X-Ray Crystallography of DNA-Drug Complexes

in press
Methods in Enzymology

Submitted by

Mary Elizabeth Peek and Loren Dean Williams

School of Chemistry and Biochemistry
Georgia Institute of Technology
Atlanta, GA 30332

Keywords: Intercalation, Stability, Minor-Groove, Anticancer, etc.

Version: Thursday, June 8, 2000

INTRODUCTION

Macromolecular x-ray crystallography is the quintessential method for determining three-dimensional structures of biological macromolecules to atomic resolution. Several recent textbooks and reviews provide excellent practical and experimental treatments of protein crystallography.[1-3] DNA crystallography has the same fundamental theoretical and experimental underpinnings as protein crystallography. As in a protein crystallography experiment, DNA crystallography is performed by the following steps:

- Crystal growth,
- X-ray diffraction data collection and reduction,
- Phase determination by molecular replacement, multiple isomorphous replacement (MIR), or multiple wavelength anomalous diffraction (MAD) , and
- Refinement

However, DNA crystallography differs somewhat from protein crystallography in methods of crystal growth, data collection and reduction, and phase determination. Those distinctions will be the primary focus of this report.

CRYSTAL GROWTH

To initiate crystal growth, a solution is slowly brought to supersaturation.[4] Small aggregates act as

nuclei, allowing crystal growth to commence. One impediment to growth of high quality crystals is the prescriptive difference in optimum solution conditions for nucleation and for growth.

Optimum crystallization conditions favor conformationally stable and homogeneous populations and promote specific intermolecular interactions. DNA crystallization, in our view, is best understood by analogy with DNA condensation into toroids. DNA toroids are less ordered than crystals but more ordered than aggregates. The conditions that promote nucleation and growth of crystals can be anticipated by those required for nucleation and growth of toroids, with the caveat that condensation involves DNA polymers while crystallization involves DNA oligonucleotides. As reviewed by Bloomfield,[5] DNA condenses in the presence of polyamines such as spermine and spermidine, or trivalent inorganic cations such as cobalt hexamine. DNA also condenses in the presence of both alcohols and divalent cations. The majority of DNA structures contained within the Nucleic Acid Structural Database (NDB)[6] were obtained from crystals grown from solution conditions favorable to DNA condensation.

To find solution conditions favorable for crystal growth, one must perform a multi-parameter search. As

described in a previous review of DNA crystallography,[7] the parameters to be varied in a search for DNA crystallization conditions include (1) metal ions and polyamines, (2) type and concentration of DNA ligands, (3) alcohol and buffer, (4) pH, (5) temperature, and (6) "precipitating agents". In our laboratory, pH, temperature and precipitating agents are not generally early search parameters for growing DNA or DNA-drug crystals.

Differences between DNA and protein crystal growth arise from the polyanionic nature of DNA and the dependence of DNA conformation and stability on cations, as opposed to a less-ionic character of most proteins. Cations are the first and generally most important parameter varied during searches for DNA or DNA-drug crystallization conditions in our laboratory. Our initial screen is invariably a spermine versus magnesium grid, with all other components held fixed. The initial DNA/ligand ratio is fixed at 1.1 moles ligand per 1.0 mole DNA binding site. Once magnesium and spermine concentrations are optimized, a fine search of the DNA/ligand ratio is performed.

In addition to cations and polyamines, certain alcohols and buffers also appear to be important for crystallization of DNA and DNA-drug complexes. The NDB presently contains 402 DNA crystal structures that lack

protein. Slightly over half (208) of those crystals were grown from solutions containing both methylpentane diol (MPD) and Mg^{++}. By contrast, few DNA-protein crystals, only 13 of 346 structures contained in the NDB, have been obtained from solutions containing MPD. Of DNA crystal structures that lack protein, 65% were obtained from solutions containing cacodylate buffer. Only 7% of DNA-protein crystals were obtained from solutions containing cacodylate.

pH plays a different role in nucleic acid crystallization than in protein crystallization. It is not feasible to crystallize DNA under conditions near the isoelectric point. Unlike proteins, DNA does not contain functional groups that change ionization state around physiological pH. Important exceptions to this rule are encountered when cytosine is protonated within non-Watson-Crick base pairing schemes, such as hemiprotonated C-C base pairs,[8,9] and Hoogsteen base pairs,[10] and when DNA ligands contain functional groups that change ionization state near physiological pH.

One cannot unambiguously eliminate tradition or other biases as the origins of differential patterns in protein and DNA crystallization conditions. However, we believe that protein crystallization conditions in general, even

DNA-protein crystallization conditions, can provide poor models for DNA crystallization.

X-RAY DATA COLLECTION AND REDUCTION

The ultimate goal of the x-ray diffraction experiment is to determine the electron density, $\rho(x,y,z)$, for each atom in the macromolecule according to Equation 1.

Equation 1.

$$\rho(x,y,z) = (1/V) \Sigma_{hkl} |F(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha_{hkl}]$$

where V = unit cell volume, $|F(hkl)|$ = structure factor amplitude, hkl = Miller index, xyz = real space coordinates, and α_{hkl} = relative phase of reflection hkl. In an x-ray diffraction experiment, one measures intensities, $[I(hkl)]$, of many thousands of "reflections" using a CCD camera or an imaging plate. Intensities are converted to structure factor amplitudes $[|F(hkl)|]$ by $I(hkl) = |F(hkl)|^2$.

Dynamic Range. Success in the solution of a structure by MIR or MAD, and the accuracy of a final refined model are critically dependent on the accuracy of $|F(hkl)|$. It is inherently more difficult to accurately determine $|F(hkl)|$ from a DNA crystal than from a protein crystal because the dynamic range in $|F(hkl)|$ from DNA crystals can be

significantly greater than from protein crystals. Moreover, certain packing arrangements within DNA crystals cause greater dynamic range than other packing arrangements. Specifically the dynamic range of $|F(hkl)|$ is greater for DNA crystals with end-to-end stacking (i.e. with pseudo-infinite helical axis) than for other packing arrangements such as end-to-groove packing (i.e. with skewed helical axes).

This trend in dynamic range is illustrated in Table 1 and Figure 1 where $|F(hkl)|$ from three crystals of the same quality (1.4 Å resolution) are compared. DNA with a pseudo-infinite helical axis is compared with DNA with skewed helical axes, and with a globular protein (containing alpha helix plus beta sheet). To quantitate dynamic range, the reflections of these data sets were independently sorted by $|F(hkl)|$ and identified with the median and maximum amplitudes ($|F(hkl)|_{med}$ and $|F(hkl)|_{max}$). The ratio $|F(hkl)|_{med}/|F(hkl)|_{max}$ varies with dynamic range. As shown in Table 1, $|F(hkl)|_{med}/|F(hkl)|_{max}$ for the pseudo-infinite helical axis DNA is more than two times greater than that for the skewed helical axes DNA, and nearly three times greater than that for the globular protein. This trend is illustrated in detail in Figure 1. If one defines weak $|F(hkl)|$ as those that are less than $|F(hkl)|_{max}/20$,

then ~30% are weak in a globular protein, ~42% are weak in the skewed helical axis DNA, and ~80% are weak in the pseudo-infinite helical axis DNA.

The effects of a pseudo-infinite helical axis of DNA on the diffraction pattern can been observed directly in the precession photograph shown in Figure 2. In this case (a DNA-porphyrin complex), the pseudo-infinite helical axis is directed nearly along the crystallographic *c-axis*. Extremely intense reflections with Miller indices 1 0 16, 1 0 -16, -1 0 16, and -1 0 -16 are observable in the photograph.

The origin of differences in dynamic ranges of |F(*hkl*)| from globular proteins and DNA crystals can be understood in part from a Patterson analysis. In a globular protein crystal, interatomic vectors generally have random lengths, directions and origins. By contrast, a B-DNA or intercalated B-DNA crystal will yield a large number of "stacking vectors" with conserved lengths (~3.4 Å, 6.8 Å, 10.2 Å, etc.), directions (along the helical axis), and semi-conserved origins (from the planes of base pairs). When the B-DNA is organized in an end-to-end fashion in the crystal (with a pseudo-infinite helical axis), then all the stacking vectors are very nearly aligned. Therefore a DNA Patterson map can contain intense

peaks spaced by ~3.4 Å. A Patterson map from a bis-intercalated DNA crystal (NDB entry DD0018)[11] with a pseudo-infinite helical axis is shown in Figure 3.

A Fourier decomposition of the electron density within a B-DNA or intercalated DNA crystal gives large amplitude waves with wavelengths of ~3.4 Å, ~6.8 Å, ~10.2 Å, etc. For example $F(hkl)_{max}$ in the data set collected by Hunter and coworkers on TGTACA-4'-epiadriamycin (Table 1) has the Miller index 0 0 16. Dividing the length of the crystallographic *c-axis* (52.39 Å) by l=16 gives a d-spacing for this reflection of 3.27 Å. The Miller plane normal is parallel to the *c*-axis*. The spacing and direction of this intense reflection are consistent with base-base and base-intercalator stacking, and alignment of the helical axis along the crystallographic *c-axis*. The second and third ranked F(*hkl*) in this data set have Miller indices 1 1 16 and 0 2 16 (3.25 Å d-spacing with Miller plane normals within 15° of the *c-axis*). Thus the Fourier decomposition of the electron density in this crystal contains three stacking waves of similar wavelength and direction. Those stacking waves have amplitudes nearly 40-fold greater than the median intensity in the data set and dominate the diffraction pattern. Their positions are tightly clustered in reciprocal space.

Data Collection Strategy. One of the goals of most diffraction experiments is to collect data to the highest resolution possible. The highest resolution data has the lowest intensities, requiring long collection times, intense x-ray beams and sensitive detectors. It is often impossible to collect weak high resolution data simultaneously with strong stacking reflections. One part of the strategy employed in our laboratory is to collect stacking intensities independently from high resolution data by collecting multiple data sets on the same crystal. The stacking intensities are collected with short exposures, low amperage on a rotating anode, or an attenuated the beam at a synchrotron. The short and long exposure data sets are scaled and merged. It is best practice to exclude the stacking reflections from the scaling, for example by merging data from 20 to 2.6 Å (short exposure) with data from 3.1 to 1.4 Å (long exposure). In this case the overlap, used for scaling, is 3.1 Å to 2.6 Å and would not contain the ~3.4 Å stacking reflections. A second part of our strategy is to identify the stacking reflections prior to initiating data collection, and to set the collection parameters to ensure accurate measurement of their intensities. Detector overload and peak overlap must be avoided by empirical

adjustment of exposure time and crystal-to-detector distance. The stacking reflections are in close proximity in reciprocal space, and are generally broader than other reflections.

Crystallographic Anisotropy. Elongated flexible molecules that lack lateral structural hooks exhibit characteristic types of crystalline disorder. In the precession photo in Figure 2, the diffraction pattern of a DNA-porphyrin complex is seen to extend out to the edge of the photo along *c**, but to fade out at lower resolution along *a**. Thus the diffraction pattern is anisotropic, indicating that the disorder within the crystal is anisotropic. The DNA is more highly ordered along the helical axis than along the perpendicular to the helical axis. These systematic errors in the data can be attenuated, once the refinement is near completion, by anisotropic or local scaling of observed-to-calculated data. One cost of these corrections is that they naturally introduce additional parameters to the refinement.

Additional information about disorder within the DNA-porphyrin crystal is provided by the faint elongated "X" pattern, stretching between the intense stacking reflections (Figure 2). This elongated "X" is reminiscent of a DNA fiber diffraction pattern. Our interpretation of

the origin of the elongated "X" in Figure 2 is that a fraction of the DNA within the crystal is disordered by random rotation about the helical axis, just as in a fiber.

STRUCTURE DETERMINATION

Molecular Replacement. Determination of helical axis orientation by Patterson analysis can be combined with symmetry information to simplify structure solution by molecular replacement. For example the Patterson map and strong 0 1 16 reflection indicate that the helical axis of the CGTACG-porphyrin complex is nearly parallel to the *c-axis*. Simple volume calculations and spectroscopic analysis of a dissolved crystal indicate that the asymmetric unit contains one strand of DNA plus one porphyrin molecule. If the DNA forms a duplex, then the duplex must be centered on a crystallographic two-fold axis. The DNA-porphyrin complex must be centered on one of the two crystallographic two-fold axes (in space group $P6_{1(5)}22$). Two possible orientations, which differ by 180°, are possible on each two-fold axis. Thus the molecular replacement search is limited to four one-dimensional translations. However, the success of molecular replacement always depends on an accurate search model. The CGTACG-porphyrin structure contains an unanticipated flipped-out base. Therefore molecular replacement failed in

this case, even though location and orientation of the complex was correctly anticipated.

MIR and MAD. Structure determination by MIR or MAD requires substitution with heavy atoms. Derivatives can be generated by de novo crystal growth of modified DNA. Cytosine and uracil can be substituted with bromine or iodine at the C5 position.[11-13] A search of the NDB indicates 30 bromine-substituted DNA fragments have been crystallized. Guanines are effective targets for soaking, for example by accepting platinum at the N7 position.[13]

SUMMARY

Here we have stressed important differences between protein and DNA crystallography. Crystal growth and data collection methodologies are not directly transferable between the two subfields. In addition, we note that analysis of symmetry and packing of DNA crystals can be useful and a uniquely aesthetic exercise.

ACKNOWLEGEMENTS

This work was supported by the National Science Foundation (Grant MCB-9976498) and the American Cancer Society (Grant RPG-95-116-03-GMC).

Table 1. Dynamic range in x-ray intensity data.

	Ident-ifier	res. (Å)	space group	Character-ization	$F_{med}(hkl)$ / $F_{max}(hkl)$	refer-ence
TGTACA-4'-epiadriamycin	NDB entry DDF035	1.4 Å	$P4_12_12$	Intercalat-ed B-DNA, pseudo-infinite helical axis	0.38	14
CGCGAATTCGCG	NDB entry BDl084	1.4 Å	$P2_12_12_1$	B-DNA, skewed helical axes	0.17	15
E. *coli* uracil DNA glycosylase	PDB entry 4eug	1.4 Å	$P2_12_12_1$	alpha helix and beta sheet	0.13	16

Figure Legends

Figure 1. Dynamic range of x-ray diffraction data from crystals of a DNA complex (TGTACA-4'-epiadriamycin) with a pseudo-infinite axis, a DNA fragment d(CGCGAATTCGCG) with a skewed helical axes, and a globular protein (E. Coli uracil DNA glycosylase). The graph shows plots of frequency of observation versus amplitude, normalized to the greatest amplitude in that data set. To obtain frequency of observation, $|F(hkl)|$ were sorted into bins of

$|F(hkl)|>0.05|F(hkl)|_{max}$,

$0.05|F(hkl)|_{max} >|F(hkl)|>.10|F(hkl)|_{max}$,

$0.10|F(hkl)|_{max} >|F(hkl)|>.15|F(hkl)|_{max}$,etc.

Figure 2. Precession photo of a crystal of a DNA porphyrin complex CGATCG-CuTMPyP4 (Cu(II)meso-(4-N-tetramethylpyridyl)porphyrin, NDB entry DDF060)[13]. This 6° screened precession photo, taken with Ni-filtered Cu-radiation, recorded the h0l layer. In this crystal (a=39.49 Å, c=56.15 Å, α=90°, γ=120°, space group $P6_122$) the pseudo-helical axis is very nearly parallel to the crystallographic *c-axis* causing extremely large |F(1 0 16)|. Anisotropy in crystalline order causes the diffraction pattern to fade out at lower angle along the *a*-axis* than along the *c*-axis*. The elongated cross is

caused by rotational disorder about the helical axis. The six-fold screw axis along the *c-axis* is indicated by systematic absences along the *c*-axis*.

Figure 3. Patterson map from a crystal of duplex $[d(CGTACG)]_2$ bound to the bis-intercalator D232 (*a* = 28.24 Å, *b* = 28.24 Å, *c* = 72.74 Å, α = 90.0°, β = 90.0°, γ= 120.0°).[11] The pseudo-infinite helical axis is parallel to the crystallographic *c-axis.*

References

1. McRee, D. E., *Practical Protein Crystallography*, Academic Press, Inc., New York (1993).
2. Drenth, J., *Principles of Protein X-Ray Crystallography*, Springer-Verlag, New York (1999).
3. Carter, C. W., Jr, and Sweet, R. M., *Macromolecular Crystallography*, Vol. A & B, Academic Press, New York (1997).
4. McPherson, A., *Preparation and Analysis of Protein Crystals*, Robert E. Krieger Publishing Co, Malabar, FL (1989).
5. Bloomfield, V. A., DNA condensation by multivalent cations, *Biopolymers, 44*, 269-82 (1997).
6. Berman, H. M., Zardecki, C., and Westbrook, J., The nucleic acid database: A resource for nucleic acid science, *Acta Crystallogr. Sect. D, 54*, 1095-1104 (1998).
7. Wang, A. H.-J., and Gao, Y.-G., Crystallization of Oligonucleotides and their Complexes with Antitumor Drugs, in *Methods, A Companion to Methods in Enzymology*, Vol. 1, Carter, C. W. J., Ed., Academic Press, Inc., New York, pp. 91-99 (1990).
8. Kang, C. H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R., and Rich, A., Crystal structure of

intercalated four-stranded d(C_3T) at 1.4 A resolution, *Proc Natl Acad Sci U S A, 91*, 11636-40 (1994).

9. Chen, L., Cai, L., Zhang, X., and Rich, A., Crystal structure of a four-stranded intercalated DNA: d(C_4), *Biochemistry, 33*, 13540-6 (1994).
10. Quigley, G. J., Ughetto, G., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., and Rich, A., Non-Watson Crick C•G and A•T Base Pairs in a DNA-Antibiotic Complex, *Science, 232*, 1255-1258 (1986).
11. Shui, X., Peek, M. E., Lipscomb, L. a., Gao, Q., Ogata, C., Roques, B. P., Garbay-Jaureguiberry, C., and Williams, L. D., Effects of Cationic Charge on Three-Dimensional Structures of Intercalative Complexes: Structure of a *bis*-Intercalated DNA Complex Solved by MAD Phasing, *Current Med. Chem., 7*, 59-71 (2000).
12. Kielkopf, C. L., Erkkila, K. E., Hudson, B. P., Barton, J. K., and Rees, D. C., Structure of a photoactive rhodium complex intercalated into DNA, *Nat Struct Biol, 7*, 117-21 (2000).
13. Lipscomb, L. A., Zhou, F. X., Presnell, S. R., Woo, R. J., Peek, M. E., Plaskon, R. R., and Williams, L. D., Structure of a DNA-Porphyrin Complex, *Biochemistry, 35*, 2818-2823 (1996).

14. Leonard, G. A., Brown, T., and Hunter, W. N., Anthracycline binding to DNA. High-resolution structure of d(TGTACA) complexed with 4'-epiadriamycin, *Eur J Biochem, 204*, 69-74 (1992).

15. Shui, X., McFail-Isom, L., Hu, G. G., and Williams, L. D., The B-DNA dodecamer at high resolution reveals a spine of water on sodium, *Biochemistry, 37*, 8341-8355 (1998).

16. Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T., and Gilliland, G. L., Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited, *Proteins, 35*, 13-24 (1999).

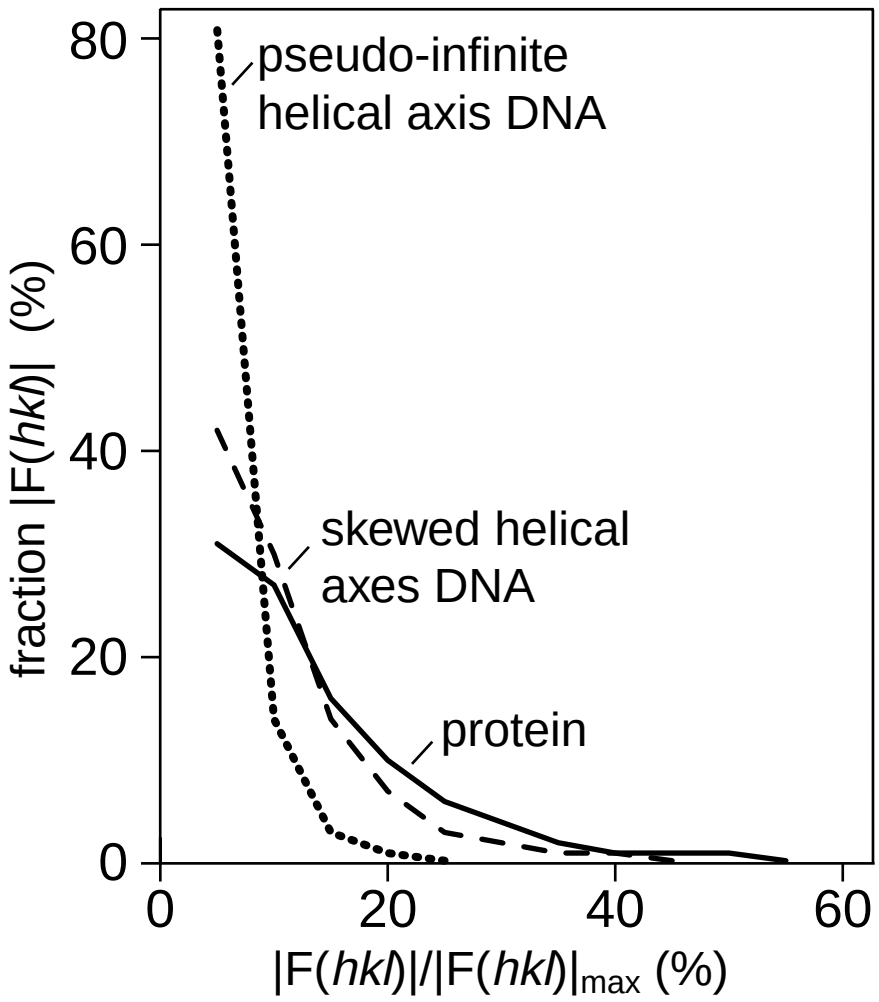
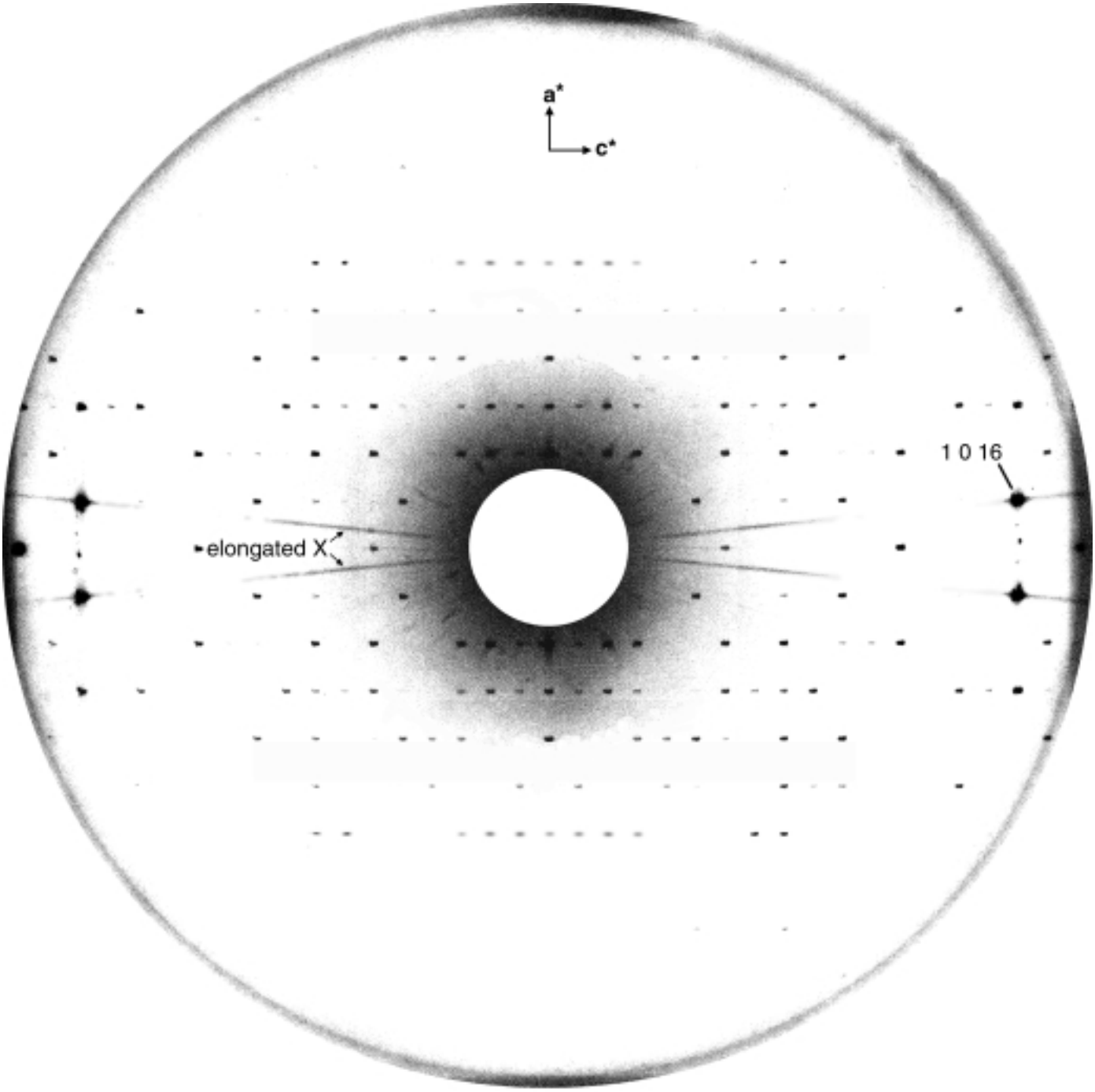


Figure 1
Peek and Williams



Peek and Williams
Figure 2

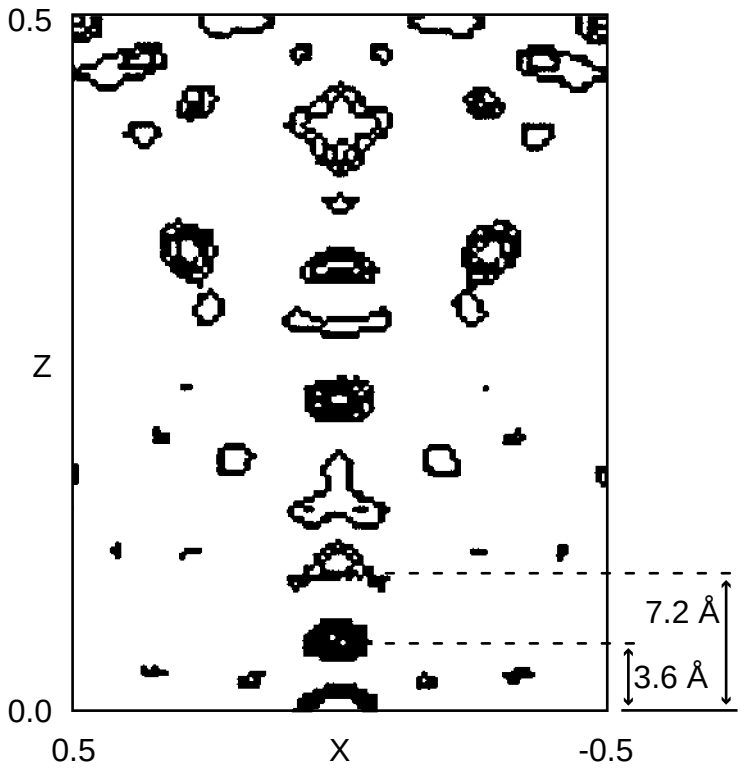


Figure 3
Williams and Peek