cocultures is gap junctional transfer of a communicator of hormonal stimulation. If this hypothesis is supported by further investigation, it is possible that the phenomenon of intercellular communication may have a biologically relevant role in the transmission and regulation of hormonal stimulation in various mammalian organs.

We thank Joan Pesek, Eleana Sphicas and Roberta Morganstern for technical assistance, Asneth Kloesman for help with figures and Madeleine Naylor for secretarial assistance. FSH and LH were provided by the NIAMDD. This work was supported by grants from the Rockefeller Foundation, the NIGMS, the NICHHD, the NHLBI and the Irma T. Hirschl Trust. N.B.G. and W.H.B. are recipients of US PHS Research Career Development Awards.

Received 14 November 1977; accepted 23 January 1978.

1. Furshpan, E. J. & Potter, D. D. *J. Physiol., Lond.* **145,** 289–325 (1959).
2. Kanno, Y. & Loewenstein, W. R. *Nature* **212,** 629–630 (1966).
3. Payton, B. W., Bennett, M. V. L. & Pappas, G. D. *Science* **166,** 1641–1643 (1969).
4. Johnson, R. G. & Sheridan, J. D. *Science* **174,** 717–719 (1971).
5. Weidmann, S. *J. Physiol., Lond.* **187,** 323–342 (1966).
6. Subak-Sharpe, H., Burk, R. R. & Pitts, J. D. *J. Cell Sci.* **4,** 353–367 (1969).
7. Rieske, E., Schubert, P. & Kreutzberg, G. W. *Brain Res.* **84,** 365–382 (1975).
8. Pitts, J. D. & Finbow, M. E. in *Intercellular Communication* (ed. De Mello, W. C.) 61–86 (Plenum, New York, 1977).
9. Simpson, I., Rose, B. & Loewenstein, W. R. *Science* **195,** 294–296 (1977).
10. Gilula, N. B., Reeves, O. R. & Steinbach, A. *Nature* **235,** 262–265 (1972).
11. Gilula, N. B. in *International Cell Biology* (eds Brinkley, B. R. & Porter, K. R.) 61–70 (Rockefeller University Press, New York, 1977).
12. Bennett, M. V. L. in *Handbook of Physiology I. The Nervous System,* Sect. 1 (ed. Kandel, E. R.) 357–416 (Williams and Wilkins, Baltimore, 1977).
13. Loewenstein, W. R. *Ann. N. Y. Acad. Sci.* **137,** 441–472 (1966).
14. Furshpan, E. J. & Potter, D. D. *Curr. Topics dev. Biol.* **3,** 95–127 (1968).
15. Sheridan, J. D. in *The Cell Surface in Development* (ed. Moscona, A. A.) 187–206 (Wiley, New York, 1974).
16. Pitts, J. D. in *International Cell Biology* (eds Brinkley, B. R. & Porter, K. R.) 43–49 (Rockefeller University Press, New York, 1977).
17. Tsien, R. W. & Weingart, R. *J. Physiol., Lond.* **260,** 117–141 (1976).
18. Rall, T. W. & West, T. C. *J. Pharmac. exp. Ther.* **139,** 269–274 (1963).
19. Tsien, R. W., Giles, W. & Greengard, P. *Nature new Biol.* **240,** 181–183 (1972).
20. Reuter, H. *J. Physiol., Lond.* **242,** 429–451 (1974).
21. Kukovetz, W. R., Poch, G. & Wurm, A. *Adv. Cyclic Nucleotides Res.* **5,** 395–414 (1975).
22. Harary, I., Renaud, J.-F., Sato, E. & Wallace, G. A. *Nature* **261,** 60–61 (1976).
23. Moura, A. M. & Simpkins, H. *J. molec. cell. Card.* **7,** 71–77 (1976).
24. Goshima, K. *Expl Cell Res.* **84,** 223–234 (1974).
25. Lane, M. A., Sastre, A. & Salpeter, M. M. *Devl Biol.* **57,** 254–269 (1977).
26. Krause, E. G., Halle, W., Kallabis, E. & Wollenberger, A. *J. molec. cell. Card.* **1,** 1–10 (1970).
27. Goshima, K. *J. molec. cell. Card.* **8,** 713–725 (1976).
28. Strickland, S. & Beers, W. H. *J. biol. Chem.* **251,** 5694–5702 (1976).
29. Beers, W. H. & Strickland, S. in *Novel Aspects of Reproductive Biology* (ed. Spillman, C. H.) (Spectrum, New York, in the press).
30. Beers, W. H., Strickland, S. & Reich, E. *Cell* **6,** 387–394 (1975).
31. Kolena, J. & Channing, C. P. *Endocrinology* **90,** 1543–1550 (1972).
32. Espey, L. L. & Stutts, R. H. *Biol. Reprod.* **6,** 168–175 (1972).
33. Merk, F. B., Botticelli, C. R. & Albright, J. T. *Endocrinology* **90,** 992–1007 (1972).
34. Albertini, D. F. & Anderson, E. *J. Cell Biol.* **63,** 234–250 (1974).
35. Hyde, A. *et al. Prog. Brain Res.* **31,** 283–311 (1969).
36. Goshima, K. *Expl Cell Res.* **65,** 161–169 (1971).
37. Schramm, M., Orly, J., Eimerl, S. & Korner, M. *Nature* **268,** 310–313 (1977).
38. Epstein, M. L. & Gilula, N. B. *J. Cell Biol.* **75,** 769–787 (1977).
39. Unkeless, J. C. *et al. J. exp. Med.* **137,** 85–111 (1973).
40. Kalderon, N., Epstein, M. L. & Gilula, N. B. *J. Cell Biol.* **75,** 788–806 (1977).
41. Pitts, J. D. & Simms, J. W. *Expl Cell Res.* **104,** 153–163 (1977).

# Three-dimensional reconstruction of the fibres of sickle cell haemoglobin

**Gene Dykes, Richard H. Crepeau & Stuart J. Edelstein**

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

*Three-dimensional reconstruction of electron micrographs of the 20-nm diameter fibres of HbS reveals an inner helical core of four strands surrounded by an outer helix of 10 strands to give a total of 14 strands. The strands are arranged with roughly hexagonal packing to produce an unusual helical structure which features a variety of intermolecular contacts and a non-circular cross section.*

SICKLE cell disease has long been recognised[1] and the role of a haemoglobin variant in the disease was specified in 1949 (ref. 2). An active effort to understand the structure of the fibres of haemoglobin S that lead to the symptoms of the disease has occurred only recently. Structural studies[3–6] with X rays and electrons have led to the proposed specific structural models. Two distinct forms of fibres of haemoglobin S have been identified by electron microscopy on negatively stained samples—17-nm diameter fibres with striations perpendicular to the fibre axis[4] and 20-nm diameter fibres with striations diagonal to the fibre axis[5,7–9]. In studies with sickled cells, both classes of fibres were observed, but the 20-nm diameter fibres were found in much greater abundance[7]. In studies on fibres from gelled haemolysates, only the 20-nm diameter fibres have been observed[5,8,9]. It therefore seems that under most conditions the 20-nm diameter fibres are the predominant form. In this report we describe the analysis of these fibres with three-dimensional image reconstruction methods, using Fourier transform techniques[10–12]. For the first time an inner core structure for the fibres is revealed.

Analysis of the 20-nm diameter fibres of haemoglobin S has gone through several stages as the quality of micrographs improved and the analysis progressed from real space measurements[5] to optical transforms[7–9], and now to computer reconstructions. The initial optical transforms revealed meridional reflections at about 1/30 Å, pairs of reflections at about 1/60 Å, and pairs of reflections close to the equator. A splitting of the 1/30 Å reflection was observed with fibres prepared from haemolysates[8,9], and the splitting has also been apparent in fibres from sickled cells. Various forms of polymorphism within the 20-nm diameter fibres have been considered, but we are now inclined to the view that the 20-nm diameter fibres are a homogeneous class, but of such complexity that very long regions of fibres must be examined. This view is strengthened by our work with computed transforms in which a consistent pattern of reciprocal maxima can be obtained from fibres prepared either from sickled cells or gelled haemolysates. The Fourier transforms give a highly complex pattern of reciprocal space maxima with approximately 30 resolvable maxima on 20 layer lines, suggesting that interpretations from the less well resolved optical transforms were incomplete and variable. The Bessel orders to each of the maxima of the transform have now been assigned, permitting three-dimensional reconstructions of the fibres to be carried out. The results of these reconstructions are presented here.

## Images of the fibres and their Fourier transforms

Electron microscopy images of the fibres of haemoglobin S were recorded on samples negatively stained with 2%

0028–0836/78/0272–0506$01.00

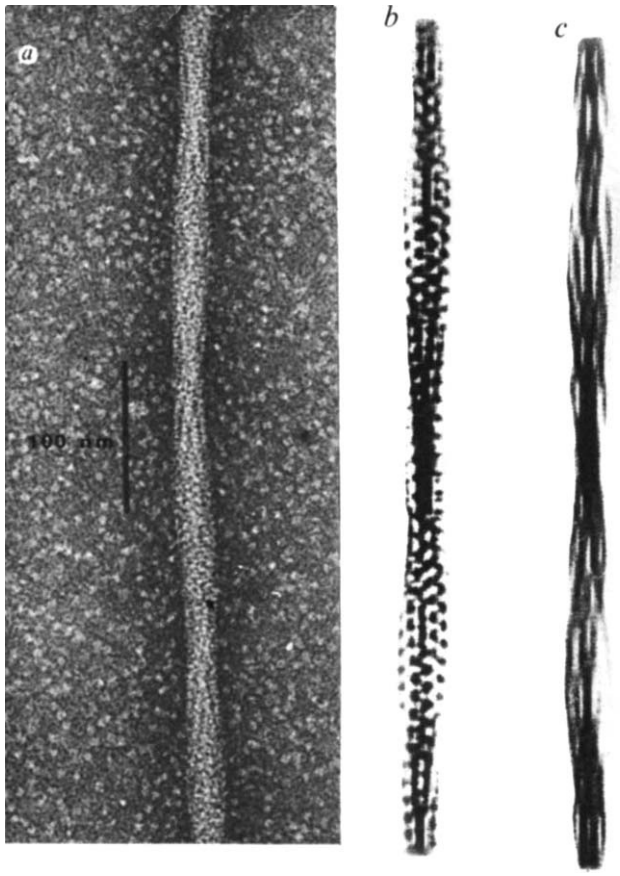



**Fig. 1** Negatively stained image and reconstructions of fibre of HbS. *a*, Micrograph of fibre of HbS prepared from a sickled cell by direct lysis with negative stain on the electron microscope grid. *b*, Two-dimensional reconstruction of the fibre of HbS using computer reconstruction techniques with the maxima from the Fourier transform displayed in Fig. 2. The output is recorded from a Tektronix graphics terminal. *c*, Two-dimensional reconstruction as in (*b*), but with only the maxima of layer lines 1–6 of the Fourier transform used.

phosphotungstate on high resolution grids, as described previously[9]. A Philips EM 301 electron microscope was used and images were recorded with minimum beam exposures. A typical fibre is presented in Fig. 1 and shows the periodic variation in apparent diameter that is a consistent feature. Short regions of striations parallel to the fibre alternating with regions having a more cross-hatched appearance are also evident. To compute Fourier transforms of the fibres, electron microscope plates were digitised with a Syntex AD-1 autodensitometer. This instrument was interfaced to a Nova computer equipped with tape and disk drives and a Tektronix graphics terminal, permitting the digitisation, transformation and reconstruction to be carried out on the same computer. The development of this on-line image reconstruction system has greatly facilitated our studies, enabling us to easily compute numerous transforms on many different fibres. Only after a consistent Fourier transform pattern of amplitudes was found repeatedly were we prepared to accept its validity, since it has several uncommon features. A transform showing these various features is reproduced in Fig. 2. Many maxima are present clustered into three regions: a region near the equator, a region in the middle of the transform (around layer line 24) corresponding to spacings in the vicinity of 1/60 Å, and a region near the top of the transform (around layer line 49) corresponding to spacings in the vicinity of 1/30 Å. These same regions were identified in earlier optical transforms[7–9], but the complexity of the various maxima within each region was not appreciated, since it was beyond the resolving power of the optical diffractometers used. Two-dimensional reconstructions, which do not require assignment of Bessel orders, give good agreement with original fibres, as seen in Fig. 1*b*. Thus, the Fourier amplitude map shown in Fig. 2 represents the 20-nm diameter fibres and the next step was to evaluate the information contained in the map.

## Assignment of Bessel orders

For most of the fibre structures of proteins previously analysed by Fourier methods, patterns considerably less complex than Fig. 2 were obtained[12] and the maxima could be reconciled with a single helical lattice. In the fibres of haemoglobin S the reciprocal space pattern is composed of contributions from several overlapping families of lattices, including contributions from structural elements at the interior of the fibres. The final assignment of Bessel orders to each maxima was therefore a difficult process and several approaches were used (a detailed description of each approach will be published elsewhere).

First, the question of whether a particular maximum arose from an even-ordered or an odd-ordered Bessel function was determined in the usual way[11]: by evaluating whether the left and right members of each pair were in phase (corresponding to even order) or out of phase (corresponding to odd order). The results of this analysis were unambiguous and indicated that the maxima in the central region of the transform (layer lines 21–28) were all of odd order, whereas all other maxima (on layer lines 1–6 and 46–51) were of even order.

Second, efforts were directed at assigning the absolute Bessel order for a given maximum as the ratio of phase change per degree of tilt (parallel to the fibre axis) for micrographs recorded at various angles with the goneometer stage, as described by Finch[13]. This method was only partially successful, principally because it required recording several images from each fibre and the quality of the computed transforms decreased markedly with successive exposures. Nevertheless some useful information was obtained, particularly for the maxima near the equator, for which the data indicated a trend from low to high Bessel orders with increasing layer line number (for layer lines 1–6). In addition, the same Bessel order was indicated for the multiple maxima of a given layer line.

Third, we developed a new approach to assigning Bessel orders based on a combination of real space and Fourier space reconstruction methods. The basis of this approach was to compute a two-dimensional reconstruction of the fibres using only the reflections from layer lines near the equator (1–6) to give a reconstructed analogue of the fibre with an array of continuous strands, as shown in Fig. 1*c*. This idealised fibre had an apparent repeat of 3,000 Å. A real space algorithm[14,15] was then used to deduce the cross section of the idealised fibre, assuming that each increment along the length of the fibre corresponded to a specified degree of rotation. The outcome was an elliptical cross section composed of four inner strands and 10 outer strands. Transforms of structural models based on this pattern gave near-equatorial reflections similar to those seen on the transforms of haemoglobin S fibres and rotations of the model permitted the Bessel orders of the reflections to be assigned[13]. The rotations indicated that layer lines 1–6 contained reflections of Bessel orders 2, 4, 6, 8, 10 and 12, respectively. Once these assignments were made, the Bessel orders of the reflections on the higher layer lines could also be determined, using the traditional reasoning of lattice building. For example, layer lines 5, 27 and 49 can be connected by a lattice line with maxima of Bessel orders 10, 5 and 0 on the three layer lines, respectively. The various assignments are summarised in Table 1. In effect, a distinct reciprocal space lattice can be drawn for each of the 12 distinct maxima on each side of the first six layer lines

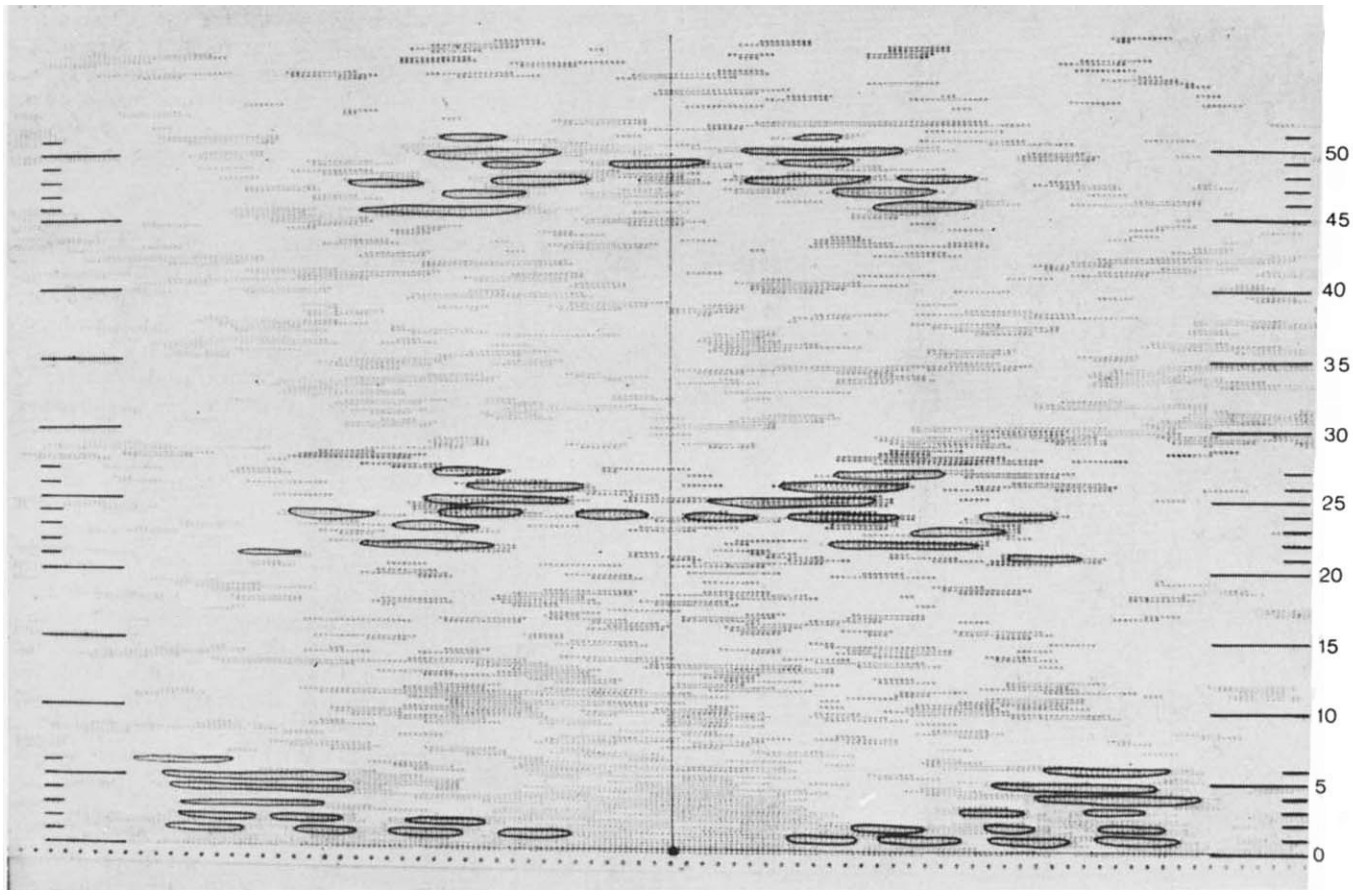



**Fig. 2** Fourier transform amplitude map of fibre of HbS. The map was obtained by computer transform of a 256 × 256 array of densities corresponding to a region of the HbS fibre digitised on the Syntex AD-1 autodensitometer. The computations were performed on a Nova computer and the output displayed as integers corresponding to increasing density levels on a high-speed printer. Regions of density above background were outlined in black to aid visualisation and the various layer lines of the maxima are numbered on the right. The first layer line corresponds to a spacing along the fibre axis of 1,500 Å.

with other points of the reciprocal space lattices lying among the maxima on layer lines 21–28, and with additional lattice points on layer lines 46–51. The clusters at the middle and higher ranges of lattice lines contain fewer maxima than in the near-equatorial region, since lattice points from more than one lattice are superimposed in several cases among the maxima in the middle and upper regions of the transform.

## Three-dimensional reconstructions

The arrangements of strands can be visualised by reconstructing the near-equatorial reflections in three-dimensions. Such a reconstruction is shown in Fig. 3. The cross-section indicates that the fibres are packed with individual strands in a hexagonal array. An important feature of the structure is that the four inner strands are not equivalent, but occur in two pairs. Thus, the distance between two of the inner strands (11 and 13 in Fig. 3) is greater than the distance between the other two inner strands (12 and 14 in Fig. 3). The outer strands also occur with a family of radial spacings. This inequivalence in the radii of the strands results in a noncircular, roughly elliptical cross section.

When three-dimensional reconstructions are computed using the reflections of Table 1 to 60 Å resolution, density varies along the length of each of the 14 strands and the appearance of the sections depends on the position along the fibre axis. By following individual strands through many such sections and noting the level of maximum density, the locations of molecules along the strands of the inner and outer core can be deduced. This information is summarised in the form of a surface lattice in Fig. 4 and as a sphere model in Fig. 5. Data at higher resolution (to 30 Å) may reveal subunit orientations in addition to the location of molecules and this point is currently under investigation.

**Fig. 3** Three-dimensional reconstruction of fibre of HbS. This reconstruction was computed using the maxima of layer lines 1–6 (corresponding to the two-dimensional reconstruction of Fig. 1*c*) in order to permit each of the 14 strands to be clearly displayed in the same reconstruction. In reconstructions which also include maxima from higher layer lines, densities from certain strands are reduced or absent and compilations of many such reconstructions from different positions along the fibre axis permitted the overall helical pattern of the fibre to be specified, as displayed in Fig. 4. This figure is a photograph from the Tektronix terminal with densities displayed in terms of contour lines.

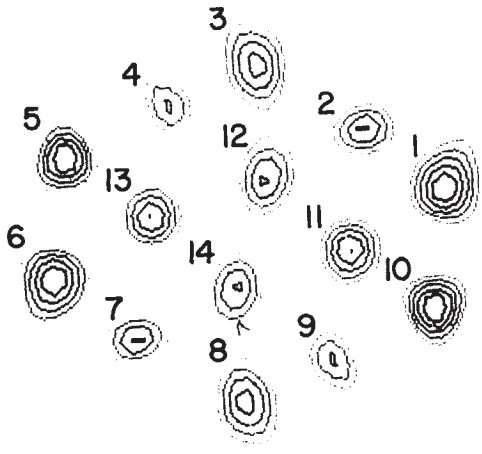


**Table 1** Layer lines and Bessel orders for Fourier transforms

| Layer line | Bessel order | No. of maxima |
|---|---|---|
| 1 | 2 | 4 |
| 2 | 4 | 3 |
| 3 | 6 | 2 |
| 4 | 8 | 2 |
| 5 | 10 | 1 |
| 6 | 12 | 1 |
| 21 | −7 | 1 |
| 22 | −5 | 1 |
| 23 | −3 | 2 |
| 24 | −1 | 3 |
| 25 | 1 | 1 |
| 26 | 3 | 1 |
| 27 | 5 | 1 |
| 28 | 7 | 1 |
| 46 | −6 | 1 |
| 47 | −4 | 2 |
| 48 | −2 | 2 |
| 49 | 0 | 2 |
| 50 | 2 | 1 |
| 51 | 4 | 1 |

The Bessel orders are as assigned to the maxima of the respective layer lines of the transforms of the fibres. For layer lines with more than one maxima, as indicated in the column on the right, each maxima was characterised by the same Bessel order. These data are a composite from several transforms, as individual transforms may have one or more reflections absent.

## Implications of the 14-stranded structure

The long-range goal of the structural studies on fibres of haemoglobin S is to bring the analysis to atomic resolution, so that the particular amino acid residues at each intermolecular contact can be specified. The complexities in the structure of the fibres revealed in this report suggest that description of the intermolecular contacts in atomic terms will be more difficult than previously imagined. It is thus not surprising that the list of sites on the haemoglobin S molecule where changes can alter fibre formation is growing[16–21]. It now seems likely that many positions on the molecule participate in the self-association process involved in fibre formation once the thermodynamics have been tipped towards fibre formation by the primary Glu→Val transition. A large number of distinct classes of intermolecular contacts are implied by the variety of nearest-neighbour interactions, but some simplifications in the distinctions between contacts may develop through application of the principle of quasi-equivalence which has been useful in other supramolecular structures, particularly viruses.

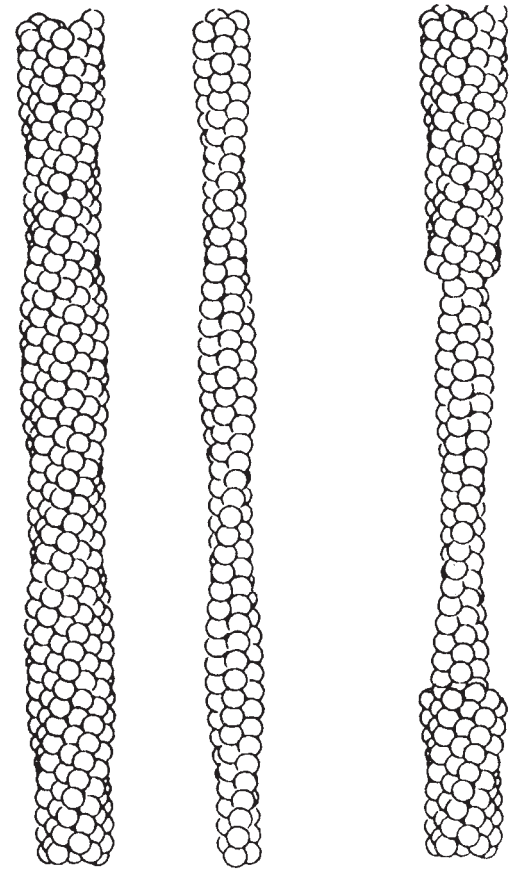
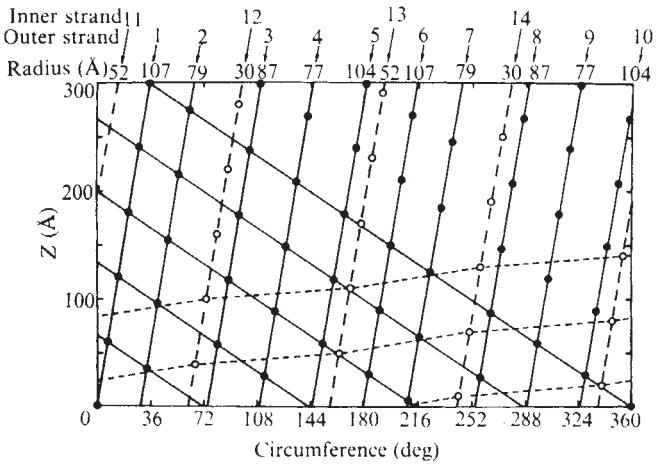


**Fig. 5** Solid sphere models of the helical elements of the fibres of HbS. The models present the outer strands (on the left); the inner strands (in the centre) in alignment with the outer strands such that the narrow diameter region of the inner strands corresponds to the narrow diameter region of the outer strands, and a cutaway version (on the right) showing both inner and outer strands in proper juxtaposition.

**Fig. 4** Surface lattice of the fibres of HbS. The 14 strands of the fibres are displayed with a numbering system corresponding to the designations in Fig. 3. For each strand the centre of its radial distance is indicated at the top of the surface lattice. The outer 10 strands are represented with filled circles and the inner 4 strands with open circles. Because of the differences in the radii of the strands, the lines of the surface lattice cannot all be straight. The data are not sufficiently precise to specify exactly the small perturbations from linearity, so the surface lattice is arbitrarily drawn with the left-handed 5-start diagonal lines linear. The inner four strands occur at two pairs of radii such that a pseudo 1-start helix (dashed line) has a slope which alternates between two angles, which differ slightly, at each intersection with an inner strand.




The new structure of the fibres may also help to explain the exceptionally high order in the kinetics of fibre assembly[22–24] and the high density of gels[25]. Our preliminary calculations based on the 14-stranded model indicate that the densities of closely packed fibres are consistent with the fibre pellet concentrations reported by Ross *et al.*[25]. In fact, the close hexagonal packing of the individual strands is an arrangement which may maximise the quantity of protein in a volume element of the fibres. It remains to be seen whether similar packing will emerge as a general pattern for helical protein assemblies. One example already reported is the seven-stranded cable of glutamine synthetase[26] which has a hexagonal cross-section. The advantages of hexagonally packed cables have been appreciated for some time[27–28], particularly units with three strands and seven strands[26]. If this series is extended, 10-stranded and 14-stranded structures would also be predicted, with the HbS structure presented here as an example of the latter.

The complexities of the 14-stranded structure now make it less likely that data on intermolecular contacts from crystals of single molecules of haemoglobin S[6] will be directly applicable to the intermolecular contacts of the

fibres. However, the fact that the bundles formed by stirring haemolysates[29] are crystalline arrays of fibres in the form described here[9] offers a possibility of obtaining data to higher resolution by crystallographic means which, in conjunction with existing atomic models for individual molecules of HbS[6], could be sufficient to specify molecular orientations within the fibres and the stereochemistry of the contacts.

This work was supported by the USNSF (grant BMS74 00012) and NIH (HL-13591 and CA-14454).

Received 30 December 1977; accepted 27 January 1978.

1. Herrick, J. B. *Arch. Int. Med.* **6**, 517–521 (1910).
2. Pauling, L., Itano, S., Singer, J. & Well, I. C. *Science* **110**, 543–548 (1949).
3. Magdoff-Fairchild, B., Swerdlow, P. H. & Bertles, J. F. *Nature* **229**, 217–219 (1972).
4. Finch, J. T., Perutz, M. F., Bertles, J. F. & Dobler, J. *Proc. natn. Acad. Sci. U.S.A.* **70**, 718–722 (1973).
5. Edelstein, S. J., Telford, J. N. & Crepeau, R. H. *Proc. natn. Acad. Sci. U.S.A.* **70**, 1104–1107 (1973).
6. Wishner, B. C., Ward, K. B., Lattman, E. E. & Love, W. E. *J. molec. Biol.* **98**, 179–194 (1975).
7. Josephs, R., Jarosch, H. S. & Edelstein, S. J. *J. molec. Biol.* **102**, 409–426 (1976).
8. Edelstein, S. J. *et al. Proceedings of the Symposium on Molecular and Cellular Aspects of Sickle Cell Disease* (eds Hercules J. I. *et al.*) 33–59 (DHEW Publication No. (NIH) 76-1007 (Bethesda, 1976).
9. Crepeau, R. H., Dykes, G. & Edelstein, S. J. *Biochem. biophys. Res. Commun.* **75**, 496–502 (1977).
10. DeRosier, D. J. & Klug, A. *Nature* **217**, 130–134 (1968).
11. DeRosier, D. J. & Moore, P. B. *J. molec. Biol.* **52**, 355–369 (1970).
12. Crowther, R. A. & Klug, A. *A. Rev. Biochem.* **44**, 161–182 (1975).
13. Finch, J. T. *Proc. 6th European Cong. Electron Microscopy*, 578–579 (1972).
14. Gordon, R., Bender, R. & Herman, G. T. *J. theor. Biol.* **29**, 471–381 (1970).
15. Klug, A. & Crowther, R. A. *Nature* **238**, 435–440 (1972).
16. Bookchin, R. M., Nagel, R. L. & Ranney, H. M. *J. biol. Chem.* **242**, 248–255 (1967).
17. Bookchin, R. M. & Nagel, R. L. *J. molec. Biol.* **60**, 263–270 (1971).
18. Kraus, L. M., Miyaji, T., Iuchi, I. & Kraus, A. P. *Biochemistry* **11**, 3576–3582 (1972).
19. Bookchin, R. M. & Nagel, R. L. *Seminars in Hematology* **11**, 577–595 (1974).
20. Benesch, R. E., Yung, S., Benesch, R., Mack, J. & Schneider, R. G. *Nature* **260** 219–221 (1976).
21. Benesch, R. E., Kwong, S., Benesch, R. & Edalji, R. *Nature* **269**, 772–775 (1977).
22. Hofrichter, J., Ross, P. D. & Eaton, W. A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3864–3868 (1974).
23. Malfa, R. & Steinhardt, J. *Biochem. biophys. Res. Commun.* **59**, 887–893 (1974).
24. Moffat, K. & Gibson, Q. H. *Biochem. biophys. Res. Commun.* **61**, 237–242 (1974).
25. Ross, P. D., Hofrichter, J. & Eaton, W. A. *J. molec. Biol.* **115**, 111–134 (1977).
26. Frey, T. G., Eisenberg, D. & Eiserling, F. A. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3402–3406 (1975).
27. Pauling, L. & Corey, R. B. *Nature* **171**, 59–61 (1953).
28. Crick, F. H. C. *Acta Crystallogr.* **6**, 689–697 (1953).
29. Pumphrey, J. G. & Steinhardt, J. *Biochem. biophys. Res. Commun.* **69**, 99–105 (1976).

# Gene *K*, a new overlapping gene in bacteriophage G4

**D. C. Shaw*, J. E. Walker, F. D. Northrop, B. G. Barrell, G. N. Godson† & J. C. Fiddes‡**

MRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK

*A third overlapping gene in the isometric bacteriophages has been identified by nucleotide sequence analysis of the G4 genome and by isolation and microsequence analysis of the new protein. Five nucleotides are used in all three translational reading frames of the DNA.*

THE single-stranded DNA bacteriophage G4 is closely related to phage ΦX174 (ref. 1); its genome has the same overall structure as that of ΦX174 (J.C.F., B.G.B. and G.N.G., unpublished), and despite a 40% base sequence difference, codes for a similar set of 10 proteins.

The complete nucleotide sequence of ΦX174 revealed that 2 of the 10 genes so far identified overlap and are translated in a different reading frame[2–4]. An examination of all three translational reading frames of the complete sequence has revealed several other possible overlapping genes (coding for proteins of more than 20 amino acid residues) which have the following characteristic features[5]. An open translational reading frame of at least 60 nucleotides comes before a termination codon in the same phase, and the ATG initiation codon is preceded by a sequence complementary to the 3′ end of the 16S ribosomal RNA. This presumed ribosome-binding site should be at least three nucleotides long and be within 15 nucleotides of the initiation codon[6,7]. Sequences very similar to at least two of these putative genes have also been found in phage G4 (J.C.F., B.G.B. and G.N.G., unpublished).

We now provide evidence from nucleotide sequencing combined with protein chemistry, for a hitherto unknown phage-specified protein whose amino acid sequence corresponds to one of these nucleotide sequences.

This gene has been named *K* following the convention adopted for ΦX174. Gene *K* overlaps genes *B*, *A* and *C* and is the third example of an overlapping gene in the isometric phages. Two regions of gene *K* briefly overlap with two other genes simultaneously so that all three translational reading frames of the DNA can be used.

## DNA sequence analysis of gene *K*

Inspection of the DNA sequence of bacteriophage ΦX174 (ref. 5) in the gene *A*/*B* and *C* region revealed a potential new translational reading frame extending from the ATG at nucleotide 51 to the TGA termination codon at position 210. The ATG was preceded by a GGA sequence eight nucleotides upstream which is complementary to the nucleotide sequence shown boxed at the 3′ end of 16S rRNA, that is, 5′ GAUCACC-UCC UUA$_{OH}$ 3′. This sequence has been shown to be involved in the initiation of protein synthesis[6,7]. This possible new gene (gene *K*) would code for a protein of 56 amino acids. The initiation codon overlaps the termination codon of gene *B* and extends through the end of gene *A* into gene *C*.

To see if this potential gene is conserved in the related bacteriophage G4 the DNA sequence in this region was determined using the 'plus and minus' method, and more recently, the new chain termination method of Sanger *et al.*[8]. Autoradiographs of three gels showing the sequence of the G4 gene *K* region are shown in Fig. 1. Because of the high homology between the ΦX174 and G4 DNA sequences in this region it was possible to use the ΦX174 *Taq* I fragment 9 as a primer with G4 viral or complementary DNA strand using the chain termination method (Figs 1*a* and 1*c*, respectively). This fragment extends from nucleotide 61 to 93 (Fig. 2). The *Taq* I site(T CGA) at the 5′ end of this fragment is conserved in G4 but the site at the 3′

*Permanent address: Department of Physical Biochemistry, John Curtin School of Medical Research, Australian National University, Canberra, Australia.
†Permanent address: Radiobiology Laboratories, Yale University School of Medicine, New Haven, Connecticut 06510.
‡Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143.