Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

FibreHelix, a program for calculating the X-ray diffraction pattern of macromolecules with helical symmetry: application to DNA coiled coils

David Dominguez-Sal,[a] Xavier Messeguer[a] and Juan A. Subirana[b]*

[a]Algorithmics and Genetics Group, Software Department, Universitat Politècnica de Catalunya, C/Jordi Girona C6, 08034 Barcelona, Spain, and [b]Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, Avenida Diagonal 647, 08028 Barcelona, Spain

Correspondence e-mail: juan.a.subirana@upc.edu

Received 6 September 2004
Accepted 23 November 2004

A program has been developed to determine the diffraction pattern given by partially ordered fibres formed by macromolecules with helical symmetry. It is particularly useful for visualizing the splitting of layer lines typical of coiled coils. The program produces as output the diffraction diagram calculated for helices that are oriented along their axis but are randomly oriented in other directions. The results can be numerically analyzed and also visualized on-screen. The program has been applied to the diffraction patterns given by DNA and protein coiled coils.

1. Introduction

We have recently found in our laboratory that under certain conditions DNA oligonucleotides with sticky ends may form rigid coiled coils. These give rise to striking diffraction patterns that cannot easily be interpreted with available software packages. Therefore, we decided to develop a new program suitable for analyzing the diffraction patterns of helices and coiled coils.

When analyzing the internal structure of a polymer fibre, many geometric configurations are possible. Helical symmetry (helices and coiled coils) is often found. *FibreHelix* is a program built to help researchers to determine whether they are dealing with a helical molecule and then to find the correct parameters for it. Moreover, the program can be used in training with simple models in order to understand the effects of each variable in the diffraction pattern in an easy and fast way.

Although the diffraction patterns of fibres with helical and coiled-coil molecular structures have been analyzed theoretically already (Cochran *et al.*, 1952; Fraser *et al.*, 1964*a*; Wilson, 1966), no computer programs have been developed for this purpose. To our knowledge, there is no software available to calculate and visualize diffraction intensities from fibres of a coiled coil. Recently, a new program called *Helix* has been presented in the frame of *CCP*13 (Knupp & Squire, 2004). However, it is a very simple program that does not take into account the molecular structure of the residues that build the helix.

FibreHelix uses the full theory for discontinuous helices: it calculates the layer-line intensities from the atomic model. So, when using it for calculating intensities we know that we are applying the complete model for the helices, avoiding simplifications that could be severe in complex molecules. Moreover, after calculating the intensities it can simulate the experimental diffraction pattern.

This paper is divided into several parts. In §2 we describe the theoretical formulas used. In §3 we present the software itself. In §4 we give two examples, a polyalanine coiled coil (which has been used as a model of keratin; Fraser *et al.*, 1964*a*,*b*) and a DNA coiled coil.

2. Theory for a discontinuous helix

2.1. The basic equation

The theory of diffraction by helical molecules has played an important part in determining the conformation of many biological molecules. It was first developed by Cochran *et al.* (1952) and by Stokes (unpublished work). The diffraction produced by a discontinous helix can be calculated with the expression

$$G\left(R, \psi, \frac{l}{c}\right) = \sum_n \sum_j f_j J_n(2\pi R r_j) \exp\left\{i\left[n\left(\psi - \varphi_j + \frac{\pi}{2}\right) + \frac{2\pi l z_j}{c}\right]\right\}, \quad (1)$$

with the restriction $nK + mN = l$, where n and m are whole numbers, G is the structure factor in a point of reciprocal space with cylindrical coordinates R, ψ and l/c, l is the layer-line number, c is the crystallographic true repeat, f_j is the atomic structure factor, J_n is the Bessel function of order n, r_j, φ_j and z_j are real-space atomic coordinates, N is the number of residues in the true crystallographic repeat c and K is the number of helical turns in c.

2.2. Average intensities

FibreHelix can calculate diffraction for any value of ψ, a situation encountered when the helical molecules have crystalline order. More frequently, the molecules in fibres have a random orientation around the fibre axis. It is then convenient to calculate diffraction from the following expression, which directly gives the intensity averaged over ψ,

$$||G||^2 = \sum_n |C_{n(R,r,l,c)}|^2, \qquad (2)$$

where

$$C_n = \sum_j f_j J_n(2\pi R r_j) \exp\left\{i\left[n\left(-\varphi_j + \frac{\pi}{2}\right) + \frac{2\pi l z_j}{c}\right]\right\}.$$

2.3. Coiled coils

A coiled coil (or superhelix) is a geometry in which the axis of the molecular helix (called the minor helix) follows a helical path (called the major helix). Therefore, it has helical symmetry as a simple helix. A fibre of a coiled coil can thus also be modelled as a helix. In order to obtain a coiled coil, the helical residue is chosen as a section of the minor helix that follows the major helix, as suggested by Fraser *et al.* (1964*a*).

2.4. Packing disorder

Helical molecules are often arranged in either fibres or partially ordered crystals with their axes parallel to the c direction of the fibre, but randomly translated/rotated. In such a situation there are Bragg reflections along the equator but continuous diffraction along the

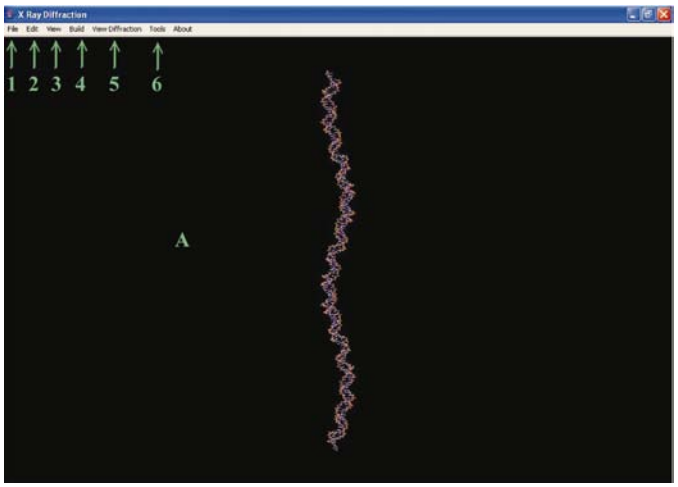

Figure 1
The main program window is shown for a DNA coiled-coil model. 'A' is the section where the three-dimensional interactive model of the molecule is rendered. '1' allows loading and saving coordinates in PDB format. '2' allows changes in the cell size. '3' contains various options for viewing the model rendered in 'A'. '4' is used to introduce the helical parameters and '5' calculates the X-ray diffraction pattern. Finally, '6' contains a comparator of images that can draw two different images one on top of another in different colours in order to quickly spot the differences between them.

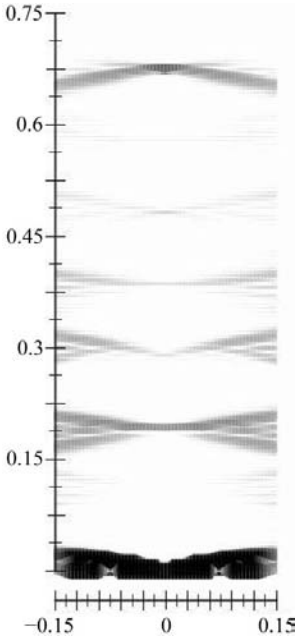

Figure 2
Diffraction pattern (Å^{-1} units) obtained by *FibreHelix* from a coiled coil of polyalanine with 126 amino-acid residues per turn of the major helix, which has a pitch of 186 Å. The coordinates given by Fraser *et al.* (1964*a*) have been used.

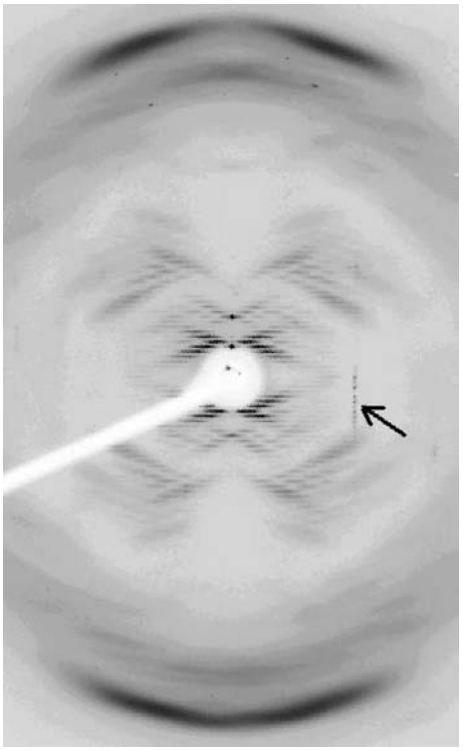

Figure 3
Rotation pattern (3°) obtained from a crystal of the duplex d(ATATATATATAT)$_2$, which associates through sticky ends to form a DNA coiled coil. Some Bragg reflections are also visible (arrow). The inclination of the stacking reflections indicates that the DNA duplexes are inclined about 19° with respect to the axis of the crystal, which is approximately vertical in the figure. The crystal appears to be a mosaic of molecules with three-dimensional order that give rise to the Bragg spots and randomly displaced molecules that give rise to continuous diffraction on layer lines. Layer lines can be clearly differentiated up to approximately $l = 30$.

layer lines. Intermediate degrees of order are also possible (Arnott, 1973). In DNA coiled coils such a situation can be modelled (as shown in Fig. 4) by excluding the terms with $m = 0$ in (1), a possibility that has also been introduced into our program.

3. Implementation details

3.1. Solving the addition of infinite terms

Equation (1) has infinite terms. From the computation point of view an infinite addition cannot be carried out, so some simplification is needed.

Thanks to the shape of Bessel functions, this simplification can be performed without altering the result. Bessel functions have an associated order. Only for order zero does it have a value different from zero at the origin; all other orders have a flat zone near the origin. The higher the order, the longer the flat zone is. To take advantage of this fact, we estimated the point where the function finishes its flat zone with a maximum error of 10^{-6}. We modelled it with the following linear higher approximation,

$$\begin{cases} x \leq 10 & n = 22 \\ 10 \leq x \leq 30 & n\lceil 1.255x + 9.5\rceil \\ 30 \leq x & \end{cases}$$

where $x = 2\pi R r_j$.

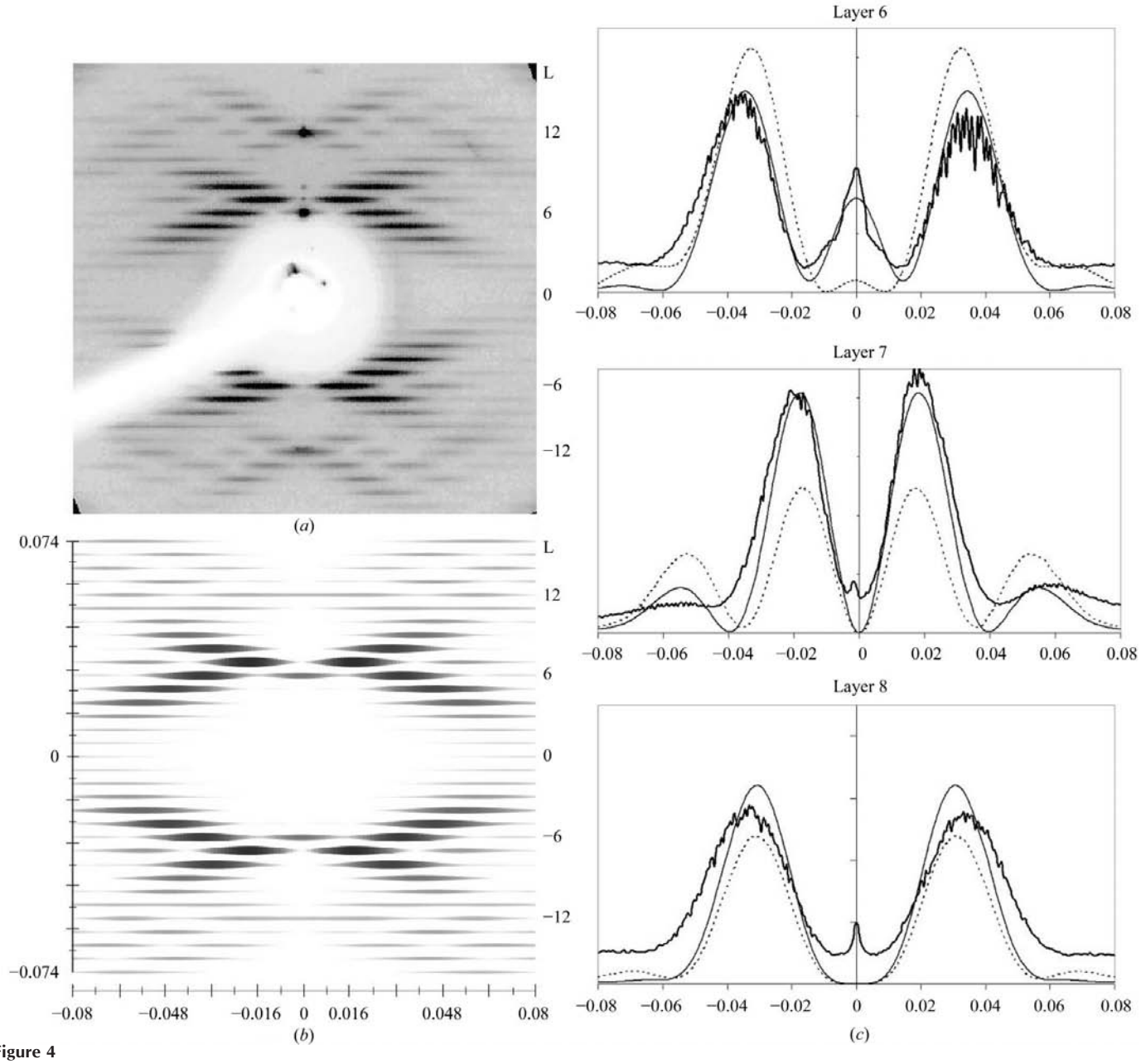

Figure 4
Comparison of the central region of the experimental diffraction (*a*) with the pattern calculated (*b*) with *FibreHelix* for a Hoogsteen DNA model (shown in Fig. 1). Disorder is assumed, so that diffraction in the equatorial region is not included. Meridional Bragg reflections (0 0 6 and 0 0 12) are also visible in the experimental diffraction pattern. Layer lines (L) are indicated. (*c*) Comparison of the experimental diffraction intensities for the three most intense layer lines of the Hoogsteen and Watson–Crick models. The latter was built from the experimental coordinates of a crystallized $(ATATAT)_2$ sequence (Yoon *et al.*, 1988). The experimental curves are given in black as heavy lines, the Hoogsteen model as continous lines and the Watson–Crick model as dotted lines.

3.2. Scattering factors

Scattering factors for the most common atoms (about 50 different atoms) have been approximated as described in Wilson & Price (1999). For ions, the most common form in a biological context has been selected.

3.3. Use of the program

The program can be used in any computer through a Java interface. It has a main window in which instructions can be introduced and the results viewed, as indicated in Fig. 1. Coordinates of the residues must be entered in PDB format. The program will be available on request from JAS.

4. Examples

4.1. Polyalanine coiled coil

Fraser *et al.* (1964*a*) modelled keratin with a polyalanine model of seven amino-acid residues. A coiled coil can be produced from 18 such units, which generate a coiled coil with pitch 186 Å and $7 \times 18 = 126$ amino-acid residues. The result of the simulation with *FibreHelix* is shown in Fig. 2. The typical amino-acid repeat around 1.5 Å (0.667 $Å^{-1}$) is clearly visible in the figure, as well as the expected splitting of layer lines in the 0.2 $Å^{-1}$ region. The equatorial region shows very high intensities which are not present in the experimental pattern (Fraser *et al.*, 1964*b*) as a result of sampling owing to the crystalline arrangement of keratin fibrils. Scattering in the equatorial region can be eliminated if required by using the 'screw disorder' option of *FibreHelix*.

4.2. DNA coiled coils

In our laboratory, we have found that short DNA oligonucleotides with sticky ends form partially ordered structures which apparently arise from the formation of coiled coils. Full details will be reported elsewhere (Subirana *et al.*, in preparation). An example of a diffraction pattern is given in Figs. 3 and 4. It shows a mixture of Bragg reflections and continuous layer lines. With the help of the program *FibreHelix*, we have interpreted the layer-line intensities as arising from a DNA coiled-coil structure such as that shown in Fig. 1.

The intensity of the layer lines is found to emerge from the meridional region around the seventh and 12th layer lines. These regions correspond to the first and second layer lines of a standard DNA double helix. Thus, the observed layer-line pattern arises from splitting of the layer lines of the original DNA helix, as predicted by the theory for coiled coils (Fraser *et al.*, 1964*a*; Wilson, 1966). Comparison of the experimental pattern with that predicted by *FibreHelix* shows a striking overall similarity (Fig. 4).

For a more quantitative analysis, the intensities of the three most intense layer lines calculated for two different models is also shown in Fig. 4. We first adjusted the average radius of the coiled coils so that the calculated curves appeared at the position of the experimental peaks. Comparison shows that the model with Hoogsteen base pairs (Abrescia *et al.*, 2004) gives a better correspondence with the experimental results, since the seventh layer line is the strongest found experimentally. Instead, the model with Watson–Crick base pairs has a stronger sixth layer line. However, on the basis of this information alone we cannot conclude which is the best molecular model, in particular since the conformation of the junction between the sticky ends has not been accurately modelled.

We are thankful to Drs I. Fita and J. Navaza and Mr S. Gestí for advice and discussion. This work was supported by grant BIO2002-00317 from the Ministerio de Ciencia y Tecnología. This program has been developed as part of the master thesis (PFC, in Catalan) of David Domínguez Sal.

References

Abrescia, N., González, C., Gouyette, C. & Subirana, J. (2004). *Biochemistry*, **43**, 4092–4100.
Arnott, S. (1973). *Trans. Am. Crystallogr. Assoc.* **9**, 31–56.
Cochran, W., Crick, F. & Vand, V. (1952). *Acta Cryst.* **5**, 581–586.
Fraser, R., MacRae, T. & Miller, A. (1964*a*). *Acta Cryst.* **17**, 813–816.
Fraser, R., MacRae, T. & Miller, A. (1964*b*). *J. Mol. Biol.* **10**, 147–156.
Knupp, C. & Squire, J. (2004). *J. Appl. Cryst.* **37**, 832–835.
Wilson, A. & Price, E. (1999). *International Tables for Crystallography*, Vol. *C*, edited by E. Prince, pp. 572–574. Dordrecht: Kluwer Academic Publishers.
Wilson, H. (1966). *Diffraction of X-rays by Proteins, Nucleic Acids and Viruses.* London: Edward Arnold.
Yoon, C., Privé, G., Godsell, D. & Dickerson, R. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 6332–6336.