

### Lessons for Today and Tomorrow from Yesterday

S. Arnott

The University, St Andrews, KY16 9AJ, Scotland

*Many early X-ray fibre diffraction analyses were flawed by poor modelling rather than low quality data. This revelation should prompt the adoption of more standard methods of analysis to facilitate critical appraisals of both models and experimental observation in this field.*

### The needless tension between X-ray analysis of molecular structures in fibres and in crystals

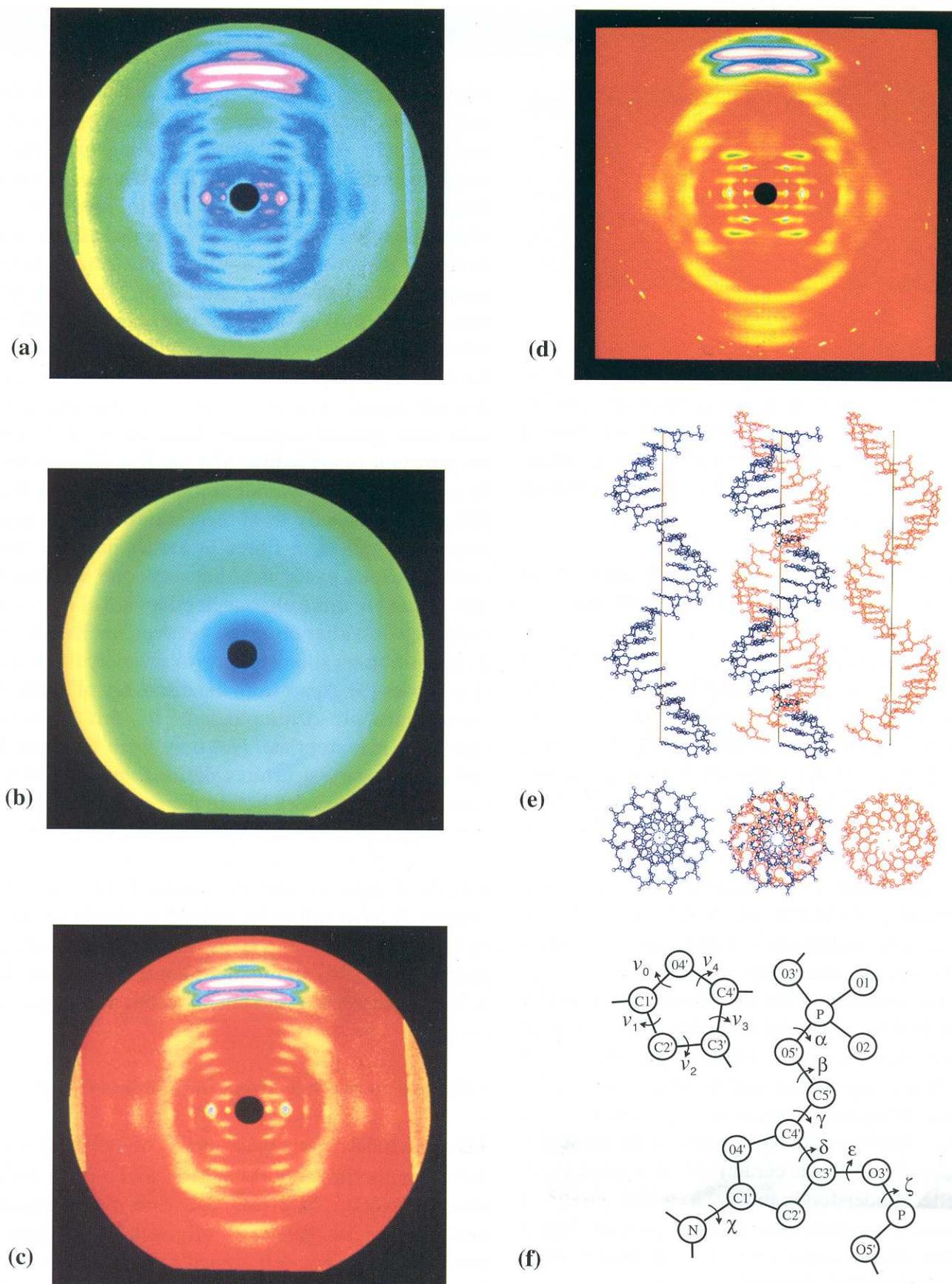
The molecular architecture of important linear polymers such as the polypeptide, polysaccharide and polynucleotide examples touched on below can be unveiled by X-ray diffraction analysis of their uniaxially oriented fibres that are sometimes further ordered in a variety of ways up to the level of microcrystals. Like almost all physical methods the technique has its idiosyncratic challenges, but its final requirement is production of a credible, parametrically parsimonious model that fits the observed data *significantly* better than any other plausible but parametrically equally parsimonious model. The accuracy of the final model should not be in doubt, but the precision of its description need be no greater than the use to which its parameters will be put. When appropriate and possible it is better to use X-ray diffraction analysis of single crystals with extensive three-dimensional order: with these the route to a unique (*i.e.* accurate) model is more facile and better understood among a wider group of practitioners; crystals also usually provide higher resolution data in larger numbers that lead to more precise and detailed structures.

For fibre diffractionists to concede the possible superiority of crystallographic analyses in certain circumstances is not to concede this in all circumstances, nor to provide a licence for later, merely authenticating crystallographic studies to airbrush from history the earlier pioneering discoveries of fibre diffractionists. To use

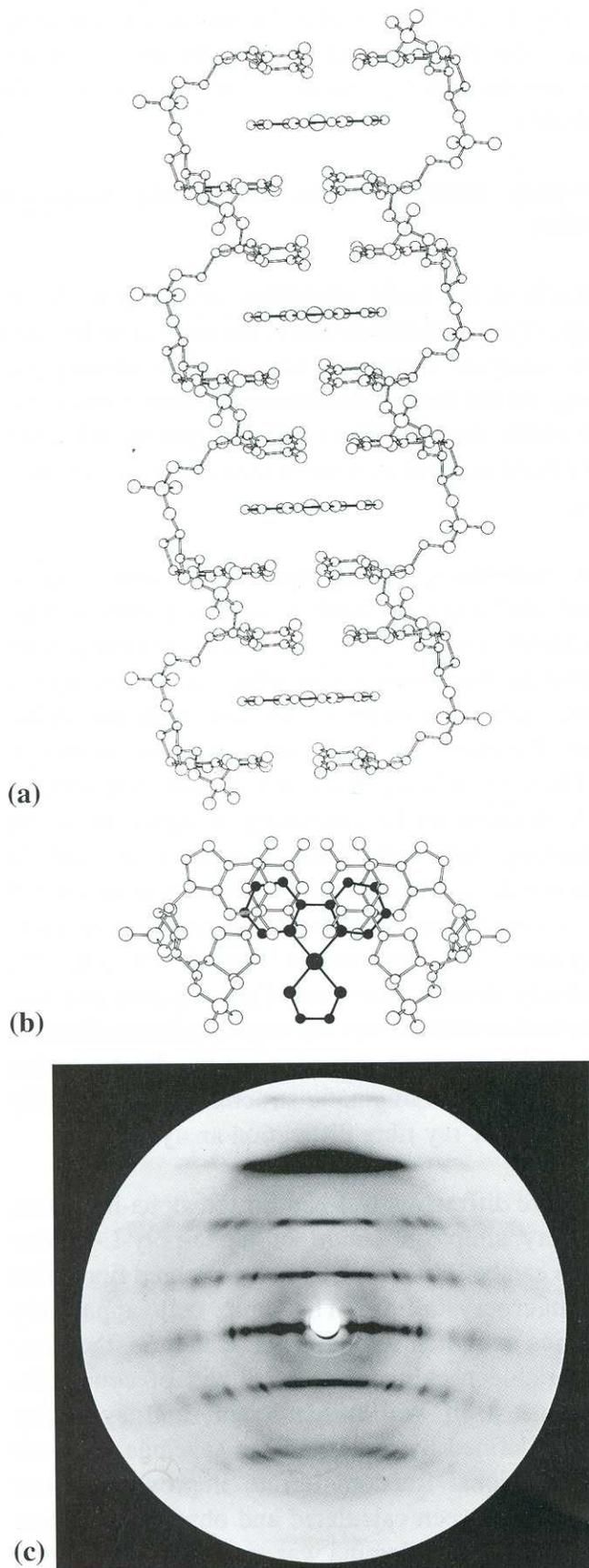
crystallographic analyses in such a fashion is not only intellectually fraudulent but also economically pernicious in diverting public funding from truly novel investigations. It is true, for example, that two decades of X-ray crystallographic analyses of oligonucleotide duplexes were launched by the dramatic discovery of a novel left-handed helix morphology accessible to oligo d(GC) duplexes [1] and, in a reversal of the usual sequence of events, a subsequent fibre study authenticated this discovery for similar polynucleotides [2]. But it is also the case that subsequent studies of oligonucleotide duplexes with more general sequences have shown that most are reluctant to crystallize and the few variants that do crystallize have overall secondary structures that differ insignificantly from the various B and A-type structures defined in fibres in earlier decades. The conformational excursions between one nucleotide and the next in these oligonucleotides are of the same order as was earlier observed between nucleotides in different fibrous allomorphs with a variety of helical symmetries [3]. Apart from the discovery of Z-DNA, crystallographic triumphalism in this area is especially inappropriate. Not only were the main B and A secondary structures defined and refined by fibre analyses decades before their re-description in oligonucleotides, but novelties like the heteromeric poly dR. poly rY structures [4] for DNA-RNA hybrids (Figure 1), and the achiral L-DNA duplex (Figure 2) have still to be sighted in single crystals. In such circumstances a sympathetic synthesis of the results of the two techniques should have been the more productive aim and outcome, but this has yet to occur.

### Collagen — another arena for collaboration?

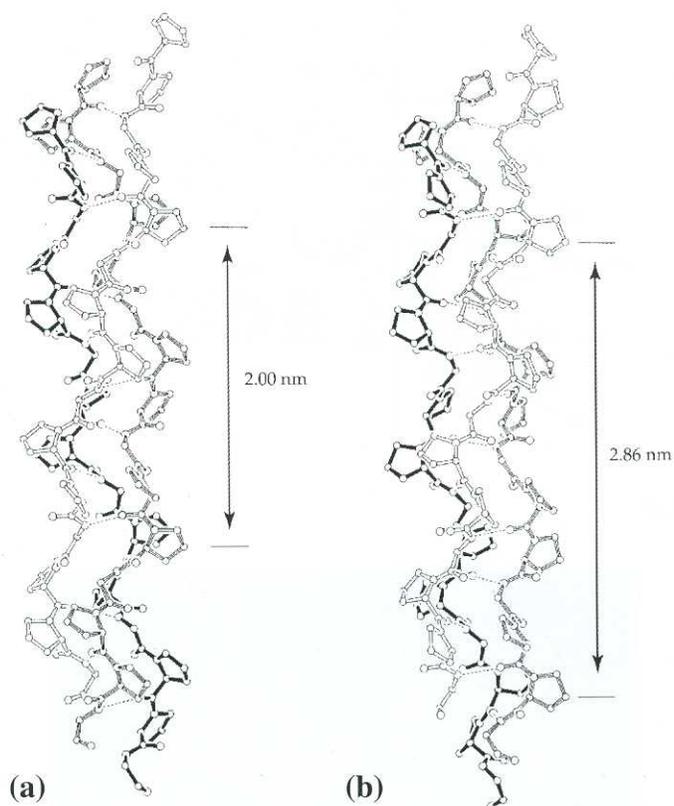
The canonical  $10_7$  triplex helix of collagen was defined authoritatively [5] in 1961 and refined using fibre diffraction data [6]. The first single crystal structure of a relevant oligopeptide, (hydroxyproline)<sub>10</sub>, was not determined until 33 years later [8] but its  $7_5$  triple helices merely authenticated Okuyama's (1981) quasi-fibrous structure [7] of (proline-proline-glycine)<sub>10</sub> (Figure 3(a)) which was the first to reveal the alternative  $7_5$  triplex symmetry and show that collagen helices could be significantly dimorphic. These  $7_5$  triplexes intriguingly have the same unit translation vertically as the  $10_7$  triplexes (Figure 3(b)), but dramatically different rotations per



**Figure 1:** Modern methods of data collection using scanners and further processing by computers are illustrated by the sequence showing (a) a badly recorded X-ray fibre pattern from the synthetic DNA-RNA hybrid poly dI.poly rC, (b) the asymmetric background determined by sampling the interdiffraction space on the film record, and (c) the cleaned-up version of the pattern which shows the mixture of Bragg and non-Bragg diffraction from a screw-disordered uniaxially oriented array of 10-fold helices. A similar pattern from the 11-fold helices of poly dA.poly rU is shown in (d). Both patterns when analysed reveal duplex structures with non-identical iso-helical chains such as those for poly dA.poly rU shown (e) in two mutually perpendicular projections as a duplex and its deconstruction into two anti-parallel components. Such polymorphism even within a rather regular duplex should be a not unexpected feature of polynucleotides where even one residue (f) has many degrees of conformational freedom.



**Figure 2:** An exotic allomorph of the DNA duplex has a ladder structure (a) with no net twist over the periodicity of  $3 \times 3.4 \text{ \AA} = 10.2 \text{ \AA}$  which is the layer line spacing in the diffraction pattern (c). The bipyridyl-ethylene-diaminyl platinum intercalant that induces and supports this DNA structure is shown as filled circles in (b) which depicts the projection of the intercalated DNA complex down the long molecular axis.

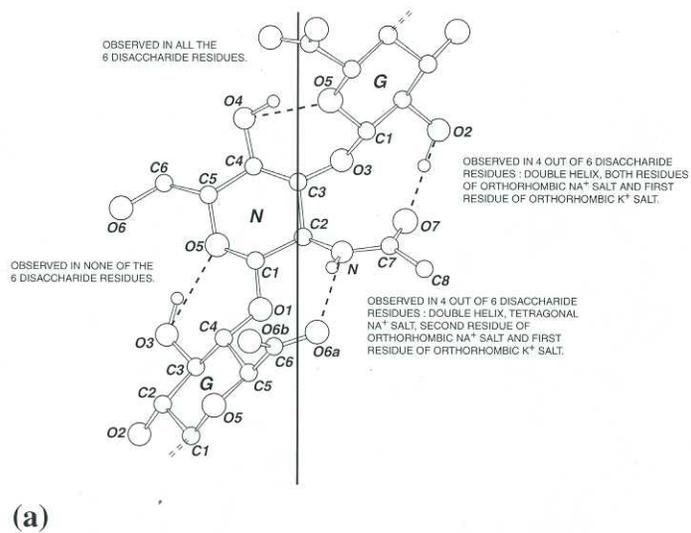


**Figure 3:** Two allomorphs of the collagen triple helix observed: (a) in a quasi-fibrous crystal of  $(\text{pro-pro-gly})_{10}$  [7] and in single crystals of  $(\text{hyp-pro-gly})_{10}$  [8] in which one gly in the middle of the sequence has been replaced by ala; and (b) in a native collagen. In the first allomorph the three polypeptide chains with  $7_1$  symmetry (twist  $51.4^\circ$ ) are nested to produce a triplex with  $7_1$  symmetry and unit height  $2.86 \text{ \AA}$  which is the same as in the canonical  $10_7$  triplex model in which the individual chains have  $10_1$  symmetry and a rather different twist ( $36.0^\circ$ ).

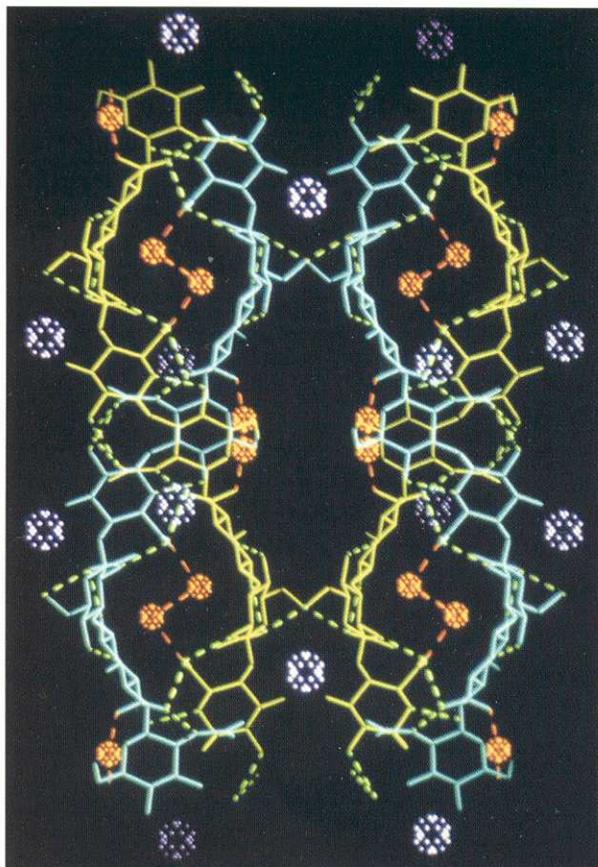
tripeptide ( $51.4^\circ$  and  $36.0^\circ$  respectively) in their individual  $7_1$  and  $10_1$  helical chains. This could be the mechanism whereby arrays of collagen helices of fixed length might optimize lateral inter-triplex contacts by different local rotations. More structures of differently sequenced, collagen-like structures in either fibres or crystals or both are needed to explore this peculiar polymorphism.

### Polysaccharide structures are almost completely dependent on fibre studies

It is not surprising that extensive crystals are not found for such floppy, multifunctional, often polymorphic molecules [9], even when they contain multiple sources of chain stiffening as illustrated for hyaluronate in Figure 4(a). As a result of further stiffening by duplex formation (Figure 4(b)), which involves hemi-protonation, hyaluronate solutions can form a "putty" at  $\text{pH} \sim 3$  [10] when, theoretically,



(a)



(b)

**Figure 4:** It is possible (a) for the polydisaccharide units of hyaluronate and related connective tissue polysaccharides to be stiffened by multiple intramolecular hydrogen bonds [9] and (b) for hyaluronate uniquely to be further buttressed by double helix formation when hemi-protonated and interchain carboxyl-carboxylate and other hydrogen bonds are formed with the aid of special pairs of water molecules as revealed by fibre diffraction analysis [12]. These provide extra scaffolding for the duplex.

all the chains could be paired. *In vitro* at higher pH hyaluronate solutions exhibit anomalous viscosities, presumably because there is a persistent population

of stiff double helices. It is the anomalous viscosity that is the property that is exploited *in vivo* in the vitreous humour of the eye and in the synovial fluid of joints.

### Wrongly misprised data and rightly traduced models

Since it is the facile credibility accorded to X-ray single crystal structures that is the main contrast with fibre analyses it may be important to identify the source of the lack of confidence in fibre studies and if possible show that the credibility gap should never have been as wide as it seemed or is now less than it was.

Any community of experimental scientists that is small and wayward finds it hard to establish high credibility for its results, but those concerned with important biopolymer secondary structures had a particularly inauspicious start both with the alpha-helix of proteins and with commonest (B) allomorph of DNA. In both cases the initial molecular models were deemed to be essentially accurate from the beginning, but both failed to provide a good fit between the calculated amplitudes of diffraction and the corresponding amplitudes measured from fibre diagrams. The convenient conclusion that the relatively strange data were of poor quality and low information content was not easy to refute at the time and so the moments passed when the determination of two major paradigmatic structures could be fully credited to X-ray fibre diffraction analysis.

The fibre diffraction test-bed for the alpha-helix was the very simple structure of alpha-poly-L-alanine which can be spun into uniaxially oriented fibres that are microcrystalline. The unit cell apparently contains one pitch of one helical chain *i.e.* there are no variable packing parameters and of course the molecule itself with pitch length and symmetry readily defined by the layer line spacings has little conformational freedom left to improve the poor initial fit between calculated and observed structure amplitudes. The dilemma was resolved only in 1966 when it was shown [11] that the packing of alpha-poly-L-alanine consists of an array of its polar helices pointing up and down *at random*. The mutual displacement and orientation of the two sets of helices provide the two extra parameters needed to define the overall structure and the best values for these dramatically and significantly improve the fit between the model and observations.

The disjunction between model and data in the case of B-DNA was even more severe ( $R > 0.8$ ) than for the alpha-helix but had a different source in the molecular structure itself and not the packing. The original Crick and Watson model [13] was entirely accurate in its principal features — the isomorphous AT and GC base-pairs, the right-handed  $10_1$  helix symmetry, the mutually intertwined, anti-parallel polynucleotide chains — but in one respect there was an error — the sugar rings were puckered in the C3'-endo fashion appropriate for A-DNA-like allomorphs rather than in the C2'-endo fashion characteristic of B-DNA-like structures. This error, of little interest at the time, has crucial consequences for the gross morphology of DNA models incorporating it: in particular, base-pairs have to be sited 4-5 Å further away from the helix axis than is indeed the case in B-DNA; this in turn produces a diffraction pattern very different in its intensity distribution from the one actually observed.

Yet again the quality of the unfamiliar diffraction data was made the scapegoat for the difficulty. But it could have been shown that the information content of the misprised data is in fact very high. Even when the observed amplitudes are combined disadvantageously in a strongly biased fashion with X-ray phases calculated from the morphologically flawed, aboriginal DNA model [14] the resulting Fourier synthesis clearly reveals the deficiencies of the model and makes clear that the remedy is a major relocation of the base-pairs (Figures 5(a), (b) and (c)).

Again the revisions took place rather too late for the damage to the credibility of X-ray fibre studies to be easily remedied. In the meantime, encouraged by the Watson and Crick modelling coup, which owed little to local experimental effort, many other analyses of fibrous polynucleotide systems were undertaken with just as little investment in experiment but with much less insight. Deservedly, most of the conclusions from these forays were wrong, but from them grew the myth that accurate X-ray fibre studies were impossible because of the appalling negative record that not one of these fibrous nucleic acid structures produced in the 1950s and 1960s by a laboratory not of Maurice Wilkins' school, has survived critical re-examination [16].

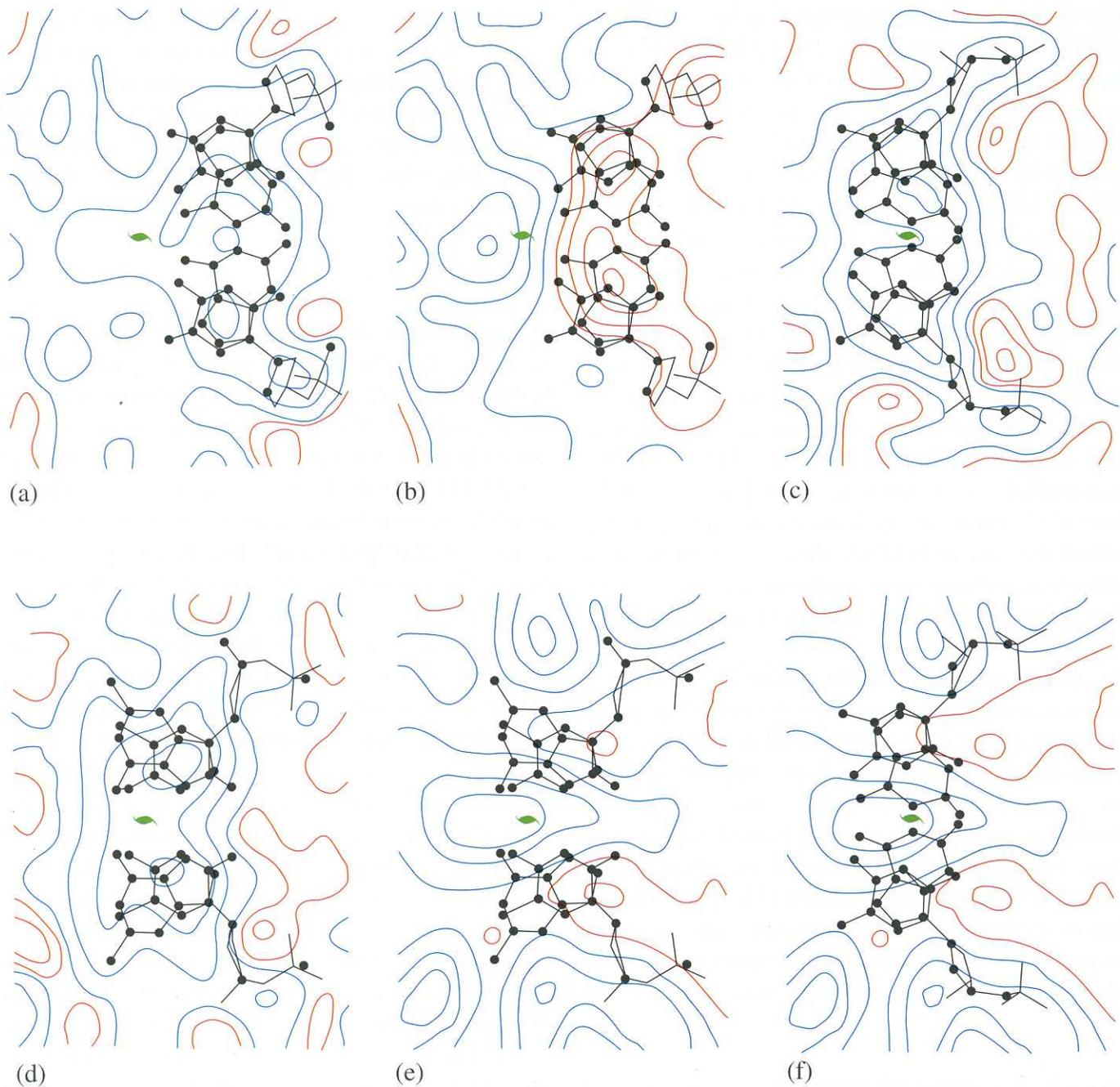
Despite this, the good news not only for fibre diffractionists, but also for all who need to rely on fibre-derived molecular structures, is that modelling

such structures with scrupulous application of today's technologies allows most gross errors to be detectable if enough effort is invested in alternative models. Very detailed structures and their subtle structural features can now be teased out routinely from data once unfamiliar to narrowly educated diffractionists.

## Conclusion

What should be clear is that the information content of X-ray diffraction data from fibres can be much higher than generally is conceded. What is needed is extra credibility for the conclusions derived from these data that has to come, in part, from analyses conducted within as standard a framework as possible so that results can be authenticated by a larger community of knowledgeable and sympathetic peers. The components of a modern fibre diffraction analysis include better specimen preparation, more intense X-ray sources, better data collection and handling, copious and fast computer facilities that trivialise the burden of refining and comparing all plausible alternative models (when the X-ray phase problem is not capable of experimental solution), standard frameworks for analysis such as linked-atom least-squares followed by statistical tests to adjudicate among competing models, a better awareness of the existence of structures with statistical or screw-disordered or other less than crystalline packings, and an ability to handle quantitatively the mixture of Bragg and non-Bragg diffraction that arises from such systems (*e.g.* Figure 1). A major effort to make and maintain all these procedures and processes within as uniform and standard a framework as possible would be of major benefit to the whole field.

Diffraction analysis of fibrous materials is too important for this effort to be postponed or evaded. Many linear polymer systems inherently are incapable of forming crystals of any great lateral extent. Indeed, many of the properties of gel-forming polysaccharides (for example) arise specifically from molecular bundling that is not very extensive. These and other repetitious polymers also are often quasi-cylindrical so that even the limited lateral order in their fibres is not periodic. The structures of all of these less-than-fully-crystalline systems are capable of being analysed quantitatively by X-ray diffraction methods, some more easily than others and to different levels of resolution. In all these circumstances it seems needlessly limiting for



**Figure 5:** (a) The electron density distribution in the plane of an (average) Watson-Crick base pair obtained with observed amplitudes and phases derived from the original Crick and Watson coordinates for DNA. The image is not only of poor resolution but discordant with the model. (b) The difference map shows that the major deficiency of the model is the position of the base-pairs in the negative (red contoured) region rather than astride the helix axis. (c) When the molecular morphology had been changed by Wilkins' refinements a new electron density synthesis with better phases shows a more concordant fit between model and experiment. (d) A biased electron density distribution calculated [14] using phases from a DNA model with Hoogsteen pairs. (e) The corresponding difference map shows that density is needed in places that cannot be accessed without changing the type of base-pairing as illustrated in (f) where a refined Watson-Crick-type model is superposed on the same difference map.

progressive structuralists to concentrate only on systems that crystallize and even more suspect to focus only on surrogate oligomers that also happen to crystallize. To do so must result in the selection of a very small subset of structures for detailed analysis and the postponement for decades of the discovery of other structures, thereby distorting the global view. If

we take only the fibre structures illustrated in this article as examples we can highlight the predicament: DNA-RNA hybrid duplexes with non-identical anti-parallel chains have been observed only in fibres; the same is true of the ladder-like DNA duplex complexed with an intercalated drug; collagen triplexes were discovered to be dimorphic

in fibres fourteen years before this was confirmed in a crystal; the behaviours of starch, cellulose, as well as connective tissue and other gel-forming polysaccharides have been rationalized exclusively from fibre diffraction studies; the structures of DNA and RNA single, double, triple and quadruple helices were defined from fibre analyses up to a quarter of a century before analogous structures were disclosed in single crystals.

Fibre diffraction analysis has many past achievements to be proud of and many future triumphs to anticipate.

## References

- [1] Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Broom, J.H., van der Marel, G. and Rich, A., *Nature* (1979) **281**, 680.
- [2] Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G.W. and Ratliff, R.L., *Nature* (1980) **287**, 561.
- [3] Dickerson, R.E., *Meth. Enzymol.* (1992) **211**, 67.
- [4] Arnott, S., Chandrasekaran, R., Millane, R.P. and Park, H.-S., *J. Mol. Biol* (1986), **188**, 631.
- [5] Rich, A. and Crick, F.H.C., *J. Mol. Biol* (1961) **3**, 483.
- [6] Fraser, R.D.B., MacRae, T.P. and Suzuki, E., *J. Mol. Biol* (1979) **129**, 463.
- [7] Okuyama, K., Arnott, S., Takayanagi, M. and Kakudo, M., *J. Mol. Biol.* (1981) **152**, 427.
- [8] Bella, J., Eaton, M., Brodsky, B. and Berman, H.M., *Science* (1994) **266**, 75.
- [9] Chandrasekaran, R., in *Adv. in Carbohydrate Chemistry and Biochemistry* (1997) **52**, 311.
- [10] Balazs, E.A., *Fed. Proc* (1966) **25**, 1817.
- [11] Arnott, S. and Wonacott, A.J., *J. Mol. Biol* (1966) **21**, 371.
- [12] Arnott, S., Mitra, A.K. and Raganathan, *J. Mol. Biol* (1983) **169**, 861.
- [13] Crick, F.H.C. and Watson, J.D., *Peoc. R. Soc. (London) Ser. A* (1954) **223**, 80.
- [14] Arnott, S. in *Bull. Inst. Phys. and Phys. Soc* (1964), 169-178.
- [15] Arnott, S., Wilkins, M.H.F., Hamilton, L.D. and Langridge R., *J. Mol. Biol* (1965) **27**, 391.
- [16] Arnott, S., in *Oxford Handbook of Nucleic Acid Structure* (1998), 1-36.

## X-Ray Studies on Biological Fibres

A. Miller, J.P. Orgel and T.J. Wess

University of Stirling

The most important scientific paper of the second half of the twentieth century was a report of X-ray fibre diffraction studies on DNA. Biological fibres have well developed crystallinity in the direction of the fibre axis, but are not so well ordered in the directions normal to the fibre axis. This simple one-dimensional order meant that fibres allowed the first recognition of the secondary structure of proteins - the alpha helix, beta sheet, coiled-coil alpha helix and collagen helix - as well as the double helix of DNA, long before X-ray crystallography yielded the full three-dimensional structures of biological macromolecules and complexes. In addition X-ray fibre diffraction studies gave information about the molecular packing in the native biological fibres of muscle, tendons and the cytoskeleton and it is this window into structural molecular biology which is the subject of this note.

Advances in the technology of X-ray fibre diffraction have always been the basis of its value. Small angle cameras with good resolution between the undeflected beam and the scattered beam closest to it, good reciprocal spatial resolution throughout the scattered pattern, intense X-ray sources of high spectral brilliance and efficient and effective X-ray detectors have all played their part. So we will start by discussing some aspects of synchrotron radiation which are undergoing developments which will be relevant to fibre work - as well as to macromolecular crystallography - in the near future. We then summarise our recent work on the molecular packing of Type I collagen in tendons and comment on recently published claims about features in the X-ray diffraction patterns from the alpha-keratin of hair and their relationship to breast cancer.

## Synchrotron Radiation

Third generation synchrotron sources such as the ESRF at Grenoble introduced the feature of flexibility of insertion devices which could be custom designed for specific scientific applications. In second generation synchrotrons the flexibility lay mainly in the design of the beam lines. Hence at ESRF, once the machine energy was decided as 6