X-*Rays and the Stoicheiometry of the Proteins, with Special Reference to the Structure of the Keratin–Myosin Group.*

A Lecture Delivered before the Chemical Society on November 20th, 1941.

By William Thomas Astbury.

The analysis of molecular structure implies two principal steps: discovering the nature and numbers of the various constituents, and discovering the plan on which they are put together. For the smaller molecules this is a question of relatively few atoms and radicals, but for the giant molecules that shoulder the burden of molecular biology we have hardly yet gone beyond the stage of thinking wistfully in terms of a multitude of sub-molecules—" residues " as they are commonly called, residues of sugars, amino-acids, nucleotides, and the like. Proteins are polypeptide chain systems, alone or in combination with accessory molecules or groupings. These systems are often heavily disguised, it is true, but it is still the considered verdict of most workers in the subject, as it has been since the beginning of the century, that they provide the best and most comprehensive interpretation yet devised. Put more precisely in the light of recent developments, proteins are built from *folded* polypeptide chains cross-linked through combinations and interactions between their side chains. The problem of protein structure, therefore, resolves itself in the first place into discovering the nature and numbers of the amino-acid residues present and their order in the chain or chains, and then above all—for a protein may readily lose its specific properties (become " denatured ") without suffering more than an almost negligible change in chemical constitution —determining the *configuration* of the chain or chains, and what it is that induces or maintains that configuration. The title of this lecture suggests that it relates to the numerical proportions of the amino-acid residues in proteins, and so it does, among other things; but it is not sound structure analysis, nor indeed is it possible, to consider the distribution and properties of these residues apart from the patterns they form; and so in the end we shall be talking about the whole architecture of proteins. To revert to our opening metaphor, the two principal steps in the molecular analysis of the proteins merge into one great steep climb.

In some ways we know more now about the general plan underlying protein architecture than about their exact chemical constitution. That they are most beautifully planned, and frequently perhaps down to the last atom, is undeniable when we contemplate the elaborate *X*-ray diffraction patterns of feather (Fig. 1) or porcupine quill (Figs. 5 and 6) for instance, or recall the facts of immunology and admire anew the ultra-centrifugal studies of Svedberg and his collaborators, or even when we sift the seemingly discordant chemical data, if we put aside unworthy doubts and do not look for the impossible. The major trouble at the moment, though, does rest with the chemical data, partly on account of inadequate or laborious technique, but also on account of analytical variations well outside the limits of experimental error. Certain protein analysts have maintained that the proportions of the amino-acids in proteins often approximate to simple fractions involving 2's and 3's, and Bergmann and Niemann [1] have put forward the hypothesis that the numbers of the various residues, and also the total number of residues in the molecule, are always expressible in the form 2^n3^m, but there are other biochemists who feel that the numerical agreements claimed are fortuitous, resulting from inaccurate analyses. Neither of these two extreme views can be wholly correct and neither wholly incorrect. For consider, *e.g.*, Bergmann's proposed analytical data for gelatin (Table I), from which, in conjunction with *X*-ray data, it has been found possible to deduce the general form of the intramolecular pattern along the polypeptide chains of the great collagen group of protein fibres.[2] Whether or not we accept the smaller estimated yields of

Table I.

Gelatin.

Estimated average residue weight = 94. No. of gram-residues in 100 g. of gelatin = 100/94 = 1·06.

Amino-acid.	Yield, %.	*M.*	G.-mol.	Frequency.
Glycine	25·5	75	0·34	$3\ (2^0 . 3^1)$
Proline	19·7	115	0·17	$6\ (2^1 . 3^1)$
Hydroxyproline	14·4	131	0·11	$9\ (2^0 . 3^2)$
Alanine	8·7	89	0·098	$9\ (2^0 . 3^2)$
Arginine	9·1	174	0·052	$18\ (2^1 . 3^2)$
Leucine–*iso*Leucine	7·1	131	0·054	$18\ (2^1 . 3^2)$
Lysine	5·9	146	0·040	$24\ (2^3 . 3^1)$

amino-acids—and it is obvious that high accuracy is necessary before we may affirm in these cases that the frequency is truly expressible in the form 2^n3^m—there is no such order of difficulty with the larger yields, and it is beyond doubt that almost exactly one-third of the residues in gelatin are glycine residues and nearly another third ($\frac{5}{18}$) are proline and hydroxyproline residues. Similar conclusions hold for

preponderating acids in other proteins: too many examples are known to satisfy convincingly the theory of sheer accident. To anyone with a grain of stoicheiometric faith it is abundantly clear that the proteins are at least trying to behave. The important point is to find out what they are aiming at, and not to devote meticulous pains just yet to the second or third decimal place, or to concentrate always on a few particular amino-acids. It is not as though the situation were new to us—it is not. In its way it is the old problem of the silicates, alloys, and mixed crystals all over again. *X*-Rays have shown that, however complicated the empirical formula for a silicate may be, and some colossal formulæ have been put forward in the past, it is the atom or radical *sites* that count. These always rest on a stoicheiometric basis, and the pattern they form in space is strict and relatively simple. Let us then approach the problém of protein structure from a similar point of view. Let us look for the stoicheiometric ideal, the intramolecular *pattern* on which the details are embroidered.

The problem is to reconcile irregularity of constitution with regularity of structure, and we postulate an intramolecular interchange, or potential interchange, not between individual amino-acid residues, but between certain *types* of residues.[3]

If there is anything at all in this idea, the fibrous proteins that constitute the keratin–myosin group present a striking example. There is no need to detail their properties again: [4] it will be sufficient to recall that this group includes all the mammalian hairs, horn, spines, etc., the fibrous proteins of the epidermis, and also the chief protein of muscle tissue, myosin; *i.e.*, practically all the fibrous proteins we know except those belonging to the collagen group.[2] Moreover, all the members of the group give similar *X*-ray diffraction patterns and have closely related long-range elastic properties. Their normal, unstretched form (α) is built from regularly folded polypeptide chains, whereas their stretched form (β) arises from the same chains pulled out almost straight. When the stretched fibres are let go in the presence of water or other polar molecules, they return spontaneously to the folded form: the basis of their long-range elasticity is an intramolecular, reversible transformation—the β-form, in fact, may be called a " mechanical stereoisomer " of the α-form with an elongation of roughly 100%.

It follows from the common features of their diffraction patterns (how very much alike they are will be evident from Figs. 2a and 2b) and their similar elastic properties that the fibrous molecules of the keratin–myosin group are all built to a common plan, the α-keratin configuration we may call it, that is remarkably persistent as regards its main outlines and dimensions—and this in spite of equally remarkable variations in chemical constitution, ranging from myosin at one extreme, through the epidermal proteins, to keratin at the other. For the purposes of this discussion I have collected together all the available analytical data for wool keratin in Table II, whereas Table III shows less complete data for rabbit myosin.

TABLE II.

Wool keratin.

I, Amino-acid. II, Molecular weight. III, Reported yields from 100 g. of wool. IV, Yield chosen as probably most nearly correct. V, Gram-residues in 100 g. of wool.

I.	II.	III.	IV.	V.
Glycine	75	0·58,[5] 6·5 [7]	6·5	0·0870
Alanine	89	4·13 [5]	4·13	0·0464
Valine	117	2·8,[5] 4·8 [8,m]	4·8	0·0410
Leucine, etc.	131	11·5,[5] 11·3 [8,b,m]	11·3	0·0863
Phenylalanine	165	4·0,[7] 3·75 [8,m]	3·75	0·0227
Proline	115	4·4,[5] 6·8 [8,m]	6·8	0·0591
Methionine	149	0·44—0·66,[9] 0·7 [8,m]	0·7	0·0047
Cystine/2	120	7·3,[5] 13·1,[6,7] 11·9 [12,c,f]	11·9	0·0989
Serine	105	0·1,[5] 10·3 [8,c,m]	10·3	0·0981
Threonine	119	6·4 [8,c,m]	6·4	0·0538
Tyrosine	181	2·9,[5] 4·8,[6] 4·5 [7]	4·65	0·0257
Aspartic acid	133	2·3,[5] 5·84,[12,c] 7·3 [10,c]	6·57	0·0494
Glutamic acid	147	12·9,[5] 15·3 [10,c]	14·1	0·0959
Arginine	174	10·2,[6] 8·7,[7] 10·4 [11]	10·3	0·0592
Lysine	146	2·8,[6] 2·5 [7]	2·65	0·0182
Histidine	155	6·9,[6,a] 0·7 [7]	0·7	0·0045
Tryptophan	204	1·8,[6] 0·7 [7]	1·8	0·0088
Total gram-residues from 100 g. of wool				0·8597 *
Amide nitrogen	14	1·2,[6] 1·37,[c,m,8] 1·37 [10,c] 1·14,[13,c,m,d] 1·12 [13,c,e]	1·13	0·081

Notes.—Numbers refer to literature references.

a, Without doubt excessive. *b*, Martin and Synge's work suggests a leucine : *iso*leucine ratio of about 4 : 1. *c*, Cotswold wool. *m*, Merino wool. *d*, 1·37 value corrected for decomposition of serine. *e*, Mild hydrolysis. *f*, Estimated as total –S–S– and –SH.

* Not, of course, significant to four figures. These have been retained throughout so as to average out in the addition.

FIG. 2*a*.

X-Ray sector comparison photograph of α-horn and α-myosin (M, *myosin ;* H, *horn*).

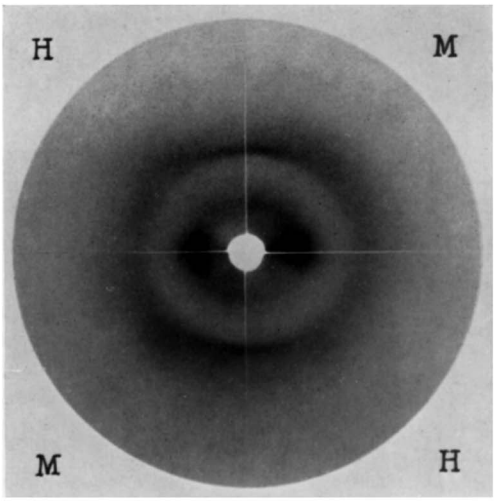
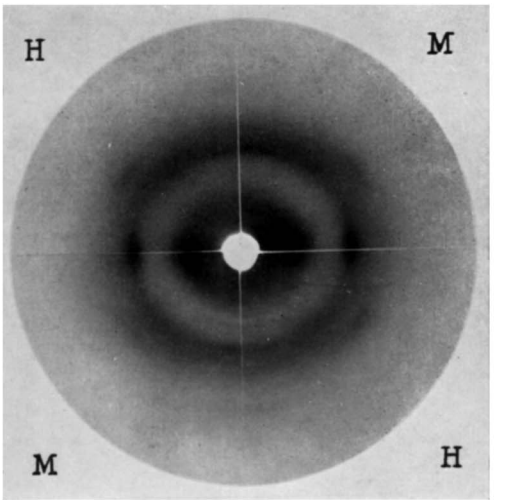


FIG. 2*b*.

X-Ray sector comparison photograph of β-horn and β-myosin (M, *myosin ;* H, *horn*).



Cu-K_α *rays. Film-to-specimen distance* = 4 *cm.*

Each of these figures corresponds to two photographs taken on one and the same film, diagonally opposite quadrants being part of the same photograph. Each meridian or equatorial reflection is thus divided into two halves, one half arising from keratin and the other from myosin.

FIG. 1.

X-Ray "fibre photograph" of the quill of a sea gull's feather ; Cu-K_α *rays ; film-to-specimen distance,* 5 *cm. ; X-ray beam parallel to the surface of the quill.*

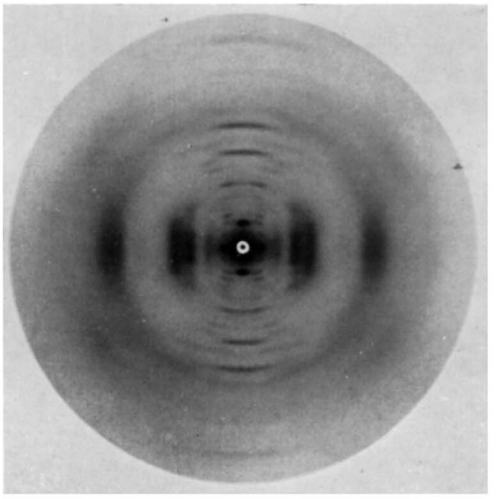
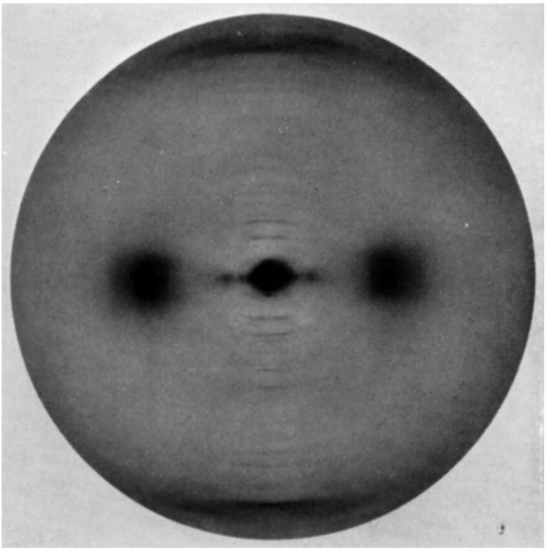


FIG. 5.

X-Ray fibre photograph of the tip end of a porcupine quill, dried in a vacuum over phosphoric oxide. Cu-K_α *rays. Specimen-to-film distance* = 9·7$_9$ *cm. Collimator* 13 × 0·025 *cm.* (*MacArthur.*)



[*To face p.* 338.

TABLE III.

Rabbit myosin.

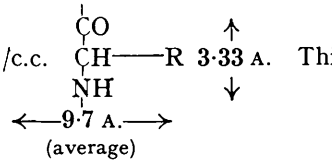
I, Amino-acid. II, Molecular weight. III, Most recently reported, and probably best, yields from 100 g. of myosin. IV, Gram-residues in 100 g. of myosin.

I.	II.	III.	IV.
Cystine/2	120	1·39 [15]	0·0116
(+ cysteine)	(121)		
Methionine	149	3·4 [14]	0·0228
Serine	105	3·57 [8]	0·0340
Threonine	119	3·81 [8]	0·0320
Tyrosine	181	3·4 [14]	0·0188
Aspartic acid	133	8·9 [16]	0·0669
Glutamic acid	147	22·1 [16]	0·1469
Arginine	174	7·0 [16]	0·0402
Lysine	146	10·3 [17]	0·0705
Histidine	155	1·7 [16]	0·0110
Tryptophan	204	0·82 [14]	0·0040
Glycine	75	1·9 [16]	0·0253
Alanine	89	5·1 [17]	0·0573
Valine	117		
Leucine, etc.	131		
Phenylalanine	165 (*b*)	—	0·2412
Proline	115		
Total			0·7825
Amide nitrogen	14	$1\cdot19_3$,[17,a] $1\cdot19_5$ [15,a]	0·0852

Notes.—The yields quoted in the upper half of the table are more reliable than those quoted in the lower half. *a*, Mild hydrolysis. *b*, Sharp's yields of these four acids give together 0·1506 g.-residue, but Martin and Synge,[18] using their new chromatographic method,[19] have estimated for me that these four together with methionine account for probably some 22% of the total nitrogen. From Sharp's value of the total N (16·8%), and allowing for the methionine yield quoted above, this means that valine, leucine, phenylalanine, and proline give altogether 0·2412 g.-residue. Sharp recognises, however, that most of his shortage lies with the monoamino-acids, and the 9% missing even now may reasonably be ascribed chiefly to the glycine and alanine determinations

At first sight it might appear rather hopeless, especially in view of the element of incompleteness and uncertainty, to try to co-ordinate these chemical findings with the *X*-ray data on keratin and myosin, but actually the latter give us at once the average amino-acid residue weight, and thence that important quantity that we must know from the outset, namely, the number of gram-residues in 100 g. of protein. The deduction follows from the average residue dimensions, as inferred from *X*-ray photographs of the fully-extended, or β-, form, and the density, thus:

β-*Keratin.*



Average residue mass = $3\cdot33 \times 9\cdot7 \times 4\cdot65 \times 1\cdot3 \times 10^{-24}$ g. One-sixteenth of the mass of an oxygen atom = $1\cdot65 \times 10^{-24}$ g. Therefore average residue weight = $\dfrac{3\cdot33 \times 9\cdot7 \times 4\cdot65 \times 1\cdot3}{1\cdot65} = 118$ (approx.) and number of gram-residues in 100 g. of keratin = 0·85.

Similarly, for

β-*Myosin*:

Average residue weight = $\dfrac{3\cdot3 \times 9\cdot8 \times 4\cdot65 \times 1\cdot275}{1\cdot65} = 116$ (approx.).

and number of gram-residues in 100 g. of myosin = 0·86.

From these calculations we see that the tale of amino-acid analyses for wool keratin is now on the average reasonably sound and complete: indeed, the grand total of the yields that I have chosen as most plausible amounts to somewhat more than the predicted. It seems best to assume that this error is fairly evenly distributed, and to base our subsequent calculations on the *X*-ray value. For myosin we see that the total yields still amount to only about 91% of the possible maximum, and so *a fortiori* we shall rely on the *X*-ray value. In any case, it must be understood quite definitely that the argument developed in this lecture does not rest at this stage, and is not intended to rest, on mathematically precise chemical analyses. We are trying to discover first the broad molecular plan underlying the keratin–myosin

group. In the end nothing will suffice chemically but complete analyses carried out each *on one and the same specimen of protein.**

The " broad plan " of the keratin–myosin group will have only an approximately constant weight, of course, but it should be founded on always the same number of amino-acid residues, and these should be susceptible of certain ideal distributions that, either singly, in combination, or after residue interchange, reproduce the various members of the group. For several reasons, none of them convincing, it must be granted, if taken alone, but impressive enough when weighed together, the number of residues typical of the group would appear to be either 576 or a multiple of 576. These reasons are as follows :

(1) Given an ideal plan—our fundamental postulate—it is not likely that this is going to be based on a series of mostly incommensurable numbers of residues, especially in molecules of such great size, which *X*-rays show in any case to be possessed of considerable internal regularity of pattern. The chemical analytical indications already noted, the *X*-ray evidence on porcupine quill to be adduced below, the recent stimulating findings of Pacsu [20] on the poly-condensation of certain peptide esters into chains of 3×2^n residues, the genesis of close-packed structures and more generally of the 32 crystallographic point-groups from aggregates involving factors of 2 and 3, all justify us in giving every chance, so to speak, to any attempt on the part of the data to conform to some such numerical system.

(2) The work of Svedberg and his collaborators, and related physicochemical investigations, point to the existence of groups of protein molecules of the same or closely similar architecture; and probably these groups are built in the first place from sub-groups containing at the most 144 ($2^4.3^2$) residues.

(3) Myosin incorporates as part of its more permanent make-up a small proportion of phosphorus. Bate Smith and Davis [21] have reported percentages ranging from 0·04 to 0·06, and Bailey's [12, 22] determinations include 0·043—0·050, 0·054, 0·055, 0·055, 0·048, and 0·067. The percentage corresponding to one atom of phosphorus per 576 residues of average weight 116 is 0·0463.

(4) Table IV, col. I, shows the " frequencies " † of the various amino-acid residues in wool keratin corresponding to the chosen experimental yields listed in Table II. They are remarkably close to the powers of 2 and 3 shown in col. II, which give residue numbers totalling exactly 576 (col. IV). (The residue numbers shown in col. III are derived from the experimental frequencies in col. I, and so add up to rather more than 576 because the total of the chosen experimental yields is a little too high—see above.) In a combination of, or an interchange between, ideal distributions of residues, we should not in general expect such close agreement between experimental and ideal frequencies; there is thus a further implication here, *viz.*, that the structure of the average wool approximates to an ideal distribution at the keratinous end of the keratin–myosin series.

Provisionally, then, we may suppose that the smallest " pattern weight " in wool keratin is about $576 \times 118 = 68{,}000$, and in rabbit myosin $576 \times 116 = 66{,}800$.

We are not yet in a position to draw up for myosin as complete a table as Table IV, but Table V shows the experimental numbers of residues for the more reliable amino-acids in myosin, side by side with corresponding keratin numbers abstracted from Table IV in order to facilitate comparison. In what follows, partly to disarm criticism on the score of making free with debatable Bergmann–Niemann numbers and partly because our argument makes it unnecessary to assume that either wool keratin or rabbit myosin conforms strictly to an ideal distribution, *we shall use only experimental values for both.* Tables IV and V illustrate at a glance the sort of stoicheiometric problem with which we are confronted. We have to correlate this wide range of chemical constitution with a common molecular pattern and similar long-range elastic properties. No simple correlation based on individual amino-acids seems at all possible: something broader is required. We have to apply the principle that the more things there are to be explained, the more fundamental must be the explanation. We must look for some structural property of supreme generality.

Let us turn now to the new α-model,[23] which certainly has this property of generality. It is based on the polypeptide " grid " and the close-packing of side chains, and it was derived from crystallographic, elasticity, and density considerations without reference to any particular side-chain distribution. A satisfactory atomic model has been constructed to scale,[23] but the diagrammatic representation in Fig. 3 will serve our purpose. The important point to notice is that the side chains are packed together in triads that occur first on one side of the main chain and then on the other.

In an attempt to give more purely chemical substance to this model, we first thought [23] that it meant

* It is almost useless to continue to estimate certain amino-acids on one kind of keratin and other acids on another kind; and if it is asked why I seem to have glossed over the question of the variable cystine content in keratin, I can only say that for want of data on *accompanying* variations in the other amino-acids I have done the next best thing, taken a cystine value representative of wool on the average—again, it must be repeated, with the object of trying to deduce the general scheme on which there are so many and so obvious variations.

† A frequency f for any amino-acid means that the residues of that acid amount to $1/f$th of the total residues in the molecule.

TABLE IV.

Wool keratin.

I, Experimental frequencies. II, Possible "ideal" frequencies corresponding to I. III, Experimentally determined approximate numbers of residues referred to a total of 576 ($2^6 . 3^2$). IV, Numbers of residues, totalling exactly 576, corresponding to the "ideal" frequencies given in II.

Amino-acid.	I.	II.	III.	IV.
Glycine	9·8	9	59	64
Alanine	18·3	18	31	32
Valine	20·7	24	28	24
Leucine, etc.	9·9	9	58	64
Phenylalanine	37·4	36	15	16
Proline	14·4	16	40	36
Methionine	181	192	3	3
Cystine/2	8·6	9	67	64
Serine	8·7	9	66	64
Threonine	15·8	16	36	36
Tyrosine	33·1	36	17	16
Tryptophan	96·3	96	6	6
Aspartic acid	17·2	16	34	36
Glutamic acid	8·9	9	65	64
Arginine	14·4	16	40	36
Lysine	46·7	48	12	12
Histidine	188	192	3	3
Totals			580	576
Amides	10·5	—	55	—

TABLE V.

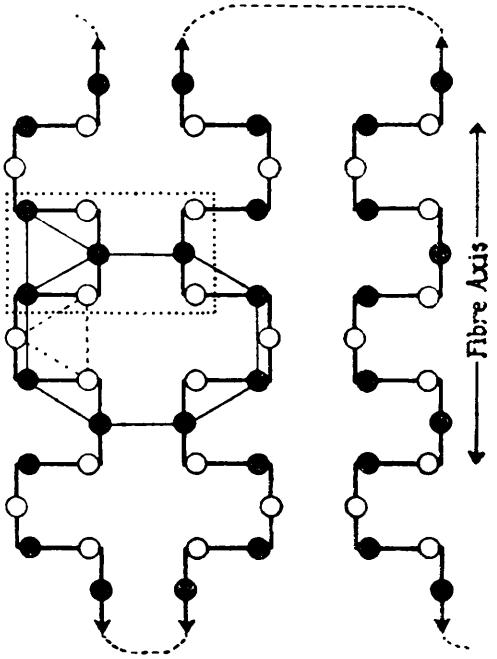
Comparative chemical constitutions of wool keratin and rabbit myosin for the more reliable amino-acids.

Amino-acid.	Observed approximate numbers of residues in 576. Wool keratin.	Rabbit myosin
Cystine/2 + cysteine	67	8
Methionine	3	15
Serine	66	23
Threonine	36	21
Tyrosine	17	13
Tryptophan	6	3
Aspartic acid	34	45
Glutamic acid	65	98
Arginine	40	27
Lysine	12	47
Histidine	3	7
Amides	55	57

that the residues in both keratin and myosin fall into three equal groups, all the members of any one group being either chemically similar (like aspartic and glutamic acids) or probably related biogenetically, and therefore potentially interchangeable. From this point of view each triad of side chains would always include one member of each of the three groups. Further opportunity of examining the problem in the light of new chemical data [18, 24] does not favour this interpretation, however, but suggests rather that the paramount considerations are these: (*a*) The triads are formed from individual side chains that lie alternately on one side and the other of the main chain. (*b*) Martin and Synge [24] find that, in the dipeptides formed on incomplete hydrolysis of wool, the basic amino-acids, at least, are linked with all sorts of non-polar amino-acids—in fact, there seems to be no really outstanding difference between the distribution of acids in the dipeptides and that in the original polypeptide. (*c*) Polar side chains will tend to aggregate with polar side chains, and non-polar with non-polar.

FIG. 3.

Illustrating the packing of side chains in the α-fold of the keratin–myosin group.



— *represents the direction of the main chains;* ● *represents a side chain pointing* up *from the plane of the diagram;* ○ *represents a side chain pointing* down *from the plane of the diagram.*

Suppose now that the great common property linking keratin and myosin is simply this, that *individual side chains are alternately polar and non-polar.* We see then that, not only do we satisfy condition (*b*) and, because of (*a*), bring about (*c*) automatically, but also we endow the structure with potential long-range elasticity! First, though, let us make sure of the corollary that in spite of everything there are in both keratin and myosin approximately equal numbers of polar and non-polar side chains. Table VI shows that this is so—fortunately we can draw up such a table without undue worry about inadequate data on the non-polar side chains, since it happens that the estimates of all the polar side chains are sufficiently good. Theoretically, of course, we might expect 288 side chains of each kind in the total of 576, but we have no right to insist on exact agreement, partly because of analytical errors that have doubtless not yet been eliminated from Table VI, but chiefly because what we have just postulated about the sequence of side chains, though fundamental, may not be the whole story: it may be that the

TABLE VI.

Summation of polar side chains in keratin and myosin (experimental values).

Acid.	Gram-residues. Keratin.	Gram-residues. Myosin.	Nos. of residues. Keratin.	Nos. of residues. Myosin.
Arginine	0·0592	0·0402	40	27
Lysine	0·0182	0·0705	12	47
Histidine	0·0045	0·0110	3	7
Aspartic acid	0·0494	0·0669	34	45
Glutamic acid	0·0959	0·1469	65	98
Serine	0·0981	0·034	66	23
Threonine	0·0538	0·032	36	21
Tyrosine	0·0257	0·0188	17	13
Totals	0·4048	0·4203	273	281
	out of 0·85	out of 0·86	out of 576	out of 576

Note.—In myosin there are also on the average about 2 cysteine residues.[25]

plan is adhered to in general, but certain divergences are permissible in detail, or are essential to serve ends we do not yet appreciate.

As for the explanation the theory offers of the long-range elasticity of the keratin–myosin group, it is easy to raise points of detail to which there are no immediate answers, but nevertheless the new outlook appears more promising in scope than anything that has gone before. It must be repeated that we have to explain similar long-range elastic properties and the same α–β transformation in the face of startling variations in chemical constitution,* and we are suggesting quite broadly that the phenomenon is to be traced throughout the group to a common mode of side-chain aggregation involving the formation of polar triads on one side of the folded main chain and non-polar triads on the other. There is nothing here, be it noted on the one hand, to say that *all* one side of a folded main-chain shall be polar and all the other side non-polar—rather might we expect alternations of polar and non-polar surfaces at intervals decided perhaps by the intervention of proline residues—and there is no question of dispensing with the initial driving force of contraction associated with the stressed side-chain linkages of the β-form,[26] or of minimising the hydrogen bridges between CO– and NH– groups that appear in the atomic model both within and between the α-folds.[23] On the other hand, there is room for an eventual interpretation of the contraction of muscle in terms of increased facilities for aggregation induced by comparatively simple reversible changes in the relative distributions of polar and non-polar side chains, and above all, there is a prospective harmonising of the concepts that myosin is both the working elastic mechanism in muscle and also a principal enzyme (adenosine triphosphatase [27]) in the elastic cycle. By the incorporation of specific reactive groupings among the polar side chains, the characteristic elastic features of the group would rest unimpaired, but the versatility or sensitivity of the system might very well be increased in relation to a particular chemical environment.

We continue with the constitution of the individual polar triads, on which depend in the main the power of aggregation and the strength of inter-chain linkage. Tables IV and V, at the lowest stoicheiometric assessment, discourage any idea of arbitrariness, but quite apart from that there must be some rational scheme of combination or interaction between the polar side chains of neighbouring main chains in the polypeptide grid. Let us consider, then, of what kinds these cross-linkages may be, granted that we have at our disposal only side chains whose end groups are acidic, basic, hydroxyl or amides. Besides salt-like linkages between the bare acidic and the basic side chains, obviously there can be a variety of polar attractions (or hydrogen bridges), and principally the latter will lie between amide and hydroxyl end groups and between hydroxyl and hydroxyl end groups.† Table VII shows the comparative distribution of the different side chain end groups in wool keratin and rabbit myosin, together with a proposed scheme of cross-linkage between opposed polar triads of the polypeptide grid. Fig. 4 illustrates this scheme diagrammatically within the framework of the α-fold as depicted in Fig. 3.

We are converging now, it will be seen, on our original analogy between the stoicheiochemistry of the proteins and that of mixed crystals, such, for example, as the plagioclase felspars, built from continuously varying proportions of albite ($NaAlSi_3O_8$) and anorthite ($CaAl_2Si_2O_8$). Further, just as in these felspars

* It was surprising, too, how many liberties could be taken with the side chains of wool, *e.g.*, deamination of the basic side chains, without seriously upsetting the α–β transformation and the principal elastic properties. All that is now, at least, not surprising. Deamination is in effect no more than the replacement of one polar side chain by another.

† It may be recalled that the strength of the cellulose fibre rests largely on interactions between the hydroxyl groups of neighbouring β-glucose residues.

TABLE VII.

Comparative distribution of polar side chains in keratin and myosin.

	Keratin.		Myosin.	
	Analysis.	Titration.	Analysis.	Titration.
Basic	55	51—55 [28.a].	81	90,[29] 98,[30] 100 [31]
Acidic	44	—	86	87 [31]
Amide	55	—	57	—
Hydroxyl	119	—	57	—

Suggested approximate scheme of cross-linkage between opposed triads in keratin :

1. Acid–base.
2. Amide–hydroxyl (with some replacement by acid–base).
3. Hydroxyl–hydroxyl (with some deficiency probably made up by cystine linkages).

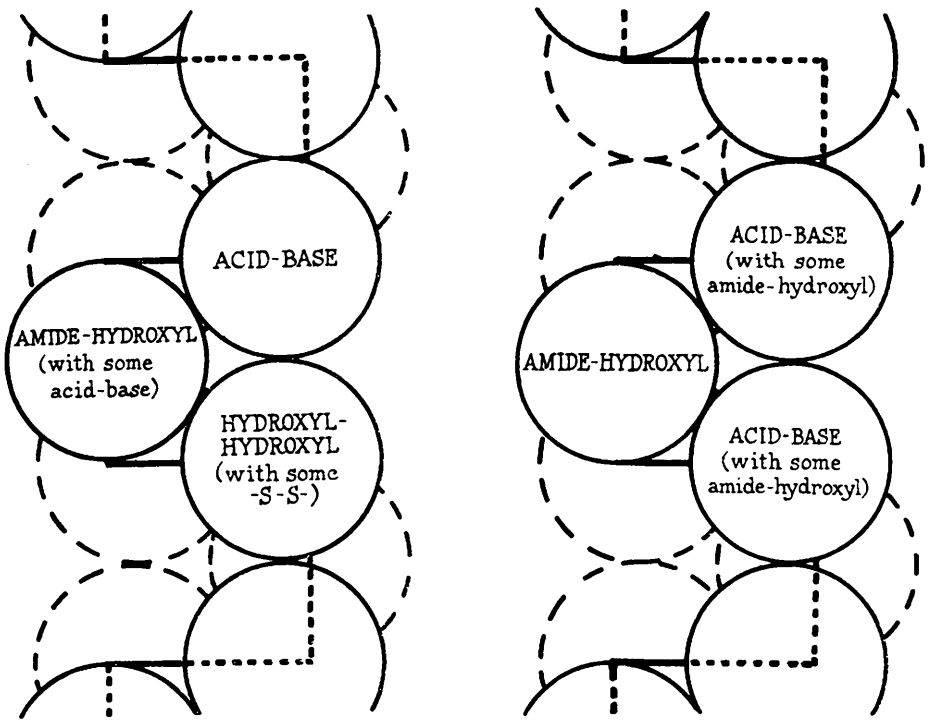
Suggested approximate scheme of cross-linkage between opposed triads in myosin :

1. Acid–base (with some replacement by amide–hydroxyl).
2. Amide–hydroxyl.
3. Acid–base (with some replacement by amide–hydroxyl).

Note (*a*).—A variety of wools and human hair. The best titration value for Cotswold wool [32] is probably 79·6 c.c. of N-HCl per 100 g. of wool, corresponding to approximately 54 basic side chains.

FIG. 4.

Proposed plan for the packing and cross-linkage of the polar side chains in keratin and myosin.



KERATIN. MYOSIN.

we have these two " ideal " structures at each end of the series, so in effect Table VII and Fig. 4 postulate an ideal limiting keratin and an ideal limiting myosin. Of the polar cross-linkages in the limiting keratin, one-third are salt linkages, one-third amide–hydroxyl bridges, and one-third hydroxyl–hydroxyl bridges; whereas in the limiting myosin two-thirds are salt linkages and one-third amide–hydroxyl bridges. Rabbit myosin itself is not far removed from the limiting myosin, which would require 96 acids, 96 bases, 48 amides, and 48 hydroxyls; but the situation with regard to wool keratin is hardly so simple just yet, possibly because the available analyses do not refer all to one and the same specimen of wool. There is a striking apparent deficiency of acids, even with the recent much improved yields, but whether this represents a true unbalance or whether, as suggested by titration * and the evidence for salt linkages advanced by Speakman,[33] it is an erroneous finding, is difficult to decide without more comprehensive experimental data. In this connection, however, it is noteworthy, as pointed out by Speakman,[33] that the proportions of the basic amino-acids vary little in the wools, so it is much less likely that the error—

* The titration curves of keratin and myosin are of a similar type.[29]

if there really is an error—lies with the bases than with the acids. It may be noted also that biochemists seem to be agreed that the amide nitrogen determinations are now fairly sound.

The position of cystine has long excited interest, so much so, in fact, that it has tended to blind protein chemists to the importance of the sixteen or so other amino-acids in wool. It is well known that keratin yields relatively enormous and highly variable quantities of cystine, as opposed to the small and almost constant proportion in myosin, but such knowledge is all but valueless until we know what happens simultaneously to all the other acids. Until we know that, the protein conscience is troubled and confidence is undermined. In the keratin–myosin group there are two ways of looking at the cystine content.[34] The first is to suppose it to consist of two parts, a small and constant proportion of " constitutional " cystine as in myosin, together with a large and variable proportion of " supplementary " cystine incorporated during the process of " keratinisation " to form a kind of cementing substance surrounding chain-bundles whose inner structure remains more or less unchanged. This view has been considered elsewhere,[34] but here we adopt the second, in which the cystine is judged to be always " constitutional," taking its place with the other acids in a molecular plan peculiarly adapted to their mutual interchange. The second hypothesis appears the more satisfying now that fuller chemical analyses are available, for the first does not accord at all well with the revolutionary changes, as we pass from myosin to keratin, in the *relative* proportions of the acids other than cystine, or with the fact that wools of different sulphur content show a roughly constant acid-combining power. It might be argued, of course, that " supplementary " cystine is combined with other acids in the supposed inter-myosin cementing substance; but such a contention begins to encroach on the second hypothesis and might well bring us back again only to the formation of another keratin.

Table VII and Fig. 4 make it clear why myosin is so much more reactive and so much more susceptible to inter-chain (" intra-micellar ") swelling by polar reagents than keratin is. In myosin there are few or no vacancies for disulphide bridges among the polar cross-linkages, but in keratin we are obliged to use some of the cystine to make up the deficiency in the hydroxyl bridges. On this view the rest of the cystine in keratin, and all or practically all of the cystine in myosin, is distributed among the non-polar triads on the other face of the α-fold, where, as a sulphur acid, it is presumably just as acceptable as among the polar triads. It may or may not be of significance that the $-CH_2-S-S-CH_2-$ bridges on the polar face are to be found among the hydroxyl cross-linkages, but it is certainly suggestive that the serine side chain, $-CH_2-OH$, is the one that stands in closest chemical relation to these bridges.

It would not be justifiable at this stage to try to wring from our model any more chemistry, stoicheiometric or otherwise. The interpretation of the keratin–myosin group that we have just examined co-ordinates more facts and is far more promising than anything advanced before, but its chief function at the moment must be to draw ever more attention to the splendid opportunities offered by the group. Above all we need exact chemical analyses, and especially of the keratin of porcupine quill, which gives such excellent *X*-ray fibre photographs. Our earlier photographs of porcupine quill, though they indicated beyond doubt the presence of higher spacings, were like those of horn, for example, shown in Fig. 2; but now MacArthur,[35] working with greatly improved technique, has brought to light more than 100 reflections, ranging in spacing from 1 A. to 100 A., that together build up one of the most perfect protein fibre patterns yet encountered. Fig. 5 is a reproduction of the central region of the porcupine quill photograph. The prominent group of reflections near the top and bottom corresponds to the arcs—the only arcs—seen in Fig. 2, so it will be appreciated how very much more crystallographic information is now at our disposal. To emphasise this point, Fig. 6 is a composite diagram that includes most of the α-keratin pattern as we know it now—it is inspiring, to say the least. Given adequate chemical data, it should be possible to work out this pattern in considerable detail. With such data as are available the task of interpretation is already being proceeded with, of course, and an interim report may be expected shortly.

The perfection of the porcupine quill diffraction diagram is an expression of regularity of structure : this type of keratin, at any rate, is no haphazard affair. The molecular pattern is obviously approaching an " ideal," and surely there will be found a corresponding perfection of stoicheiometry among the constituent amino-acids. May it not be that the keratin of porcupine quill is actually the limiting keratin envisaged in Fig. 4, or something near it? We know little yet of the chemistry of porcupine quill, but there is one experimental fact known that lends colour to this suggestion, and that is that the sulphur content is quite low as compared with that of wool.*

The molecular pattern along the fibre axis of porcupine quill repeats at probably 658 A. (this is a minimum distance to date), and the strong meridian reflection of spacing 5·14 A. that is so characteristic of the keratin–myosin group is the 128th (2^7th) order of that period. If 5·14 A. represents the length of

* Rimington [36] found 1·35% of sulphur for the whole quill, 1·50% for the cortex, and 1·24% for the medulla. These analyses refer to complete quills, whereas the most perfect *X*-ray photographs are given by the tips, the rest of the cortex giving photographs approximating more to those of wool and hair. Possibly, then, the tips will be found to be still more deficient in sulphur.

an intramolecular fold comprising three residues, as we believe, then a single chain of the pattern-length revealed by *X*-rays would contain 384 ($2^7.3$) residues, or a multiple thereof. Chemical analysis indicates 576 ($2^6.3^2$), it will be recalled, but there is no real conflict, not only because both the *X*-ray and the chemical analytical numbers quoted must be considered for the present as the smallest compatible with the experimental data, but also because the chain may well be folded into long loops—it is easy, in fact, to suggest a type of loop that would conform with both numbers. The important thing to notice is that 384 is of the form 2^n3^m; and moreover that *almost all the dominant reflections along the fibre axis of porcupine quill are orders of* 658 A. *that are exactly or nearly of the form* 2^n3^m. These reflections are tabulated in Table VIII, due to MacArthur. In the left half of the table we see that *all* meridian reflections of appreciable strength included between the centre of the photograph and the reflection at 5·14 A. follow the rule closely, while in the right half there are still seven other dominant reflections, of smaller spacing, that show a similar tendency.* It cannot be argued that all this is pure coincidence, for in the crystallographic spacings listed there is none of the wide margin of experimental error overlapping hoped-for results that has been used as an argument against similar chemical agreements.

Here then, from a new and independent quarter, are our 2's and 3's all over again. It is impossible to dismiss them lightly as an example of wishful thinking. What they signify in the case of α-keratin is this, that the full molecular pattern is made up of smaller pseudo-patterns, just as that of the tobacco mosaic virus is,[37] and these pseudo-patterns subdivide the main pattern, sometimes more sharply, sometimes less sharply, into fractions of the form $1/2^n3^m$. It is to be noted that this is not the same thing as the Bergmann–Niemann theory [1] of the way in which the amino-acid residues follow one another along the polypeptide

FIG. 6.

Composite diagram of the α-keratin diffraction pattern given by porcupine quill (Cu-K_α *rays*). *The spacing of the most prominent arc on the meridian is* 5·14$_7$ A. (*MacArthur.*)

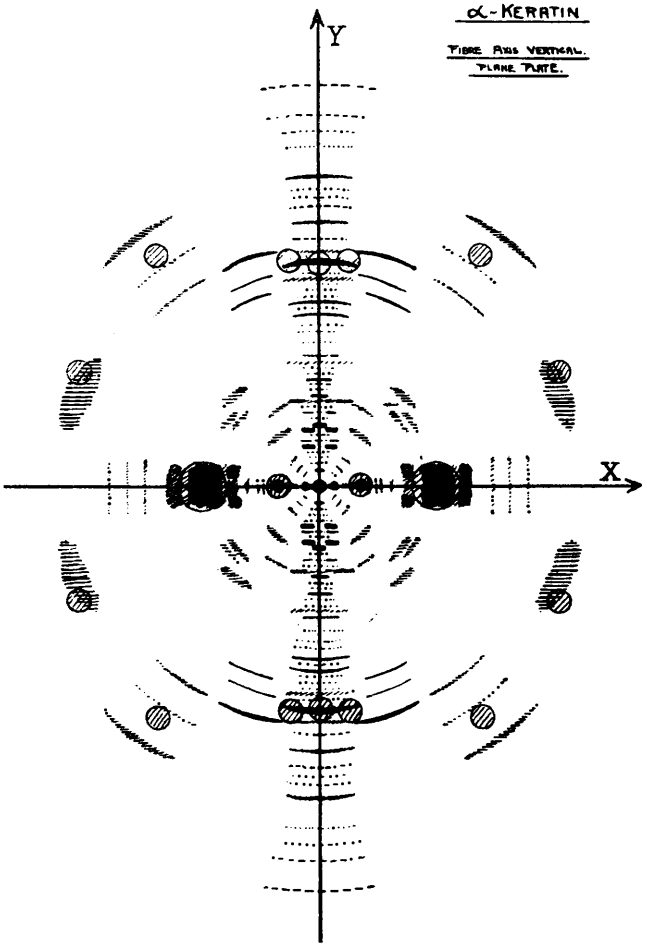


TABLE VIII.

Distribution of the main reflections along the fibre axis in the X-ray diffraction pattern of porcupine quill (*α-keratin*).

All meridian, or very near meridian, reflections ($d \geq 5{\cdot}1$ A.) with intensity $\geq$ W +:				Other dominant meridian reflections:			
Intensity.	Spacing, A.	Order of 658 A.	Nearest 2^n3^m.	Intensity.	Spacing, A.	Order of 658 A.	Nearest 2^n3^m.
M	66·3	10	9	M	4·45	148	144
M	27·8 *	24	24	W−	4·00	164	162
M	$19{\cdot}9_5$ *	33	32	W−	$3{\cdot}41_5$	193	192
WM	18·10	36	36	W+	3·08	214	216
W+	13·11	50	48	v.W	2·55	258	256
M−	12·35	53	54	(?)	2·29	287	288
W+	10·44	63	64	M+	1·49	442	432
M−	6·16	107	108				
Strong	5·14	128	128				

* Layer line spacing.

Intensities: M = moderately strong; W = weak; v.W = very weak.

* We should expect, in any case, the influence of the relative proportions of the residues to die out in reflections of smaller spacing, because the intensities of the latter will be decided more by the arrangement of atoms *within* the residues.

chain; and further, that it raises the status of col. IV, Table IV, to something higher than might be perceived through the gloom of stoicheio-pessimism.

Our last point in this lecture concerns the interpretation to be put upon the large side-spacings (up to 81 A., at least, and possibly including one of 109 A.) found in the *X*-ray diffraction pattern of porcupine quill. Large side spacings are found also in the diffraction diagrams of collagen, feather keratin, and other fibrous proteins, and beyond doubt they are of the very essence of the stoicheiometric problem. Unfortunately, they are mostly harder to resolve in the photographs than the meridian spacings are, and we have no really satisfying idea yet of their precise geometrical relationships. They show that the complete structural unit is built from a number of different, or differently oriented, polypeptide chains lying side by side. It follows from this that we are not restricted to the use of a single chain in our scheme of residue interchange; a similar end can be achieved by using varying proportions of certain types of chain, or of certain types of grid. Each of these types might be characterised by a constant stoicheiometric make-up, perhaps relatively simple, and perhaps based unequivocally on powers of 2 and 3. This approach again reminds us of some of the felspars and their recent *X*-ray interpretation in terms of a system of fine lamellæ.[38] There is much to be said for it, but the time is not yet ripe for a full discussion. I wish merely to stress again the crystallographic analogy, and to show once more how, by combining in this way the *X*-ray and the chemical viewpoint, we may come in time to see our problem in its true light.

That is the prime object of this lecture. I do not mean that I have not weighed the evidence carefully before making any suggestion—on the contrary—but I do mean that so many things are far from clear, and that the present interpretation is no more than a working hypothesis, an attempt to see just how we stand and just how we should proceed in this greatest of stoicheiometric quests. There is a touch of propaganda in this lecture, too, intentional propaganda; but it is of the benevolent kind, directed towards a closer and more understanding co-operation between the physics and chemistry of proteins. This is not the time, I know, in the midst of this war, to call on people for full-time, disinterested research, but it is good to keep the torch of knowledge alight, if only dimly.

I should like to acknowledge my indebtedness and express my thanks to Dr. Florence O. Bell for her unfailing personal help in this and similar investigations, to Dr. I. MacArthur for the use I have made of his invaluable *X*-ray results on the structure of porcupine quill, and to Mr. W. R. Atkin and Drs. K. Bailey, A. J. P. Martin, and R. L. M. Synge for helping me to collect my chemical data and for putting their own results at my disposal. Further, I should like to absolve all chemists of responsibility should it appear to some that I have misused the chemical data.

References.

[1] *J. Biol. Chem.*, 1936, **115**, 77; 1937, **118**, 301; *Science*, 1937, **86**, 187.
[2] Astbury, *J. Intern. Soc. Leather Trades' Chem.*, 1940, **24**, 69 (First Procter Memorial Lecture); Astbury and Bell, *Nature*, 1940, **145**, 421.
[3] Astbury, *ibid.*, 1937, **140**, 968.
[4] See, *e.g.*, Astbury and Dickinson, *Proc. Roy. Soc.*, 1940, *B*, **129**, 307.
[5] Abderhalden and Voitinovici, *Z. physiol. Chem.*, 1907, **52**, 348.
[6] Marston, Council of Scientific and Industrial Research, Commonwealth of Australia, 1928, Bulletin 38.
[7] Block, *J. Biol. Chem.*, 1939, **128**, 181.
[8] Martin and Synge, *Biochem. J.*, 1941, **35**, 294.
[9] Barritt, *ibid.*, 1934, **28**, 1.
[10] Speakman and Townend, *Nature*, 1937, **139**, 411.
[11] Vickery, *J. Biol. Chem.*, 1940, **132**, 325.
[12] Bailey, private communication.
[13] Gordon, Martin, and Synge, *Biochem. J.* (in the press).
[14] Bailey, *Biochem. J.*, 1937, **31**, 1406.
[15] *Idem, ibid.*
[16] Sharp, *ibid.*, 1939, **33**, 679.
[17] *Idem*, private communication.
[18] Private communication.
[19] *Biochem. J.* (in the press).
[20] *Nature*, 1939, **144**, 551; *J. Franklin Inst.*, 1940, **230**, 132.
[21] Rep. Food Investigation Board, D.S.I.R., p. 22.
[22] *Biochem. J.*, 1942, in the press.
[23] Astbury, *Chem. and Ind.*, 1941, **60**, 41, 491 (Jubilee Memorial Lecture); Astbury and Bell, *Nature*, 1941, **147**, 696.
[24] Gordon, Martin, and Synge, *Biochem. J.* (in the press).
[25] Mirsky, *J. Gen. Physiol.*, 1936, **19**, 559; Todrick and Walker, *Biochem. J.*, 1937, **31**, 292; Greenstein and Edsall, *J. Biol. Chem.*, 1940, **133**, 397.
[26] See, *e.g.*, Astbury and Woods, *Phil. Trans.*, 1933, *A*, **232**, 333.

[27] Engelhardt and Ljubimowa, *Nature*, 1939, **144**, 668; Ljubimowa and Engelhardt, *Biochemia*, 1939, **4**, 716; Engelhardt, Ljubimowa, and Meitina, *Compt. rend. Acad. Sci. U.R.S.S.*, 1941, **30**, 644; Needham, *Biochem. J.*, in the press; Bailey, *Biochem. J.*, in the press.
[28] See references in various papers by Speakman and collaborators, and by Harris.
[29] Atkin, private communication.
[30] Bailey, private communication.
[31] Salter, *Proc. Soc. Exp. Biol. Med.*, 1926, **24**, 116.
[32] Speakman and Stott, *Trans. Faraday Soc.*, 1935, **31**, 1425.
[33] *J. Text. Inst.*, 1941, **32**, 3 (Mather Lecture).
[34] Astbury and Dickinson, *Proc. Roy. Soc.*, 1940, *B*, **129**, 968.
[35] *Wool Rev.*, Jan. 1940, p. 9; Reports of the International Wool Secretariat; and in the press.
[36] In Bekker and King, *Biochem. J.*, 1931, **25**, 1077.
[37] Bernal and Fankuchen, *J. Gen. Physiol.*, 1941, **25**, 111.
[38] Chao and Taylor, *Proc. Roy. Soc.*, 1940, *A*, **176**, 76.