

## The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content

BY ROSALIND E. FRANKLIN\* AND R. G. GOSLING

*Wheatstone Physics Laboratory, King's College, London W.C. 2, England*

(Received 6 March 1953)

A qualitative survey has been made of the types of X-ray diagram given by highly orientated specimens of sodium thymonucleate at different humidities. From the nature of the structural changes that occur when the humidity is varied, certain general conclusions have been drawn concerning both the way in which the sodium thymonucleate molecules are bound to one another and the part played by water in the structure.

### Introduction

The fundamental substance of cell nuclei, nucleoprotein, consists of a fairly loose conjugation of desoxyribose nucleic acid (DNA) and simple proteins. The nucleic acid component can be separated from the protein and precipitated in the form of its sodium salt by a mild process (Signer & Schwander, 1949). A knowledge of the structure of DNA would therefore not only be interesting on its own account, but might well prove to be a valuable aid in the study of the nucleoproteins.

DNA is known to exist in the form of long-chain molecules of very high molecular weight; when sufficient precautions to avoid degradation are taken, values up to 8 million are obtained (Signer & Schwander, 1949; Katz, 1952). The chains are believed to be un-branched or nearly so. They are built up of four different nucleotides, each nucleotide being the phosphoric ester of a nucleoside, and each nucleoside the desoxypentose derivative of an amino-purine or amino-pyrimidine. The phosphate group is linked to two nucleotides through the C<sub>3</sub>' and C<sub>5</sub>' positions on the sugar ring (Davidson, 1950).

It was for some time believed that DNA contained the four bases, adenine and guanine (purines) and cytosine and thymine (pyrimidines) in equal proportions, and the suggestion was made that the fundamental unit was a tetranucleotide. However, careful analyses by Chargaff *et al.* (1949) have shown this to be untrue and their results, together with the more recent studies of Sinsheimer & Koerner (1952) and of Markham & Smith (1952) of the fragments obtained by enzymatic digestion of DNA, suggest that the sequence of nucleotides in the chain may be complicated.

Sodium thymonucleate, the sodium salt of the nucleic acid extracted from calf thymus (referred to below as NaDNA), is a fibrous solid which is capable of taking up large quantities of water to form a gel.

\* Present address: Department of Physics, Birkbeck College, London W.C. 1, England.

By means of suitable mechanical treatment and drying, this gel can be used to obtain highly orientated specimens of NaDNA and these give characteristic X-ray fibre-diagrams (Astbury, 1947; Wilkins *et al.*, 1951). The X-ray powder photographs of non-orientated aqueous systems of NaDNA have been studied by Riley & Oster (1951).

None of the X-ray diagrams obtained from NaDNA shows discrete reflexions at angles corresponding to spacings smaller than 2.5 Å. Even a complete quantitative study will not, therefore, be capable of yielding the atomic positions in the structure. But, since the chemical composition and structure of the constituent sugar and base components are well-established, this need not necessarily prove an insuperable obstacle to structure determination. It was felt that, in any case, a comparative study of the different diagrams obtained under different conditions, combined with a detailed quantitative investigation of those diagrams which show the highest degree of crystalline order, should yield a certain amount of new information.

In particular, the diagrams show that the structure is highly sensitive to the humidity of the surrounding atmosphere. A general qualitative survey of this effect is described in the present paper. We believe that it throws some light on the way in which the long-chain molecules are bound to one another in the structure and this, in turn, may possibly have some bearing on the way in which nucleic acid and protein are united in nucleoprotein.

A quantitative study of the fibre diagram given by NaDNA in its most highly ordered state will be described in later papers.

### Preparation of fibres

The material used throughout this investigation was the highly purified sodium salt of calf thymus desoxyribose nucleic acid, which was kindly supplied to us by Prof. Signer.

Fibres were prepared by the method of Wilkins;

sufficient distilled water is added to a small piece of the fibrous solid to form a stiff gel and a needle-point is then placed in the gel and slowly withdrawn. By suitably varying the speed of withdrawal and the water-content of the gel, fibres of diameter from about  $100\ \mu$  to less than  $1\ \mu$  can be obtained at will.

In general, the smaller the diameter of the fibre the greater the degree of orientation obtained.

### Apparatus and method

The first photographs (Wilkins *et al.*, 1951) were taken with a Unicam single-crystal camera and a Raymax tube, using a bundle of about 30 fibres of diameter  $10\text{--}30\ \mu$  prepared as described above. Under suitable conditions (see below) fibre diagrams showing a high degree of crystallinity were obtained. It soon became evident, however, that a photographic system of higher resolving power might be expected to show more fine-structure in the diagram. Moreover, the bundle of fibres could not easily be maintained in perfect parallel alignment, and it was evidently desirable to be able to work with a single fibre of diameter not greater than  $40\ \mu$ .

In order to use such a small specimen and a photographic system of high resolving power, and yet to avoid prohibitive exposure times, a well-collimated micro-beam of high intensity was required. This was provided by an Ehrenberg-Spear fine-focus X-ray tube (1951), used in conjunction with a North-American Philips micro-camera. The tube, which gives a focal spot of about  $40\ \mu$  diameter, was run at  $35\text{--}40$  kV. and  $0.4$  mA. The Philips camera consists essentially of a lead-glass capillary collimator of length  $10$  mm. and bore  $100\ \mu$  or  $50\ \mu$ , and a flat plate carrying the film. The latter can be set at either  $10$  mm. or  $15$  mm. from the end of the collimator.

For most of the work the  $100\ \mu$  collimator and  $15$  mm. specimen-film distance were used, together with nickel-filtered copper radiation. The specimen generally consisted of a single fibre placed directly over the front face of the collimator and fixed with cellulose glue.

When single fibres of  $30\text{--}40\ \mu$  diameter were used, exposure times were of the order of  $50\text{--}100$  hr. For fibres of diameter about  $100\ \mu$ , showing, in general, less good orientation than the finer fibres, exposure times were  $20$  hr. or less.

In using such small specimens of weakly diffracting material it is important to eliminate scattering by air. Throughout each experiment a steady stream of hydrogen was therefore passed through the camera. The moisture content of the specimen was controlled by operating at constant humidity. For this purpose the hydrogen, before entering the camera, was bubbled through a saturated solution of a suitable inorganic salt, a little of the same saturated solution being placed in a small container inside the camera. Each specimen

was left for at least one hour in the camera at constant humidity before starting the exposure.

### Results

Wilkins & Gosling (unpublished) observed that high humidity was required to obtain from the NaDNA fibres a photograph showing a high degree of crystallinity. Attempts to achieve a further increase in crystallinity by working at still higher relative humidity (r.h.) resulted in the observation that at very high humidities a well-defined structural change occurs, leading to a new type of fibre diagram. The two types of diagram are shown in Figs. 1 and 2. The structures represented by these two diagrams will be referred to in what follows as structures *A* and *B*.

Fig. 1 was taken at  $75\%$  r.h. and Fig. 2 at  $92\%$  r.h. It must be emphasized, however, that the type of diagram obtained depends not only on the relative humidity but also on the past history of the specimen. A strong hysteresis is observed, both in the quantity of water taken up by the bulk NaDNA and in the structural change in the fine fibres. A diagram of type *A* may frequently be obtained at  $92\%$  r.h. if the specimen has previously been subjected to prolonged drying over  $P_2O_5$ .

On passing from structure *A* to structure *B* the layer-line spacing increases from  $28\ \text{\AA}$  to  $34\ \text{\AA}$ , and this is accompanied by a similar increase in the macroscopic length of the fibre.

With some samples of calf thymus NaDNA, although it is possible to obtain well-orientated fibres (as shown by optical birefringence and X-ray diagrams), it has so far not been found possible to obtain structure *A*. Fibres from such samples give diagrams resembling Fig. 2 at all r.h. above about  $60\%$ . The failure to obtain structure *A* in these specimens may possibly be due to excessive degradation during extraction of the DNA, or to the presence of some impurity. But, for specimens in which structure *A* can be obtained the change from *A* to *B* is, in general, reversible. When r.h.  $75\%$  is approached from the dry side, structure *A* is obtained, and when r.h.  $92\%$  is approached from the wet side the resulting diagram is always that of structure *B*. Intermediate states consist of a mixture of *A* and *B*. An example of this is shown in Fig. 3.

Here a further reservation must, however, be made. It sometimes happens that a fibre which has been in use in the camera for some days or even weeks, and which has passed through the reversible  $A \rightleftharpoons B$  transformation a number of times, suddenly passes irreversibly to structure *B*. Such a fibre then behaves subsequently like those mentioned above, from which structure *A* cannot be obtained. The photograph shown in Fig. 4 was obtained at r.h.  $75\%$  from a fibre which had passed irreversibly to structure *B*. The difference between Figs. 2 and 4 is probably mainly due to the difficulty of photographing a very wet fibre in a highly

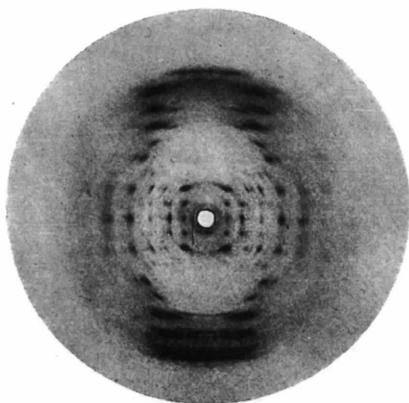


Fig. 1.

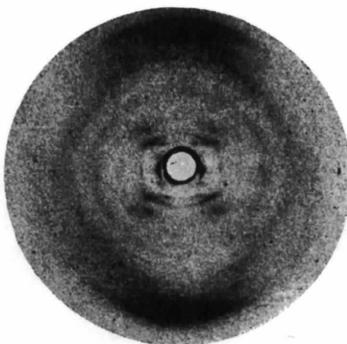


Fig. 2.



Fig. 3.

Fig. 1. Three fibres of NaDNA, diameters 18–30  $\mu$ . Specimen–film distance 15 mm.; exposure 116 hr., r.h. 75%.  
 Fig. 2. Bundle of fine fibres. Specimen–film distance 15 mm.; exposure 6 hr.; r.h. 92%.  
 Fig. 3. Bundle of fine fibres. Specimen–film distance 15 mm.; exposure 3½ hr.; r.h. 84%.

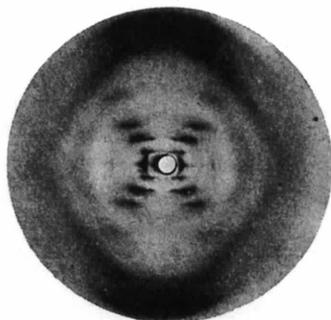


Fig. 4.

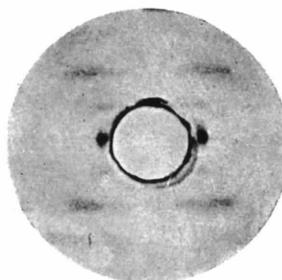
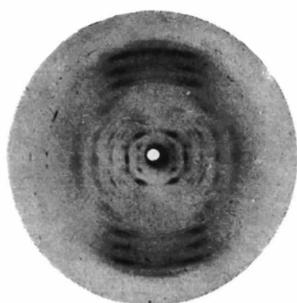


Fig. 5.

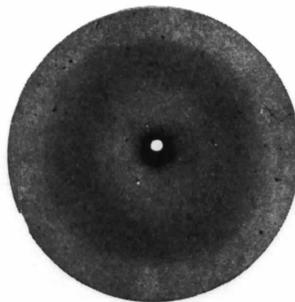
Fig. 4. Single fibre, diameter 50  $\mu$ . Specimen–film distance 15 mm.; exposure 62 hr.; r.h. 75%. This specimen had previously given photographs similar to Fig. 1.  
 Fig. 5. Specimen and conditions as for Fig. 4. Exposure 62 hr.



(a)



(b)



(c)



(d)

Fig. 6. (a) Single fibre, diameter 120  $\mu$ . Specimen–film distance 15 mm.; exposure 22 hr.; r.h. 56%. (b) Specimen as (a). Exposure 19 hr.; r.h. 40%. (c) Specimen as (a). Exposure 24 hr.; dried over  $P_2O_5$  at room temperature. (d) Bundle of about 30 fine fibres. Specimen–film distance 30 mm. (Unicam camera); specimen dried by heating in hydrogen at 50–80° C. before and during exposure.

oriented state. The reversible change  $A \rightarrow B$  is accompanied by a length increase of about 25%. It is therefore necessary to attach the fibre to the collimator at points so close to the collimator hole that it cannot move out of the direct beam. This results in some buckling of the fibre in the wet state, and consequent deterioration of the quality of the photograph. No such difficulty is experienced with fibres which do not give structure  $A$ .

If, with a specimen showing structure  $A$ , the humidity in the camera is reduced appreciably below 75%, the intensity of the reflexions decreases relative to that of the diffuse background. This effect increases progressively with decreasing humidity and is accompanied by a slight lateral shrinkage of the structure. This is illustrated in Fig. 6. After prolonged drying at room temperature only a broad diffuse ring is observed, and after drying at 80° C. even this fades out and there remains only a diffuse scattering at low angles. The change from the crystalline state  $A$  to this disordered state represented in Fig. 6 is wholly reversible.

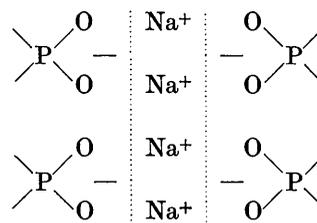
### Discussion

Quantitative work, both on the water uptake in relation to crystallinity and on the structure of the most highly ordered state, is still in progress. Nevertheless, the above qualitative observations of the structural changes that occur when the water content of NaDNA fibres is varied suggest certain general ideas which we tentatively put forward here.

The most highly ordered state is structure  $A$ , obtained at 75% r.h. and clearly possessing a high degree of crystallinity. This state has previously been obtained, in non-orientated form, by Riley & Oster (1951); the spacings reported by them for the 'crystalline' structure in Signer & Schwander's preparation agree to within 0.1 Å with those of strong spots or of the mean values of doublets or triplets in the fibre diagram of structure  $A$ . Riley & Oster also observed that the transition from this state to the 'wet gel' was reversible, but they made no measurements in the high-angle region with 'wet gel' or 'moist crystalline' specimens, and therefore did not observe structure  $B$ .

The equilibrium water content of the bulk sodium thymonucleate at 25° C. and 75% r.h., for a specimen which has previously been thoroughly dried, is about 40–45% of the dry weight. It is not, of course, certain that this represents the water content of the crystallites, but the value is fairly constant for a number of different specimens and is therefore probably similar in the actual fibres examined. There may well be excess water in layers on the surface of the crystallites or in non-crystalline, inter-crystallite regions. 45% is therefore a *maximum* possible value for the structural water of the crystallites in structure  $A$ . This value corresponds to about eight molecules of water per nucleotide.

The most polar part of the NaDNA molecule lies in the phosphate groups, and it is with these that we should expect the water molecules to be associated. We may note, for example, that the compounds  $(C_2H_5O)_2PO_2Na$  and  $(C_3H_7O)_2PO_2Na$ , in which the bonding of the phosphorus is similar to that in nucleic acid, are hygroscopic (Drushel & Felty, 1918). For the same reason we should expect a tendency for the phosphate groups to associate with one another in the structure, just as carboxyl groups are associated in the structures of organic acids and their salts. We should, in fact, expect to find an inter-molecular arrangement of the type



This type of back-to-back bonding of chains through their phosphate groups leads one to expect that, in the crystalline structure, the purine and pyrimidine groups will be linked to similar groups of neighbouring chains by hydrogen bonds.

Thus we may assume that in structure  $A$  the main intermolecular forces are phosphate-phosphate bonds together with hydrogen bonding between base groups. In the gel state and in solution there is evidence for the existence of small stable aggregates of molecules linked by hydrogen bonds between their base groups, and having their phosphate groups exposed to the aqueous medium. In particular, while the phosphate groups in DNA can be titrated in the normal way, the  $-NH_2$  and  $=CO$  groups remain inaccessible unless the pH is taken above 11 or below 5 (Gulland & Jordan, 1947).

Since the change from the crystalline structure to the wet state is readily and rapidly reversible, it seems reasonable to suppose that the small molecular aggregates of the wet state can be easily derived from the grouping existing in the crystal structure. This suggests that in structure  $A$  we may expect to find a small group of chains held together as a unit by hydrogen bonds between their base groups, and these units linked in crystalline array by phosphate-phosphate bonds.

The transformation from the crystalline structure to the molecular aggregates of the wet state, in which the phosphate groups are believed to lie near the outside, must necessarily involve some breaking of phosphate-phosphate bonds. The extreme rapidity with which water is taken up or lost by NaDNA fibres when the humidity of the surrounding atmosphere is changed, the large quantity of water which can be absorbed, and the continuity of the process which leads ultimately to gel formation and solution, all clearly indicate that the phosphate groups with which the water

is associated are in a relatively accessible part of the structure. Since we are dealing with long-chain molecules in which the phosphate groups occur at regular and frequent intervals along the chain, the phosphate-phosphate bonds and associated water molecules will lie on lines or surfaces which extend throughout the structure, and hydration, swelling and solution will result from the distention of these polar regions.

When the water content of NaDNA fibres in state *A* is sufficiently increased, structure *B* always results. It is clear from the X-ray diagrams that structure *B* has a lower degree of order than structure *A*. This suggests that the additional water has in some way weakened the directional property of the phosphate-phosphate link, probably by directly associating itself with the sodium ions and forming a new intermediate in the link. This would make possible some degree of intermolecular displacement in the direction of the fibre axis, as well as an intramolecular rearrangement. Here we may note that the photograph shown in Fig. 4 is strongly characteristic of the type of diagram shown by Cochran, Crick & Vand (1952) to result from a helical structure.

Further support for these general ideas is provided by the photograph shown in Fig. 5. Here the equatorial reflexion at about 24 Å, which is a prominent feature of structure *B*, appears as a well-resolved doublet. The spacings corresponding to this doublet are 22.1 and 24.6 Å. The difference between these spacings, 2.5 Å, is approximately equal to the thickness of a single layer of water molecules, and seems to indicate, therefore, the possible coexistence in the structure of two different hydration states.

The water content of fibres showing structure *B* may vary within wide limits, and it is probable that structure *B* can co-exist with a large quantity of inter-crystallite or inter-micellar water. Increasing the water content beyond the lower limit necessary for the passage from structure *A* to structure *B* results in increased lateral swelling of the fibre, but this is not accompanied by any further appreciable length change.

It is perhaps worth pointing to the analogy which exists between the type of bonding and hydration suggested here and that which is known to occur in the clays (Hendricks & Jefferson, 1938) and in graphitic acid (Hofmann, 1931). In each case a surface rich in -OH or -OM (where *M* is a cation) groups is separated from a neighbouring similar surface by a layer of water only a few molecules thick, the actual thickness of the layer depending on the ambient humidity. The existence of the water layer probably permits some degree of dis-alignment between the neighbouring structural units.

The above ideas seem incompatible with a structure recently proposed for DNA by Pauling & Corey (1953). These authors suggest a three-strand helical structure in which the phosphate groups form a dense core. It is hard to see how the swelling and solution of DNA in water could be explained in terms of such a struc-

ture. Moreover, if such a structure existed in the crystalline state, it would be necessary to assume a radical intermolecular rearrangement in passing from the crystalline to the wet state. For in the wet state one must explain not only the accessibility of the phosphate groups and inaccessibility of the -NH<sub>2</sub> and =CO groups during titration, but also, and far more important, the availability of phosphate groups for interaction with proteins. Since, as we have seen, the transition from the crystalline to the wet state is readily and rapidly reversible, such a radical change in structure seems unlikely.

#### *The influence of drying*

When water is progressively removed from the crystalline structure by drying, the sharp X-ray reflexions gradually fade out and the general diffuse scattering increases. There is no sign of the formation of a new ordered state in the dry substance, and the crystalline state is restored immediately on exposure to a moist atmosphere. Indeed, a more perfect crystallinity in a specimen can frequently be attained by strong drying followed by re-wetting. It seems, therefore, that drying does not break the phosphate-phosphate links but, if anything, cements them more strongly. The removal of water stresses and distorts the structure, destroying its regularity, while leaving the basic three-dimensional skeleton intact. The effect on the X-ray diagrams may be compared with that of strong thermal agitation.

The diffuse low-angle scattering observed after strong drying (Fig. 6(d)) suggests that complete removal of the water results in the formation of holes in the structure.

#### *Sequence of nucleotides*

All X-ray photographs of NaDNA show a rather strong background of diffuse scattering. While it is possible that this is merely due to the presence of small non-crystalline regions separating the crystallites, an alternative explanation might be suggested. We have supposed that phosphate-phosphate bonds are responsible for the intermolecular order in the crystalline state, and that they form a rigid skeleton strong enough not to be disrupted even by the strains produced by intensive drying. It may be that, while the phosphate groups form a truly crystalline array, the base groups of the nucleotides do not, possibly owing to different bases occupying equivalent crystallographic positions.

Wilkins *et al.* (1953) have shown that fibre diagrams similar to that of structure *B* may be obtained from NaDNA from a variety of sources other than calf thymus. In these, the content of the different nucleotides is known to vary considerably. It seems, therefore, that in structure *B* all nucleotides may be crystallographically equivalent.

Structure *A*, on the other hand, has been obtained

only from calf thymus NaDNA\*, and it is natural to enquire, therefore, whether this structure is dependent on a particular nucleotide sequence. Here one obvious difficulty arises. If structure *A* is truly crystalline—that is, if a given position in the unit cell is always occupied by the same nucleotide—then the sequence of nucleotides in calf thymus DNA must be simple and invariable; its repeat unit must lie within the bounds of a single unit cell. In view of the biological properties generally attributed to DNA, this seems unlikely. On the other hand it also seems improbable that purine and pyrimidine groups, which differ from one another considerably in shape and size, could be interchangeable in a structure as highly ordered as structure *A*. A possible solution, therefore, is that in structure *A* cytosine and thymine are interchangeable and adenine and guanine are interchangeable, while a purine and a pyrimidine are not. This is suggested by the remarkably similar crystal structures found by Broomhead (1951) for adenine and guanine hydrochlorides. In this way an infinite variety of nucleotide sequences would be possible, to explain the biological specificity of DNA, while the existence of a crystalline structure in calf thymus NaDNA would be attributed to a particular purine-pyrimidine sequence.

### Conclusions

The structural changes which occur in NaDNA fibres when their water content is varied suggest that the fundamental structural unit is a group of molecules so arranged that the phosphate groups are exposed and accessible to water. This is entirely consistent with the work of Gulland & Jordan (1947), who concluded that the hydrogen-bond-forming =CO and -NH<sub>2</sub> groups, which lie at the opposite end of the nucleotides from the phosphate groups, are inaccessible to water in aqueous solutions of NaDNA at pH between 5.0 and 11.0. Presumably, hydrogen bonds link neighbouring chains through their =CO and -NH<sub>2</sub> groups and form a stable aggregate of molecules which survives as a micelle in aqueous solution.

Such a group of molecules, in which the hydrogen-bonded groups are turned inwards and the phosphate groups outwards, explains the ready availability of phosphate groups for interaction with proteins. Moreover, since both the protein in nucleoproteins and the

water in NaDNA fibres are believed to be associated with the phosphate groups, it is conceivable that the NaDNA structure might exist without great modification in the nucleic acid of nucleoproteins, the protein taking the place of the structural water.

It is hoped that further X-ray work will yield some knowledge of the precise form of the apparently stable group of NaDNA molecules.

The authors are grateful to Prof. J. T. Randall, F. R. S., for his interest and advice, and to M. H. F. Wilkins and A. R. Stokes for discussions.

One of us (R. E. F.) acknowledges the award of a Turner and Newall Fellowship.

### References

- ASTBURY, W. T. (1947). *Cold Spring Harbour Symposium on Quantitative Biology*, **12**, 56.
- BROOMHEAD, J. M. (1951). *Acta Cryst.* **4**, 92.
- CHARGAFF, E., VISCHER, E., DONIGER, R., GREEN, C. & MISANI, F. (1949). *J. Biol. Chem.* **177**, 405.
- COCHRAN, W., CRICK, F. H. C. & VAND, V. (1952). *Acta Cryst.* **5**, 581.
- DAVIDSON, J. N. (1950). *The Biochemistry of the Nucleic Acids*, p. 32. London: Methuen.
- DRUSHEL, W. A. & FELTY, A. R. (1918). *Chem. Zbl.* **89**, 1016.
- EHRENBERG, W. & SPEAR, W. E. (1951). *Proc. Phys. Soc. B*, **64**, 67.
- FRANKLIN, R. E. & GOSLING, R. G. (1953). *Nature, Lond.* **171**, 740.
- GULLAND, J. M. & JORDAN, D. O. (1947). *Cold Spring Harbour Symposium on Quantitative Biology*, **12**, 5.
- HENDRICKS, S. B. & JEFFERSON, M. E. (1938). *Amer. Min.* **23**, 863.
- HOFMANN, U. (1931). *Z. Elektrochem.* **37**, 613.
- KATZ, S. (1952). *J. Amer. Chem. Soc.* **74**, 2238.
- MARKHAM, R. & SMITH, J. D. (1952). *Nature, Lond.* **170**, 120.
- PAULING, L. & COREY, R. B. (1953). *Proc. Nat. Acad. Sci., Wash.* **39**, 84.
- RILEY, D. P. & OSTER, G. (1951). *Biochem. Biophys. Acta*, **7**, 526.
- SIGNER, R. & SCHWANDER, H. (1949). *Helv. chim. Acta*, **32**, 854.
- SINSHEIMER, R. L. & KOERNER, J. F. (1952). *J. Amer. Chem. Soc.* **74**, 283.
- WATSON, J. D. & CRICK, F. H. C. (1953). *Nature, Lond.* **171**, 737.
- WILKINS, M. H. F., GOSLING, R. G. & SEEDS, W. E. (1951). *Nature, Lond.* **167**, 759.
- WILKINS, M. H. F., STOKES, A. R. & WILSON, H. R. (1953). *Nature, Lond.* **171**, 739.

\* The ideas expressed in this paragraph have been somewhat modified in the light of recent work (see Watson & Crick (1953) and Franklin & Gosling (1953)).