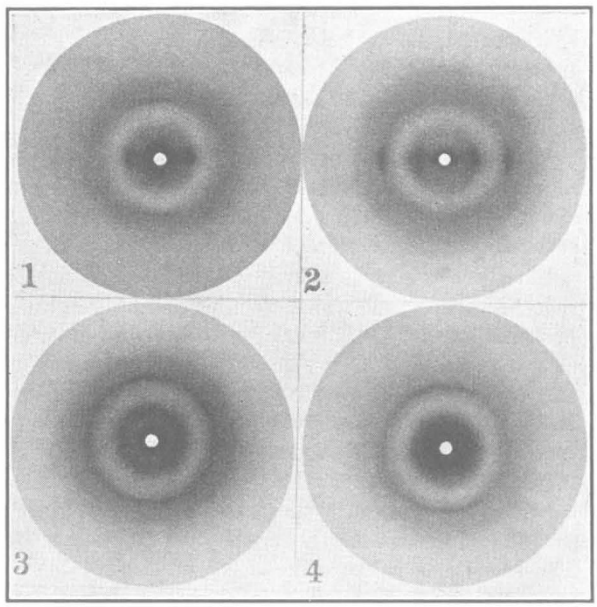
X-Ray Studies of Protein Structure*

By W. T. Astbury, Textile Physics Laboratory, University of Leeds

ACCORDING to the classical researches of Fischer and others, proteins would appear to be essentially polypeptides, giant molecules formed by the repeated condensation of α-amino-acids. This concept leads naturally to the idea of long chain-molecules like that of cellulose, the structure of which was worked out some years ago by a particularly happy combination of chemical and X-ray methods[1]. Similar methods applied to one of the simplest proteins, *fibroin*, the fibre substance of natural silk, show that, for silk at least, the hypothesis is substantially correct[1]; that, in fact, this fibre is a kind of molecular yarn or sliver built up by chain-like molecules, fully-extended polypeptides, lying roughly parallel to the fibre axis. The approximate dimensions of these chain-molecules may be predicted from atomic data already available, and they are found to fit in well with the results of X-ray analysis.

The X-ray photographs of all natural protein fibres, hair, muscle, collagen, feather, etc., show certain features in common with that of natural silk, and it seems clear that they are all built up in very much the same manner from chain-molecules lying along, or simply related to, the fibre axis. But in other respects there are well-marked differences which indicate that the straight-chain configuration found in silk cannot be true in general. Specially does this objection hold in the case of mammalian hair. On stretching hair, however, a new X-ray photograph is obtained which shows again the characteristics of fully-extended polypeptides. The chain-molecules of *keratin*, the protein of hair, must therefore be normally in some regularly folded state from which they can be pulled out straight, and to which they return when the tension is released[2].

Mammalian hairs, spines, horn, etc., are elastic over a range recalling even that of rubber, and *X-rays thus indicate that the property resides in* the keratin molecule itself, which can be transformed from a folded configuration (α-keratin) to the straight-chain configuration (β-keratin) and back again an indefinite number of times. The structure of β-keratin[3] (Fig. 2) appears to be that of a polypeptide 'grid' built up by interactions and combinations between the various 'side-chains' of neighbouring 'main-chains'. To accommodate these interactions the grid buckles, so to speak, in such a way that the main-chains fold in planes transverse to the side-chains. In the presence of water the folds can be pulled out by mechanical force, but they resume their normal configuration when the stretching force is removed. The basis of the 'setting' of hair when steamed in the stretched state is the hydrolytic breakdown of certain of the cross-linkages of the grid which are put under stress when the molecule is stretched. The broken cross-linkages then ultimately re-form in new (unstressed) positions, thereby eliminating the driving force of contraction. If, however, the cross-linkages are broken but not allowed to re-form, a more labile state is induced in which the (modified) keratin molecule can be made to contract to a length even shorter than that of normal α-keratin ('supercontraction').



FIGS. 1–4. X-ray photographs of (1) a protein in the α-configuration, (2) a protein in the β-configuration, (3) disoriented denatured protein, and (4) stretched denatured albumin. (Fibre axis or axis of extension vertical.)

The X-ray photograph of washed and dried muscle (Fig. 1) is remarkably like that of α-keratin[4,5]. The photograph in the main arises from the chief muscle protein, *myosin*[6], which must be presumed to exist normally in a folded, or α-, configuration. If, therefore, the elastic elements in muscle are by analogy with keratin the myosin chain-molecules, the contraction of muscle corresponds to the *'supercontraction' of keratin, and it should be* possible to transform both myosin and muscle into a straight-chain, or β-, configuration[4]. These transformations have now been accomplished[5], by stretching myosin film and washed muscle

* Substance of three lectures delivered at the Royal Institution on February 20 and 27 and March 5.

respectively, and the former has also been made to show 'supercontraction' after the manner of the contraction of muscle itself[5,7]. Figs. (1) and (2) are typical α- and β-photographs, respectively, (1) being that of the foot retractor muscle (washed and dried) of *Mytilus edulis*, and (2) that of stretched horn.

Considering now these two further points: (*a*), that in spite of the close resemblance between their X-ray photographs the sulphur content of myosin is quite small compared with that of keratin, which among proteins is outstandingly rich in sulphur; and (*b*), that the side-chain breakdown shown by X-ray analysis to precede the supercontracting state of keratin has been found by Speakman[8] to be largely concerned with the cystine –S–S– linkage between neighbouring main-chains, we gain the strong impression that hair protein is no other than a kind of muscle protein 'vulcanised' in order to reduce its elastic sensitivity and at the same time impart resistance to chemical attack.

The molecule of feather keratin appears to be in a slightly contracted β-configuration. X-rays show that it can be stretched continuously and reversibly over a range of some seven per cent[9].

Svedberg's investigations with the ultra-centrifuge[10] indicate that the molecules of many soluble proteins are large globular units or combinations of such units, and crystallographic examination and certain X-ray results support this view. Most X-ray photographs, however, taken without any special precautions, suggest almost the antithesis of this and indicate that the molecules either consist of, or generate spontaneously, polypeptide chains configurationally analogous to β-keratin. The idea thus takes shape that most proteins as usually examined by X-rays are in a degenerate state, that their original specific configuration has broken down partially or completely to form disoriented polypeptide chain-bundles[11]. The hypothesis is further strengthened by the observation[11] that when proteins are deliberately 'denatured', for example by heat, the resemblance to disoriented β-keratin becomes even more pronounced. This will be clear by comparing Figs. (2) and (3), the latter being a powder photograph of boiled egg-white—or for that matter, to all intents and purposes, any denatured protein. A crucial test, therefore, is to see whether it is possible to obtain an X-ray photograph like that of *oriented* β-keratin simply by stretching a denatured protein.

This test has now been carried out, and artificial fibres and films of denatured albumins and globulins have been shown to give, on stretching, X-ray photographs typical of oriented bundles of fully-extended polypeptides. The crystallographic orientation, though, of these photographs is not always the same: whereas stretched denatured edestin, for example, gives a photograph like Fig. (2), corresponding to chains lying *along* the axis of extension, a stretched film of 'poached' egg-white gives a photograph like Fig. (4), corresponding to chains lying *across* the axis of extension. The observed types of photograph can be explained only on the assumption that the polypeptide chains in denatured albumins are in general much shorter than those in denatured globulins[12].

It should be noted that films and fibres of denatured proteins are elastic, and often over a great range, presumably because the denaturation process usually results in a random, and maybe sometimes incomplete, liberation of chains which in part coalesce into bundles and in part remain in an irregularly coiled-up state from which they may be pulled out by tension, rather in the manner of the polyprene chains when rubber is stretched.

The immediate question of the future concerns the precise relation or relations between the 'globular' proteins and the chain-molecules to which they give rise so readily. Which is the more fundamental, the globular form or the chain form, and is one obtained from the other by a process of simple coiling or uncoiling, or is the chain form a consequence of the linear polymerisation or condensation of the globular form? At the moment there is evidence for both possibilities[12], though it seems more likely that in general denaturation involves little more than (*a*) the dissolution of intramolecular co-valent linkages, in particular the –S–S– linkage, (*b*) the uncoiling of a chain system, and finally (*c*) the coagulation or 'crystallisation' of the liberated chains into bundles configurationally analogous to those of β-keratin[11]. It must be emphasised, however, that though denaturation appears always to lead to the formation of polypeptide chain-bundles, it does not follow that all chain-bundles are necessarily denatured. The muscle protein, myosin, for example, may be said to be 'configurationally disposed' towards denaturation—and indeed it denatures with extreme ease—but so long as it is not allowed to dry it tends to remain soluble. It is clear that if irreversibility is to be avoided, then at least the chains must be kept from too intimate relations with one another.

What then constitutes reversibility of denaturation, if there is such a thing, and Anson and Mirsky[13] maintain that there is? At the moment, the answer given by X-rays would appear to be this, that it is determined by (*a*) the extent to which a globular protein may be uncoiled reversibly, and (*b*) the possibility of keeping the liberated chains from agglomerating into parallel bundles; or in other words, of avoiding more intimate or widespread interaction than obtained in the original globular configuration. Much seems to

depend on the way the globular proteins are built up in the first place. If they are built up piecemeal and not by the coiling of one or more polypeptide chains, then reversible uncoiling is perhaps unlikely.

A purely formal approach to the constitution of the globular proteins may be made by combining the results of X-ray analysis with the study of protein monolayers. From the X-ray examination of the fibrous proteins we must conclude that in fully-extended polypeptide chains, in β-keratin or denatured edestin, for example, the average length of an amino-acid residue is about 3½ A., its thickness is about 4½ A., and its average width in the side-chain direction is about 10 A. The density of a fibrous or denatured protein, therefore, should be about $0{\cdot}0105R$, where R is the average residue-weight of the amino-acids in question. If we take R to be of the order of 120, this gives a density of 1·26 gm. per c.c.—and it is a fact that proteins do have very much this sort of density. Similarly, protein monolayers formed of parallel arrays of more or less fully-extended polypeptides with their side-chains dipping into the substrate should have a common area (not allowing for hydration) of about $95{\cdot}5/R$ sq. metres per mgm.; that is, for $R=120$, about 0·795 sq. m. per mgm. Now Gorter and his collaborators[14], for example, find always in the region of *p*H 1 an area (extrapolated to zero pressure) of the order of 1 sq. m. per mgm., and a similar area often at the isoelectric point also, whether the molecular weight be 35,000 (Svedberg's 'unit') or a multiple of that; for example, insulin (*p*H 5), 0·875: pepsin (*p*H 2·7), 1·0: zein (*p*H 5·5), 1·07: ovalbumin (*p*H 4·7), 0·88: casein (*p*H 4·6), 1·04 sq. m. per mgm. The natural conclusion is that protein monolayers, under certain conditions at least, are formed by the liberation of polypeptide chains from an originally globular configuration in something the same way as in the process of denaturation.

A further helpful step forward was made by Gorter[15] when he showed that the area of a protein monolayer, under the conditions defined above, corresponds to that of a set of spheres of radius about 22 A., which is the radius found by Svedberg for his spherical units of weight 35,000. (More recently Bernal and Crowfoot, in the only reasonably successful X-ray analyses of unaltered single protein crystals so far accomplished, have arrived at a similar result for the molecules of pepsin and insulin[16].) Gorter's result may also be derived from first principles by means of the X-ray data given above; but it is difficult to proceed to the obvious inference that globular proteins are by way of being simply curved monolayers with the side-chains directed radially, because the calculated density (about 1·14) of such systems is too low.

It is a significant fact that the density of the globular proteins is roughly the same as that of the fibrous proteins. What other possibilities are there then that conform to Gorter's finding? One is the cylinder whose height is equal to its diameter, and whose area, therefore, is again $4\pi r^2$, but the calculated density of this (about 1) is still less satisfying. Actually there is no solution along these lines *except by building a system out of pieces of monolayer separated by the characteristic 'side-chain spacing' found by X-ray analysis. We have thus to place four disks of monolayer of diameter about* 40 *A. on top of one another at a distance of about* 10 *A. apart.* Both the weight and the dimensions then correspond to those of Svedberg's units, the density is correct, and the area of the liberated monolayer would be equal to the surface area of spheres of the same diameter. Furthermore, though X-ray data on the globular proteins are still so meagre, such an arrangement fits in well with present indications[11,12] that their structure must in many cases be somehow closely related to that of the fibrous proteins.

Just recently, Wrinch has arrived at similar conclusions along quite different lines of reasoning[17]. In effect, she has succeeded in generalising the two features of the α–β keratin transformation that it was found necessary to postulate in order to explain quantitatively the experimental facts: (*a*) that the chain should fold hexagonally at regular intervals, and (*b*) that the side-chains should stand out transverse to the plane of folding. Perhaps there is some justification then after all for the suggestion[18] that in a way keratin is the grandfather of all proteins.

[1] K. H. Meyer and H. Mark, "Der Aufbau der hochpolymeren organischen Naturstoffe", 1930.
[2] W. T. Astbury, *J. Soc. Chem. Ind.*, **49**, 441 (1930). *J. Text. Sci.*, **4**, 1 (1931). "Fundamentals of Fibre Structure", 1933. W. T. Astbury and A. Street, *Phil. Trans. Roy. Soc.*, A, **230**, 75 (1931).
[3] W. T. Astbury and H. J. Woods, NATURE, **126**, 913 (1930); *Phil. Trans. Roy. Soc.*, A, **232**, 333 (1933); W. T. Astbury and W. A. Sisson, *Proc. Roy. Soc.*, A, **150**, 533 (1935); H. J. Woods, NATURE, **132**, 709 (1933).
[4] W. T. Astbury, *Trans. Faraday Soc.*, **29**, 193 (1933); Cold Spring Harbor Symposia on Quantitative Biology, **2**, 15 (1934); *Kolloid-Z.*, **69**, 340 (1934).
[5] W. T. Astbury and S. Dickinson, NATURE, **135**, 95, 765 (1935).
[6] G. Boehm and H. H. Weber, *Kolloid-Z.*, **61**, 269 (1932).
[7] H. H. Weber, *Pflüg. Arch.*, **235**, 205 (1934).
[8] J. B. Speakman, NATURE, **132**, 930 (1933); Jubilee Issue of *J. Soc. Dyers and Colourists*, **34** (1934).
[9] W. T. Astbury and T. C. Marwick, NATURE, **130**, 309 (1932); W. T. Astbury, *Trans. Faraday Soc.*, **29**, 206 (1933); *Kolloid-Z.*, **69**, 340 (1934).
[10] See, for example, *Chem. Reviews*, **14**, 1 (1934), and numerous papers in NATURE, *J. Amer. Chem. Soc.*, *Kolloid-Z.*, etc.
[11] W. T. Astbury and R. Lomax, NATURE, **133**, 795 (1934); *J. Chem Soc.*, 846 (1935).
[12] W. T. Astbury, S. Dickinson, and K. Bailey, *Biochem. J.*, **29**, 2351 (1935).
[13] See papers in *J. Gen. Physiol.* over the last ten years.
[14] E. Gorter and F. Grendel, *Trans. Faraday Soc.*, **22**, 477 (1926): *Proc. Kon. Akad. Wetensch.*, **29**, 1262 (1926); *Biochem. Z.*, **201**, 391 (1928); E. Gorter, J. van Ormondt, and F. J. P. Dom, *Proc. Kon Akad. Wetensch.*, **35**, 838 (1932); E. Gorter and J. van Ormondt ibid., **36**, 922 (1933); *Biochem. J.*, **29**, 48 (1935); E. Gorter and G. T Philippi, *Proc. Kon. Akad. Wetensch.*, **37**, 788 (1934); E. Gorter, ibid., **37**, 20 (1934); *Amer. J. Diseases of Children*, **47**, 945 (1934); *J. Gen. Physiol.*, **18**, 421 (1935); E. Gorter and W. A. Seeder, ibid., **18**, 427, (1935); etc.
[15] E. Gorter and F. Grendel, *Proc. Kon. Akad. Wetensch.*, **32**, 770 (1929).
[16] J. D. Bernal and D. Crowfoot, NATURE, **133**, 794 (1934); D Crowfoot, ibid., **135**, 591 (1935).
[17] D. M. Wrinch, NATURE, **137**, 411 (1936).
[18] W. T. Astbury, *Kolloid-Z.*, **69**, 340 (1934); W. T. Astbury and R. Lomax, NATURE, **133**, 795 (1934).