragus lervia* Pallas, 1777) during experimental anemia. *Biochem. Genetics* **4**, 677–688.
Wrightstone, R. N., Wilson J. B., Miller, A. & Huisman, T. H. J. 1970 The structure of goat hemoglobins. IV. *Archs Biochem. Biophys.* **138**, 451–456.
Wyckoff, R. W. G. 1972 *The biochemistry of animal fossils*. Bristol.
Zohary, D., Harlan, J. B. & Vardi, A. 1969 The wild diploid progenitors of wheat and their breeding value *Euphytica* **18**, 58–65.

Discussion

Contribution to the study of the evolution of animals at the molecular level with particular reference to the domestic fowl.

C. M. Ann Baker and C. Manwell (*Zoology Department, University of Adelaide, South Australia*)

Introduction

Professor Jope's review of the primary structure of proteins has indicated a number of useful approaches to the problems of evolution of dometic animals. We are particularly interested in Professor Jope's consideration of the amino acid sequences of bovine and ovine haemoglobin chains as we discussed these at a meeting only a few weeks ago (Manwell & Baker 1976).

Nearly all intraspecific genetic variants differ by single amino-acid substitutions; by contrast some protein variants of domestic animals differ by several amino acid substitutions. This could indicate hybridization between populations which had diverged considerably from each other (Boyer *et al.* 1966; Manwell & Baker 1970, pp. 295–6, 1975).

In the case of the sheep haemoglobin chains we arranged the data in a somewhat different way (figure 1; see also Manwell & Baker 1976). This suggests that the β chain variants β^{A} and β^{B} came to domestic sheep, *Ovis aries* L., from the hybridization of two taxa which had diverged from the ancestral type. The intermediate sequence of mouflon (*O. musimon* Pallas, 1811) Hbβ^{B} suggests that it is closer to an ancestral ovine Hb chain.

Professor Jope's inclusion of the amino acid sequence of the Hbβ^{B} chain of the Barbary sheep, *Ammotragus lervia* Pallas, 1771, draws attention to the finding that the divergence of *O. aries* Hbβ^{B} from mouflon β^{B} is of the same order as the divergence of Barbary Hbβ^{B}. Thus it is possible that the mouflon β^{B} chain is close to the stem line of the caprini.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

These suggestions need the support of more data from a variety of proteins. Some proteins are more conservative in primary structure than others. Bovine and porcine cytochrome *c* are identical in their primary structure of 104 amino acids, but bovine and porcine pepsinogens differ by 22 out of 362 amino acids (reviewed by Manwell & Baker 1970, pp. 43–44). Anomalous situations may occur. For some time the known sequences of duck and chicken egg-white lysozymes were characterized by a greater similarity to each other – and even to human lysozyme – than to goose ovolysozyme. However, it has been shown recently that avian species have at least two types of lysozyme: lysozyme *c* (orignially found in chicken egg white) and lysozyme *g* (originally found in goose egg white). Both *c* and *g* may occur together in some tissues, but there seems to be a species specific regulatory mechanism determining whether a particular species has one type or both in egg white (Hindenburg, Spitznagel & Arnheim 1974).

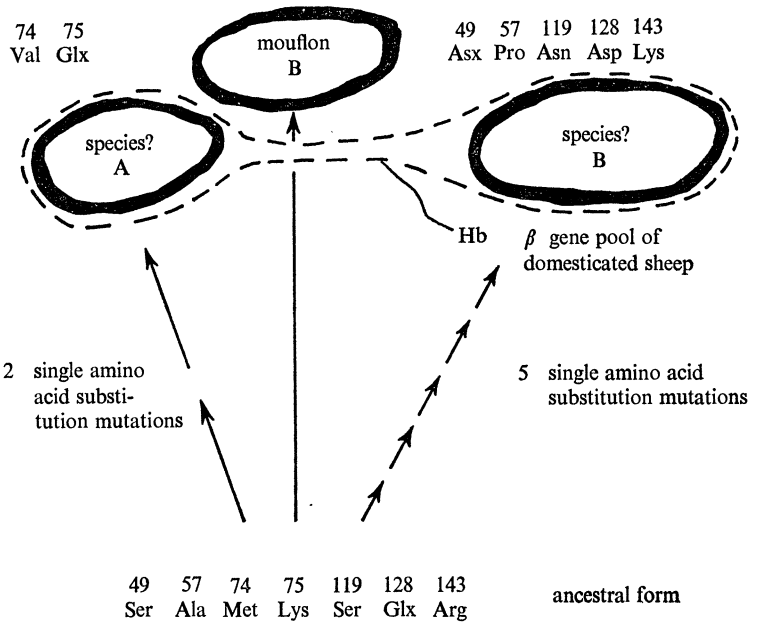

FIGURE 1. A hypothetical model of evolutionary divergence and hybridization for sheep haemoglobin β chain variants A and B. Mouflon B has retained the ancestral ovine sequence in the β chain, whereas in one hypothetical species, A, two amino acid substitutions have occurred, and in another hypothetical species, B, five single amino acid substitution mutations have accumulated. Hybridization of these two species brings the two multiply different haemoglobin cistrons into the present sheep gene pool without any of the intermediates.

Domestic birds: information from primary structure

In spite of the considerable literature on lysozyme there is little information about the primary structure of protein variants of domestic birds. At least three species have ovolysozyme variants: quail, *Coturnix coturnix* (Baker & Manwell 1967); chicken, *Gallus gallus* (Baker & Manwell 1967); chicken, *Gallus gallus* (Baker 1968*a*) and duck, *Anas platyrhynchos* (Prager & Wilson 1971). Yet the differences in primary structure have been determined in detail for only two out of a probable five of the duck variants (Prager & Wilson 1972). The variants studied, II and III, differ by six amino acids (Hermann, Jollès & Jollès 1971).

Similarly, the difference between chicken ovalbumin variants A and B is not known beyond that a single peptide is involved (Fothergill & Fothergill 1970*a*, *b*).

In other cases the number of amino acid differences has been determined but the genetic

control of the protein heterogeneity is not clear. Elleman & Williams (1970) found two amino acid differences between a major and a minor chicken transferrin but it is not certain whether the proteins represent a genetic polymorphism or non-allelic duplicated loci. Delange & Huang (1971) reported a single ambiguous amino acid position involving threonine or isoleucine in chicken egg-white avidin. It is not known if this is due to a previously undetected genetic polymorphism, coding ambiguity, or a recent gene duplication.

Domestic birds: information from protein polymorphisms

The determination of primary structure, or even amino acid composition, is expensive. However, the use of electrophoresis in supporting media such as starch or acrylamide gels provides a relatively cheap and rapid method of screening large numbers of individuals for genetic polymorphisms.

Many species of birds have been domesticated to a varying degree (Zeuner 1963). Some information on protein polymorphism is available for several species (see, for example, Manwell & Baker 1970) but most work concerns the egg-white and serum proteins of the domestic fowl, *Gallus gallus* L.

The general pattern of chicken serum after electrophoresis is typical of many birds and

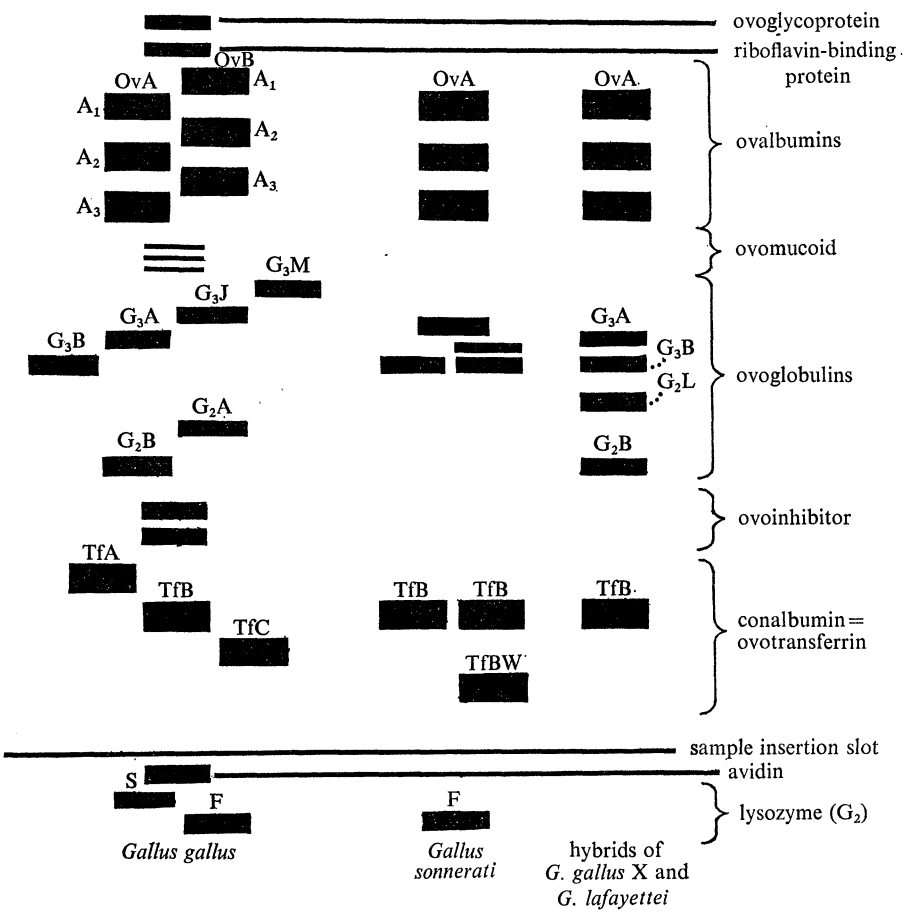

FIGURE 2. Diagram of egg white proteins of *Gallus gallus*, *G. sonnerati* and hybrids of *G. lafayettei* and *G. gallus*. * Identity on the basis of electrophoretic resolution in a variety of buffers.

mammals. Chicken egg white is more complicated. At least 10 proteins are resolved in electrophoresis; four proteins have genetically controlled polymorphism which has been found in many breeds of domestic fowl. For convenience the proteins and their genetic variants are summarized in figure 2 which is based upon information in Baker, Croizier, Stratil & Manwell (1971).

There has been considerable discussion as to whether *G. gallus* was the sole ancestor of the domestic fowl or if one or more of the other species of *Gallus* were involved. So far, no electrophoretic differences have been found between wild and domestic *G. gallus* egg proteins (Baker & Manwell 1972) and other chemical proofs of identity have been found by Feeney and his colleagues (Feeney & Allison 1969).

Sonnerat's jungle fowl, *G. sonnerati*, has egg-white patterns identical to those of *G. gallus* in many respects. Consistent differences have been found for only three proteins: G_3 ovoglobulin, G_2 ovoglobulin and glutamyl peptidase. Hybrids of the two species produce egg white which resolves to give ovoglobulin patterns identical to those from a mixture of equal parts of egg white from the two species. Glutamyl peptidase of hybrids has an electrophoretic mobility intermediate between that of *G. sonnerati* and *G. gallus*. It has not yet been determined if the intermediate position is the result of a single hybrid protein or from overlap of parental zones (Baker & Manwell 1972).

One Sonnerat's jungle fowl had a transferrin variant, Tf Bill Wilkinson, not yet observed in the domestic fowl (Baker 1968*a*). It is possible that the variant may be found in domestic birds: ovoglobulin G_3J, originally found in red Jungle fowl, has since been found in Yokohama, Japanese Bantam and Indian village fowl (Baker, Manwell, Jayaprakash & Francis 1971).

Patterns of egg white from hybrids of *G. gallus* and *G. lafayettei* only differed from typical *G. gallus* patterns in the G_z ovoglobulin region, where the hybrids had a new protein, G_zL. The hybrids also differ from *G. gallus* by a serum albumin variant, C_1 (Stratil 1968*a*, *b*).

G. lafayettei serum amylase has been found to have two variants, A and B, both of which occur in domestic fowl. Serum amylase from *G. sonnerati*, *G. varius* and wild *G. gallus* was of the A type only (Hashiguchi, Yanagida, Maeda & Taketomi 1970).

It must be emphasized that all studies involving wild species of *Gallus* consist of small numbers of samples. Further information comes from the study of populations of domestic fowl (Baker 1968*a*, *b*; Baker *et al.* 1971; Stratil 1970).

As related breeds resemble each other in the occurrence of egg-white protein variants, there is a possibility of tracing the origin of at least some alleles. The most consistent observation is that G_3B seems to be associated with Asiatic breeds and breeds known to have Asiatic stock in their make-up. This brings to mind claims for a species *G. giganteus* Temminck (see, for example, Beebe 1927) or another large, placid *Gallus*.

Many native English and Northern European breeds are homozygous for G_3A. Some older writings have contained suggestions of a European *Gallus*. Evidence for this is slight and the monomorphism of G_3 could be explained more satisfactorily in terms of less polymorphism in peripheral populations. This is supported by the occurrence of G_3J in domestic fowl in India and in Far Eastern countries and in breeds derived from these sources, but not in European breeds.

Of the other commonly observed polymorphic loci, ovalbumin and transferrin, no conclusions can be drawn. The rest of the polymorphic loci for egg and serum proteins have not yet been studied in sufficient breeds to reveal any ancestral associations.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

It is of particular interest that the frequencies of alleles at polymorphic loci vary considerably: some populations are completely monomorphic, some polymorphic at one or two loci, and others are highly polymorphic (Baker 1968*a*, *b*; Stratil 1970). This suggests a selective advantage for different variants and that the advantage may depend upon the environment in which the birds are kept. We agree whole-heartedly with Professor Jope that many 'uneconomic' breeds may be a source of valuable genetic material. For example, Professors Moav and Soller of the Hebrew University of Jerusalem have discovered one such source in Bedouin chickens from the Sinai desert. These birds have great heat tolerance and, when fed conventional laying rations, an egg production equal to the world standard of improved breeds.

We would also agree strongly with Professor Jope's insistence that many more complete amino acid sequences of proteins are necessary as primary biological data.

References

Baker, C. M. A. 1968*a* Molecular genetics of avian proteins. IX. Interspecific and intraspecific variation of egg white proteins of the genus *Gallus*. *Genetics* **58**, 211–226.

Baker, C. M. A. 1968*b* The proteins of egg white. *Egg quality: a study of the hen's egg* (ed. T. C. Carter), pp. 67–108. Edinburgh: Oliver & Boyd.

Baker, C. M. A., Croizier, G., Stratil, A. & Manwell, C. 1970 Identity and nomenclature of some protein polymorphisms of chicken eggs and sera. *Adv. Genet.* **15**, 147–174.

Baker, C. M. A. & Manwell, C. 1967 Molecular genetics of avian proteins. VIII. Egg white proteins of the migratory quail, *Coturnix coturnix* – new concepts of 'hybrid vigour'. *Comp. Biochem. Physiol.* **23**, 21–42.

Baker, C. N. A. & Manwell, C. 1972 Molecular genetics of avian proteins. XI. Egg proteins of *Gallus gallus*, *G. sonnerati* and hybrids. *Anim. Blood Groups Biochem. Genetics* **3**, 101–107.

Baker, C. M. A., Manwell, C., Jayaprakash, N. & Francis, N. 1971 Molecular genetics of avian proteins. X. Egg white protein polymorphism of indigenous Indian chickens. *Comp. Biochem. Physiol.* **40B**, 147–153.

Beebe, W. 1927 *Pheasants: their lives and homes*. London: Robert Hale.

Boyer, S. H., Hathaway, P., Pascasio, F., Orton, C., Bordley, J. & Naughton, M. A. 1966 Haemoglobins in sheep: multiple differences in amino acid sequences of three beta-chains and possible origins. *Science, N.Y.* **153**, 1539–1543.

Delange, R. J. & Huang, T.-S. 1971 Egg white avidin. III. The sequence of the 78-residue middle cyanogen bromide peptide. *J. biol. Chem.* **246**, 686–709.

Elleman, T. C. & Williams, J. 1970 The amino acid sequences of cysteic acid-containing peptides from performic acid-oxidised ovotransferrin. *Biochem. J.* **116**, 515–532.

Feeney, R. E. & Allison, R. G. 1969 *Evolutionary biochemistry of proteins*. New York: Wiley-Interscience.

Fothergill, L. A. & Fothergill, J. E. 1970*a* Thilo and disulphide contents of hen ovalbumin. *Biochem. J.* **116**, 555–561.

Fothergill, L. A. & Fothergill, J. E. 1970*b* Structural comparison of ovalbumin from nine different species. *Europ. J. Biochem.* **17**, 529–532.

Hashiguchi, T., Yanagida, M., Maeda, Y. & Taketomi, M. 1970 Genetical studies on serum amylase in fowls. *Jap. J. Genet.* **45**, 341–349.

Hermann, J., Jollès, J. & Jollès, P. 1971 Multiple forms of duck-egg-white lysozyme. Primary structure of two duck lysozymes. *Eur. J. Biochem.* **24**, 12–17.

Hindenburg, A., Spitznagel, J. & Arnheim, N. 1974 Isozymes of lysozyme in leukocytes and egg-white: evidence for the species-specific control of egg-white lysozyme synthesis. *Proc. Natn. Acad. Sci., U.S.A.* **71**, 1653–1657.

Manwell, C. & Baker, C. M. A. 1970 *Molecular biology and the origin of species*. London: Sidgwick & Jackson.

Manwell, C. & Baker, C. M. A. 1976 Protein polymorphisms in domesticated species: evidence for hybrid origin. In *Proceedings of conference workshop in population genetics, Israel*, 1975 (ed. S. Karlin & E. Nevo). London and New York: Academic Press.

Prager, E. M. & Wilson, A. C. 1971 Multiple lysozymes of duck egg white. *J. biol. Chem.* **246**, 523–530.

Prager, E. M. & Wilson, A. C. 1972 Comparison of multiple duck lysozymes. *Biochem. Genet.* 269–272.

Stratil, A. 1968*a* Transferrin and albumin loci in chickens, *Gallus gallus* L. *Comp. Biochem. Physiol.* **24**, 113–121.

Stratil, A. 1968*b* Genetická fysiologikà a chemická variabilita proteinů slepic. ČSAV – Laborator Fysiologie a Genetiky Živočichů v Liběchove. Doctoral Dissertation.

Stratil, A. 1970 A study on genetic linkage in chickens comprising six loci for polymorphic proteins and three other gene pairs. *Anim. Blood Groups Biochem. Genetics* **1**, 83–88.

Zeuner, F. E. 1963 *A history of domesticated animals*. London: Hutchinson.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

D. Boulter (*Department of Botany, University of Durham*)

Professor Jope dealt mainly with amino acid sequence data from animals, and I have been asked to discuss his paper with particular reference to plants.

Since proteins are a direct expression of the genetic information, their use avoids some of the problems which arise in establishing evolutionary relationships from comparisons of morphological and anatomical characters. However, evolutionary changes in plants may occur by selection of variants which have arisen by several different types of change in the genetic material. Thus, changes may take place in the nucleotide sequences of the genes themselves, the number and arrangement of the same genes, and the physical size and mass of the nucleotide material; these changes are referred to as genetic, genotypic and nucleotypic, respectively. Comparisons of amino acid sequences of homologous proteins only reflect genetic changes, and views differ as to the relative importance of the contribution of these processes to the morphogenetic changes which occurred during the early history of our agricultural crop plants. The recent finding of Prager & Wilson (1975) that the rate of morphological change is not necessarily paralleled by the rate of change of some metabolic proteins, may be of significance in this connection.

Professor Jope did not stress the logistic limitations in the use of amino acid sequence data. Present methods require about 5 mg of pure protein in order to determine a sequence and normally, depending upon the protein, it is necessary to extract 1–10 kg of plant material to obtain 5 mg of pure protein. In the future, new methods, such as field desorption techniques in mass spectrometry, may reduce the amount of protein needed and the time required for analysis; modern sequence methods involve mainly automated techniques.

Professor Jope discussed the extensive data from animal proteins including cytochromes *c* and haemoglobins; similarly, there are homologous protein data sets of plastocyanin and cytochrome *c* sequences available for the construction of affinity trees using plant proteins (Boulter 1974). However, evolutionary information from these data will not be discussed here, since the time-scale for the rates of change of proteins such as cytochrome *c* and plastocyanin, are of the order of millions of years for a single amino acid change. Consequently, these proteins evolve too slowly to be of use in helping to unravel the early history of agricultural plants. For example, the number of amino acid sequence differences between wheat and barley is 3, and between maize and wheat is 8 (D. L. Richardson, unpublished data).

However, it may be possible to determine the progenitors of agricultural plants by examining the proteins of their primitive, wild relatives; for example, the origins of the hexaploid wheats may be established from investigations of possible ancestors among the diploid and tetraploid wheats of the Middle East.

Whether or not amino acid sequence comparisons can be used for this purpose, will depend upon the identification of rapidly evolving proteins. It appears likely from investigations such as those of Antonovics, Bradshaw & Turner (1971) that such proteins may exist. They examined populations of plants adapted to tolerance of heavy metals and compared them with adjacent, related populations living on normal soils. These investigations indicate that there may be proteins whose rate of change is sufficiently rapid that, over the course of many generations, differences in their properties can be detected.

Johnson (1972*a*, *b*) has already used acrylamide gel electrophoretic techniques to provide seed protein profiles, in an attempt to elucidate the origin of the hexaploid wheats. He showed that all sub-species of *Triticum aestivum* have a uniform protein profile which could be closely simulated by the pattern produced from a 2:1 protein mixture of the specific profile types of

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

the ancient tetraploid cultivar, *T. dicoccum*, and the wild diploid, *Aegilops squarrosa*. He has also given important evidence from protein profiles as to the possible source of the B genome.

Generally, the protein profile of a hybrid is the summation of those of its two parents, and usually about twenty protein staining bands appear on the gel, these belonging to the major proteins. Since there are very many different proteins in plants, a disadvantage of the technique is that the homology of protein bands from different taxa is not always certain, although this problem can be overcome, to some extent, by using specific enzyme stains (Lawrence 1964; Wilkinson 1965), or by Edman degradation of proteins extracted from the gel (Weiner, Platt & Weber 1972). Even so, since the storage proteins of seeds are made up of several subunits, polymorphism may occur, which further complicates interpretation of gel patterns. Compared with amino acid sequence determinations, very little material is required for gel electrophoresis i.e. of the order of 200 μg of protein; in fact, a single seed provides enough material.

It is unlikely that sufficient material will be available from archaeological sites to allow the determination of complete amino acid sequences, but other methods such as gel electrophoresis, which use small amounts of material, may be applicable. Even so, success will depend on whether or not material from archaeological sources, such as carbonized material, seeds from bogs and graves, etc., contains relatively unchanged proteins, and little information is available. It is interesting to note that Palozzo & Jaffé (1965) have shown that *Phaseolus* seeds from Indian graves, more than 1000 years old, have detectable protein.

In evolutionary studies, when ancestors are predicted from comparisons of present-day plants, the veracity of the predictions can only be established with certainty from fossil datings. Similarly, in studies of the origin of crop plants, when progenitors are predicted from comparisons of present-day plants, the accuracy of these predictions can only be established with certainty from plant material of dated origin from archaeological sites. While most of the discussion has been on the ways in which molecular approaches may be useful in archaeological studies, this last point illustrates how archaeological findings are important in some biological studies.

Lastly, the methods used to process amino acid sequence data, especially those based on the computation of ancestral amino acid sequences (Dayhoff & Eck 1966), are of interest to archaeologists since they could be used to construct the time-scale of relationships of patterns, e.g. pottery designs.

References

Antonovics, J., Bradshaw, A. D. & Turner, R. G. 1971 Heavy metal tolerance in plants. *Adv. Ecol. Res.* **7**, 1085–1179.

Boulter, D. 1974 The use of amino acid sequence data in the classification of higher plants. *Nobel* 25, *Chemistry in botanical classification*. Uppsala: Almqvist & Wiksell.

Dayhoff, M. O. & Eck, R. V. 1966 *Atlas of protein sequence and structure*, vol. 2. Maryland: National Biomedical Research Foundation.

Johnson, B. L. 1972*a* Seed protein profiles and the origin of the hexaploid wheats. *Am. J. Bot.* **59**, 952–960.

Johnson, B. L. 1972*b* The origin of the B genome of wheat. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1398–1402.

Lawrence, S. H. 1964 *The Zymogram in clinical medicine*. Springfield, Illinois: Thomas.

Palozzo, A. & Jaffé, W. T. 1965 Reacciones inmunológicas de extractos de semillas de habichuelas recuperadas de tumbas prehistóricas en el Perú. *Bol. Soc. Quim. de Peru* **31**, 1–5.

Prager, E. M. & Wilson, A. C. 1975 Slow evolutionary loss of the potential for interspecific hybridization in birds: a manifestation of slow regulatory evolution. *Proc. natn. Acad. Sci., U.S.A.* **72**, 200–204.

Weiner, A. M., Platt, T. & Weber, K. 1972 Amino acid terminal sequence analysis of proteins purified on a nanomole scale by gel electrophoresis. *J. biol. Chem.* **247**, 3242–3251.

Wilkinson, J. H. 1965 *Isoenzymes*. London: Spon.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

D. J. S. Hulmes and A. Miller (*Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford OX1 3PS*)

Professor Jope and Professor Boulter have both discussed the value of studying genetic variation at the molecular level. One of the main problems though is the relatively long time it takes to determine the protein amino acid sequences. The sequence of the so-called $\alpha 1$ chain of rat/calf skin collagen, for example, took at least five years to determine (see Hulmes *et al.* 1973).

The collagen sequence contains 1052 amino acid residues and, for most of its length, the smallest amino acid, glycine, occurs in every third position. This must be so as only glycine can be accommodated in the core of the triple helical molecule. At each end of the sequence, however, glycine does not occur in every third position and there is probably considerable inter-species sequence variation in these regions (Bornstein & Nesse 1970), though there is probably some variation in the non-glycine residues of the 'helical' region. To detect such variation by conventional chemical sequencing methods would be a very slow and laborious business. Recently, however, our group in Oxford have developed a relatively quick way of visualizing the sequence, at least in part, through an understanding of the characteristic 'banding pattern' of collagen seen in the electron microscope.

The banding pattern is obtained when a heavy metal stain (e.g. phosphotungstic acid or uranyl acetate) is applied to the approximately 100 nm wide 'fibrils' which, when packed together in bundles, constitute a tendon. The result is a pattern of darkly stained bands perpendicular to the length of the fibrils which repeats every 67 nm. The significance of this is that, since the stain is only attached to those amino acid residues having a net electrical charge, the band pattern shows the distribution of charged residues characteristic of the sequence and arrangement of the collagen molecules in the fibril. The molecular arrangement is, in fact, sufficiently well understood to make it possible to simulate the banding pattern of rat tail tendon on the basis of the sequence (Doyle *et al.* 1974). The pattern thus represents an interpretable 'fingerprint' of the amino acid sequence.

There are at least two possible uses of this technique in archaeological studies:

(1) If the banding pattern can be observed in sufficient detail to differentiate between collagen from different animal species one may have a quick technique for identifying the source of even very small fragments of animal remains. The technique may be improved by the use of another procedure, so-called 'negative staining', which is very sensitive to inter-species or inter-tissue differences in molecular length.

(2) It would be possible to study the development of tanning methods by a study of the banding pattern of collagen from various leathers. The tanning process involves dipping hides into solutions containing salts of heavy metals. These solutions would stain collagen like the electron microscope stains, except that particular heavy metals would probably produce a characteristic band pattern. The pattern would therefore be a way of identifying the tanning process.

References

Bornstein, P. & Nesse, R. 1970 *Archs Biochem. Biophys.* **138**, 443.
Doyle, B. B., Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A. & Woodhead-Galloway, J. 1974 *Proc. R. Soc. Lond.* B **187**, 37.
Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. & Woodhead-Galloway, J. 1973 *J. molec Biol.* **79**, 137.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES