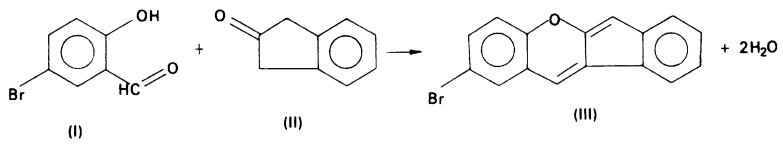
Examples of Crystal Structure Determination

9

9.1 Introduction

In this chapter we draw together, by means of actual examples, some of the material presented earlier in the book. It may be desirable for the reader to refer back to the previous chapters for descriptions of the techniques used, since we shall present here mainly the results obtained at each stage.

The first three examples can be solved by either the heavy-atom method or direct methods. Nowadays, it is quite commonplace to attempt the solving by direct methods of those structures which, at one time, would have been treated by the heavy-atom method. Where a powerful and sophisticated computer



package is available, direct methods frequently provide the most expeditious route to the solution of a structure. However, in order that we may illustrate the methods described, we shall use both techniques in this chapter.

9.2 Crystal Structure of 2-Bromobenzo[*b*] Indeno[1,2-*e*] Pyran [1]

2-Bromobenzo[*b*]indeno[1,2-*e*]pyran (BBIP) is an organic compound which was prepared by heating a solution in ethanol of equimolar amounts of 3-bromo-6-hydroxybenzaldehyde (I) and 2-oxoindane (II) under reflux in the presence of piperidine acetate. The two molecules condense with the elimination of two molecules of water. Upon recrystallization of the product from toluene, it has an m.p. of 176.5–177.0°C. Its molecular formula is $C_{16}H_9BrO$, and its classical structural formula is shown by III.

9.2.1 Preliminary Physical and X-Ray Measurements

The compound was recrystallized from toluene by slow, isothermal evaporation of the solvent at room temperature. The crystals were red, with an acicular (needle-shaped) habit, with the forms

M. Ladd and R. Palmer, *Structure Determination by X-ray Crystallography: Analysis by X-rays and Neutrons*, DOI 10.1007/978-1-4614-3954-7_9,

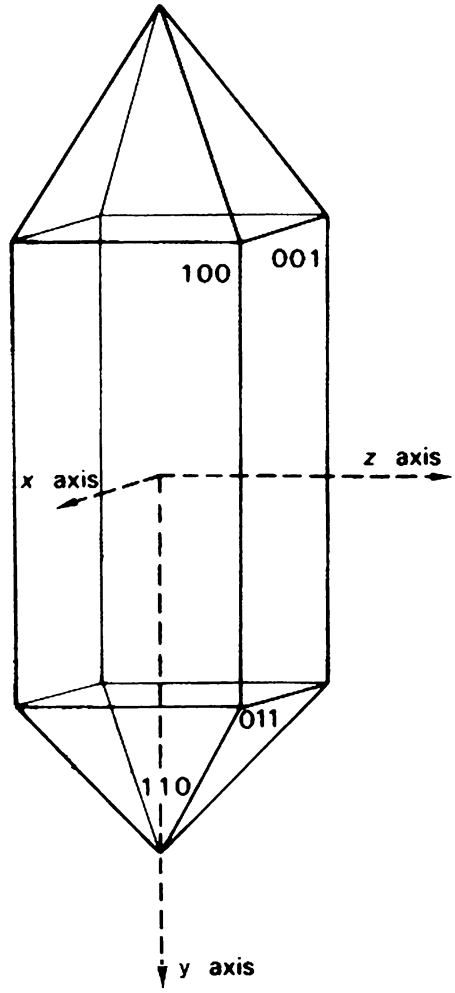


Fig. 9.1 Crystal habit of BBIP with the crystallographic axes drawn in; the forms shown are {100}, {110}, {001}, and {011}

(subsequently named) {100}, {110}, {001}, and {011} predominant, Fig. 9.1. The red color is characteristic of the chromophoric nature of a conjugated double-bond system.

The density of the crystals was measured by suspending them in aqueous sodium bromide solution in a stoppered measuring cylinder in a thermostat bath at 25°C. Water or concentrated sodium bromide solution, as necessary, was added to the suspension until the crystals neither settled to the bottom of the cylinder nor floated to the surface of the solution. Then, the crystals and solution were of the same density, and the density of the solution was measured with a pyknometer. A convenient variant here is to measure the refractive index of the final solution with an Abbe refractometer, having first prepared a graph of refractive index against density from data in the literature [2].

Under a polarizing microscope, the crystals showed straight extinction on (100) and (001), and oblique extinction (about 3° to a crystal edge) on a section cut normal to the needle axis (y). These observations suggested that the crystals were probably monoclinic. Although the diffractometer is now the main instrument of data collection, photographic work is still carried on in some parts of the world, and X-ray photographs can be useful in deciding upon the suitability of the crystal specimen and for

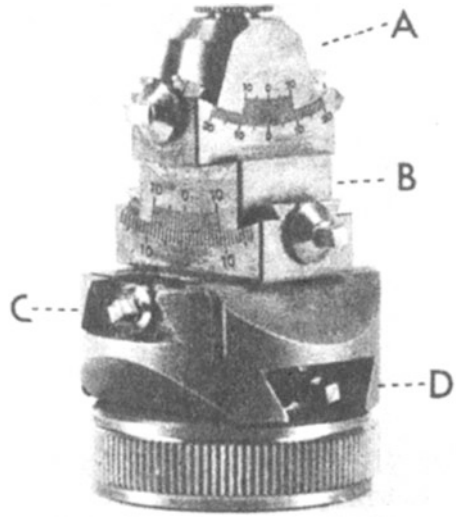


Fig. 9.2 Standard goniometer head; *A* and *B* are two arcs for angular adjustments; *C* and *D* are two sledges for horizontal adjustments (courtesy of Stoe et Cie)

information on its symmetry. In this example, we consider the use of photographic methods as a preliminary to the diffractometer measurements.

The crystals chosen for X-ray studies had the approximate dimensions 0.2, 0.4, and 0.3 mm parallel to *a*, *b*, and *c*, respectively. A crystal was mounted on the end of an annealed quartz fiber with "Araldite" (or "Eastman 910") adhesive and the fiber attached to an X-ray goniometer head, or arcs, Fig. 9.2, with dental wax. The arcs were affixed to a single-crystal oscillation camera, and the crystal was set with the needle axis accurately parallel to the axis of oscillation, first by eye and finally by X-ray methods. Copper $K\alpha$ radiation ($\lambda = 1.5418$ Å) was used throughout the work.

A symmetrical oscillation photograph taken about the *b* direction is shown in Fig. 9.3. The horizontal mirror symmetry line indicates that the Laue group of the crystal has an *m* plane normal to the needle axis. Further X-ray photographs, for example, the Laue photograph in Fig. 9.4, showed that the only axial symmetry was 2 parallel to *b*, thus confirming the monoclinic system for BBIP.

Notwithstanding Weissenberg photography is rarely employed nowadays, these photographs were used in this work, and a sample is shown in Figs. 9.5, 9.6, and 9.7. The straightforward indexing of the reflections can be understood with reference to Figs. 9.8 and 9.9. There are no systematic absences for the *hkl* reflections, so that the unit cell is primitive, but systematic absences do arise for *h*0*l* with *l* odd and for 0*k*0 with *k* odd. These observations confirm the monoclinic symmetry, and the systematic absences lead unambiguously to space group $P2_1/c$.

Measurements on the X-ray photographs gave the approximate unit-cell dimensions as:

$$a = 7.51 \text{ A}, \quad b = 5.96 \text{ A}, \quad c = 26.2 \text{ A}, \quad \text{and} \quad \beta = 92.5^\circ$$

The Bragg θ-angles of 20 high-order reflections of known indices, distributed evenly in reciprocal space, were measured to the nearest 0.01° on a four-circle CAD-4 diffractometer. From these data, the unit-cell dimensions were calculated accurately by the method of least squares; see Sect. 8.4ff. The complete crystal data are listed in Table 9.1. The calculated density D_c, for $Z = 4$, is in good agreement with the measured value D_m, which indicates a high degree of self-consistency in the

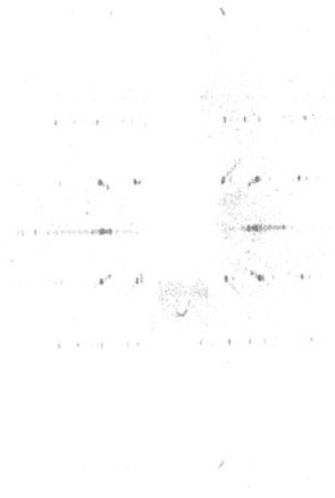


Fig. 9.3 Symmetrical oscillation photograph taken with the X-ray beam normal to *b*. The horizontal *m* line indicates an *m* plane in the Laue group of the crystal, normal to the axis of oscillation

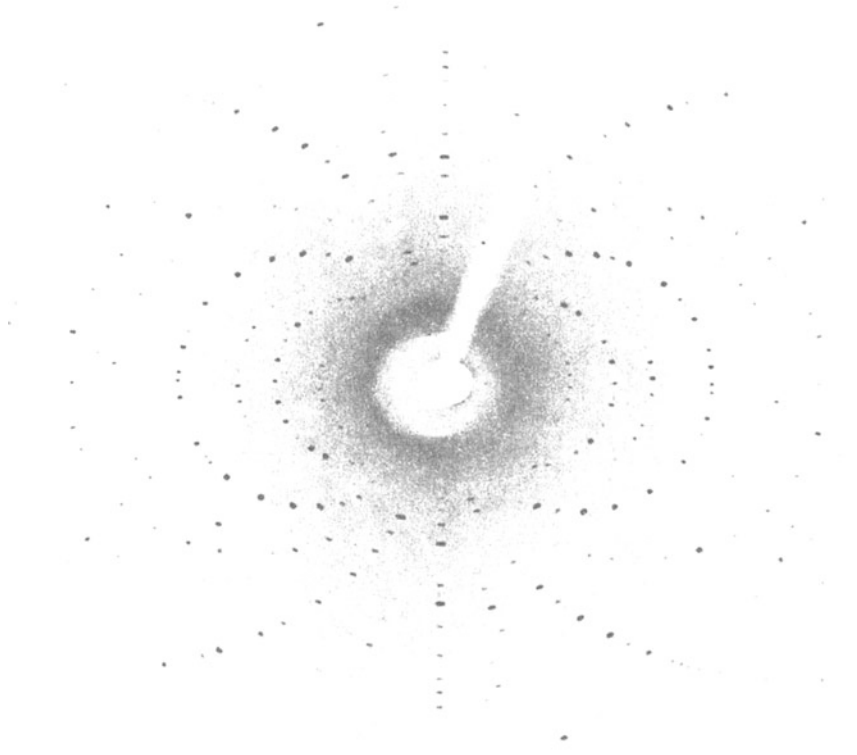


Fig. 9.4 Laue photograph taken with the X-ray beam along *b*, showing clearly the twofold symmetry axis along this direction

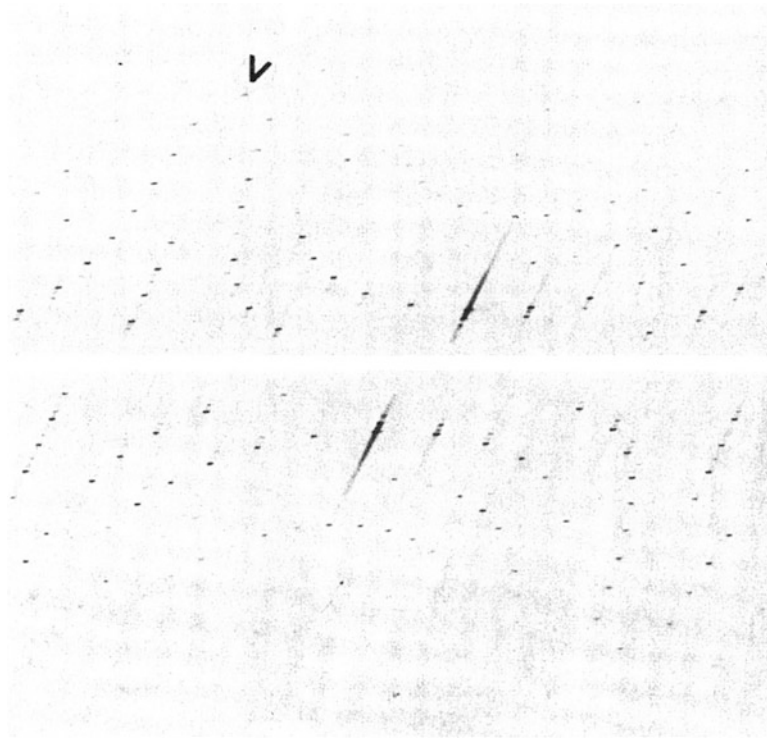


Fig. 9.5 Weissenberg photograph of the *h*0*l* layer. The more intense reflections show spots arising from both Cu $K\alpha$ ($\lambda = 1.5418$ Å) and Cu $K\beta$ ($\lambda = 1.392$ Å) radiations. In some areas, spots from W $L\alpha$ radiation ($\lambda = 1.48$ Å) arise due to sputtering of the copper target in the X-ray tube with tungsten from the filament; see Sect. 3.1.1. In each case, the spots for a given *h*0*l* reflection lie along the directions of axial rows at decreasing values of sin θ, in proportion to the change in λ. The continuous streaks (Laue streaks) arise from the "white" radiation; filtering (Chap. 3, Sect. 3.1.4) is never perfect

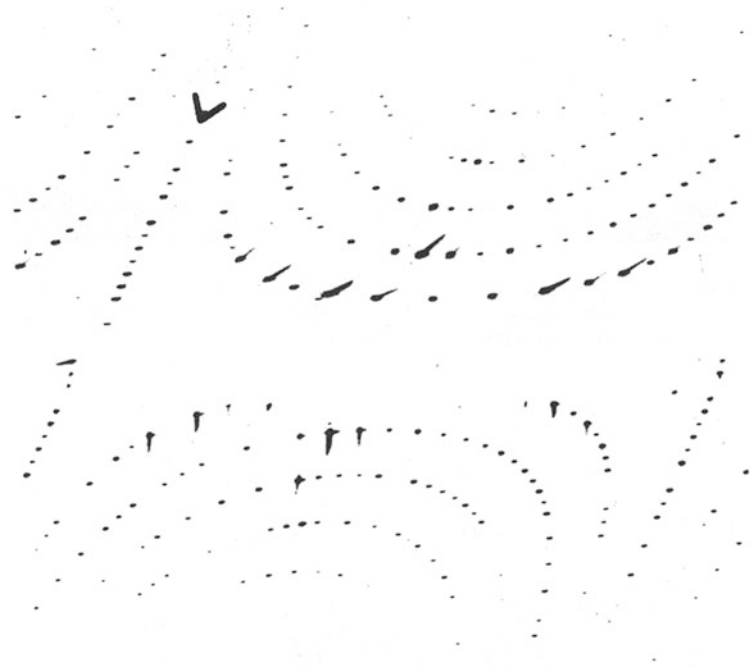


Fig. 9.6 Weissenberg photograph of the *h*1*l* layer. The 01*l* reciprocal lattice row, indicated by an arrowhead, illustrates clearly the effect of slight mis-setting of the crystal

Fig. 9.7 Upper part of a Weissenberg photograph of the 0*kl* layer, Cu $K\beta$ spots, can be seen for the more intense reflections. The 00*l* (z^*) reciprocal lattice row, indicated by an arrowhead, is common to this photograph and that of the *h*0*l* layer

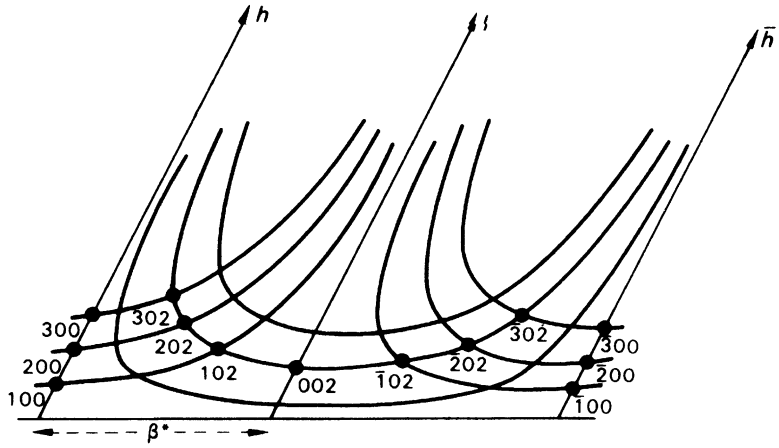


Fig. 9.8 Sample of indexed reflections on an *h*0*l* Weissenberg photograph diagram

parameters involved. The estimated standard deviations of the measured and calculated density values are important, because a significant difference ($>3\sigma$) may indicate the presence of unsuspected solvent of crystallization.

9.2.2 Intensity Measurement and Correction

Intensity data were collected on the diffractometer up to $\sin\theta = 65°$. The number N of data to be expected may be calculated from the formula $N = (4\pi/3)[2(\sin\theta_{max})/\lambda]^3 V_c/mG$, where mG is the number of general positions in the Laue group of the crystal, and the other symbols have their usual meanings. Since $mG = 4$ in the present example, $N = 1990$. From this number we subtract the number of systematic absences, 165, to give 1825. In practice, 1724 data were collected, indicating 101 accidental absences. Thus, a total of 1623 reflections were used for the structure analysis, giving ten reflections per parameter, assuming anisotropic temperature factors for the non-hydrogen atoms and a single scale factor in the least-squares refinement.

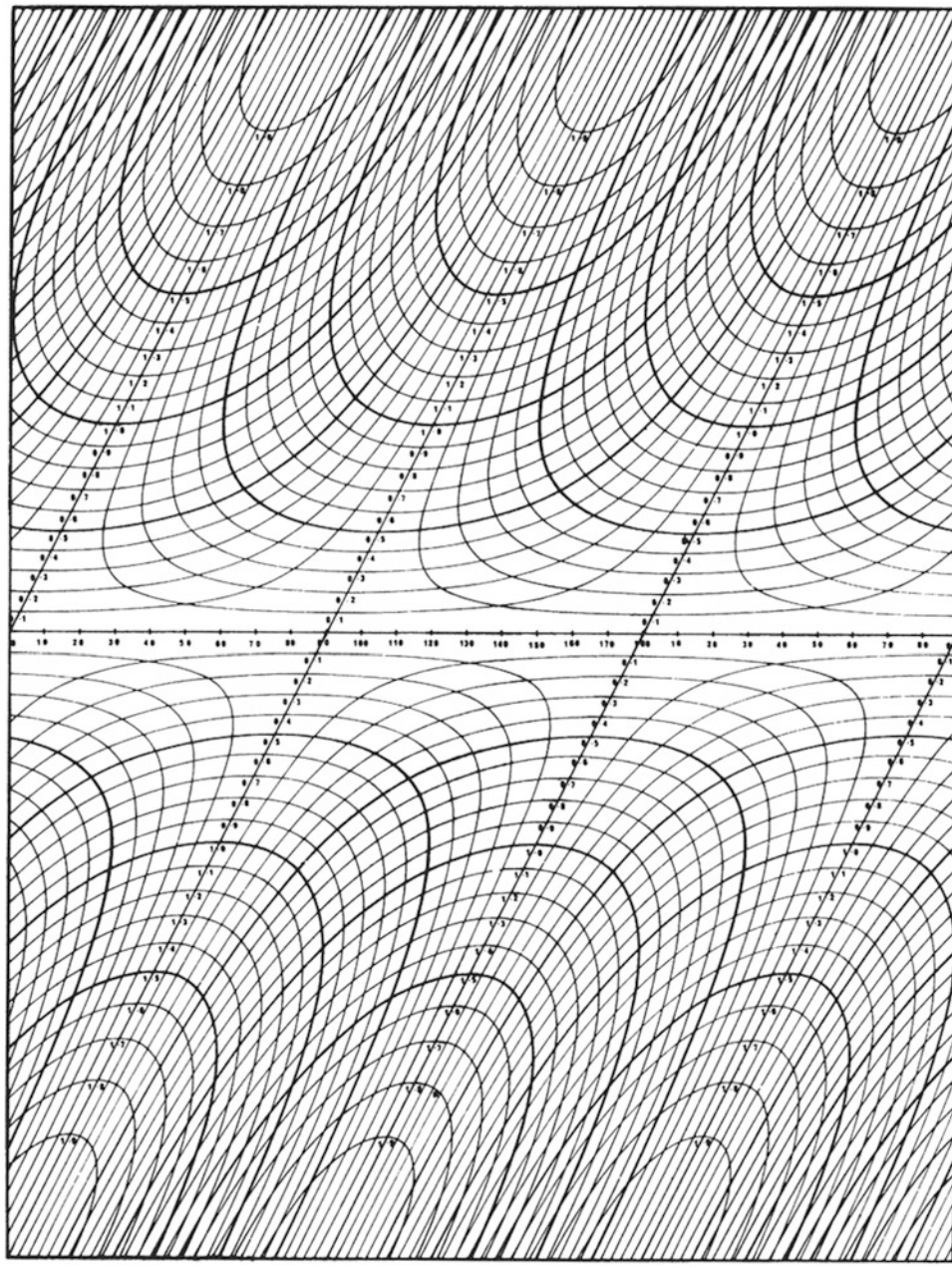


Fig. 9.9 Weissenberg chart: camera diameter 57.30 mm, 2° rotation per mm travel (reproduced with the permission of the Institute of Physics and the Physical Society, London)

The data were corrected for Lorentz and polarization effects, but not for absorption. Approximate scale K and overall isotropic temperature B factors were obtained by Wilson's method, Sect. 4.2.1. The parameters were fitted by least squares, and the line obtained had the equation

$$\ln\left\{\frac{\sum_j f_j^2}{\overline{F_o^2}}\right\} = -1.759 + 3.480\,\overline{\sin^2\theta} \tag{9.1}$$

Table 9.1 Crystal data for BBIP at 20°C[a]

Molecular formula	$C_{16}H_9BrO$
M_r	297.16
Space group	$P2_1/c$
a (Å)	7.508(4)
b (Å)	5.959(5)
c (Å)	26.172(6)
β (°)	92.55(2)
V_c (Å^3)	1,169(2)
Radiation (Cu $K\alpha$)	1.5418 Å
D_m (g cm^{-3})	1.68(1)
D_c (g cm^{-3})	1.688(3)
Z	4
F(000)	592

[a]The numbers in parentheses are estimated standard deviations, to be applied to the least significant figure

From the slope $(2B/\lambda^2)$ and intercept (2 ln K), $B = 4.1$ Å^2 and $K = 0.41$, where K is the scale factor for F_o^2. The graphical Wilson plot is shown in Fig. 9.10.

9.2.3 Structure Analysis in the *xz* Projection

This projection of the unit cell has the largest area and thus would be expected to show good resolution of the molecule. It is uncommon for a normal three-dimensional study to be preceded by an analysis in projection. However, from the standpoint of introductory study, carefully chosen two-dimensional examples have much to offer.

Using Fig. 2.32, we can associate the coordinates of the general equivalent positions

$$\pm \{x, y, z;\ x, 1/2 - y, 1/2 + z\}$$

with the four bromine atoms in the unit cell. In the xz projection, these coordinates give rise to two repeats within the length c, so we may consider this projection in terms of plane group $p2$ and compute the Patterson function from 0 to $a/2$ and 0 to $c/2$, which is equivalent to one-half of the unit cell in $p2$. This portion of the projection $P(uw)$ would be expected to show one Br–Br vector at $(2x, 2z)$, as in Fig. 9.11.

The two rows of peaks indicate that, in this projection, the molecules lie closely parallel to the z axis; this conclusion is supported by the large magnitude of |F(200)|, equal to 336. This value may be compared with F(000) in Table 9.1, and, more significantly, with $\sum_j f_{j,\theta_{200}}$, which is 474. The peak arising from the Br–Br vector is marked A, and by direct measurement we obtain the fractional coordinates $x = 0.25$, $z = 0.015$ for the Br atom in the asymmetric unit.

An electron density map, in this projection, was calculated using the signs given by F_{Br} with the experimental values of $F_o(h0l)$. If F_{Br} was less than one third of the corresponding value of F_o for any reflection, the sign was assumed to be uncertain and the reflection omitted from the electron density calculation at this stage of the analysis. Figure 9.12 shows the electron density map with the molecule, fitted with the aid of a model, marked in. The resolution is moderately good, and we can see that we are working along the right lines. From the shapes of the rings, it is apparent that the molecule is inclined to the plane of this projection, and there will be a limit to the improvement of the resolution attainable in this projection. Consequently, we begin three-dimensional studies.

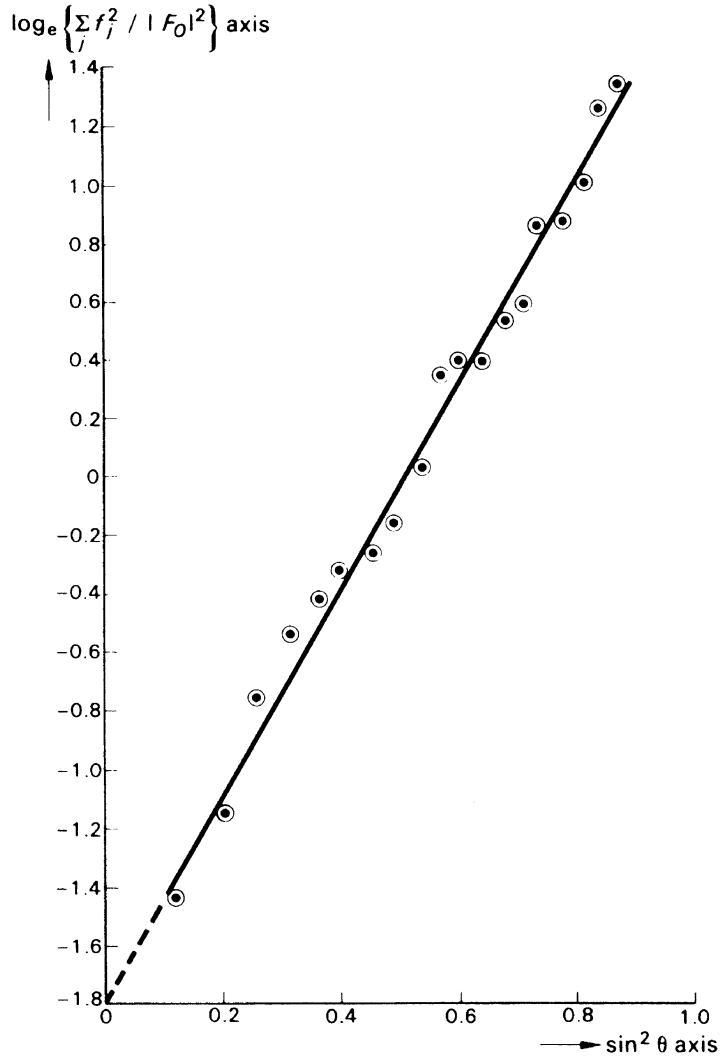


Fig. 9.10 Wilson plot for BBIP; the slope $(2B/\lambda^2)$ is 3.48 and the intercept $(2 \ln K)$ is −1.76

9.2.4 Three-Dimensional Structure Determination

In order to obtain values for all spatial coordinates, we proceeded first to a three-dimensional Patterson map $P(uvw)$, calculated section by section normal to the b axis.

The coordinates of the general positions show that the Br–Br vectors in the asymmetric unit will be found at $2x$, $2y$, $2z$ (single-weight peak), $0, \frac{1}{2} - 2y, \frac{1}{2}$ (double-weight peak), and $2x, \frac{1}{2}, \frac{1}{2} - 2z$ (double-weight peak). Hence, we must study the Patterson map carefully, particularly the Harker line $[0, v, \frac{1}{2}]$ and section $(u, \frac{1}{2}, w)$—why? Figures 9.13 and 9.14 show these two regions of Patterson space, and Fig. 9.15 illustrates a general section calculated close to the single-weight peak B. From the peaks B, C, and D, the coordinates for the bromine atom in the asymmetric unit were found to be 0.248, 0.188, 0.016.

Repeating the phasing procedure, but now for *hkl* reflections, and calculating a three-dimensional electron density map produced a good resolution of the complete structure, with the exception of the

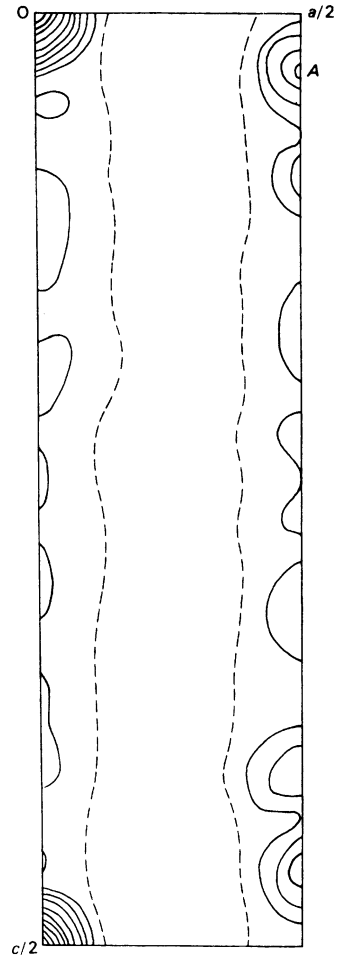


Fig. 9.11 Asymmetric unit of $P(uw)$. Since we are concerned here mainly with the Br–Br vector (A), the slight distortion arising from drawing β as 90° is inconsequential here

hydrogen atoms. Figure 9.16 illustrates a composite electron density map, which consists of superimposed sections calculated at intervals along a.

The scattering of X-rays by hydrogen atoms is small in magnitude, and these atoms are not normally resolved by the direct summation of the electron density. If the data are of good quality and all other atoms in the structure have been found, a difference-Fourier synthesis, Sect. 7.5.4, will generally result in the hydrogen atoms being located, provided the other atoms are not themselves too large in scattering power. It is possible also to calculate the positions of hydrogen atoms from the geometry of the structure, if the positions of sufficient surrounding groups are known. Nowadays, most program systems for X-ray structure analysis include routines for calculating the coordinates of hydrogen atoms, according to their nature, for example, $R_1R_2R_3CH$ and $R_1R_2CH_2$; the hydrogen atoms in a CH_3 group can be calculated if a position with respect to rotation about the R–CH_3 bond can be postulated.

Finally, we arrive at the complete structure for BBIP, as shown in Fig. 9.17 with a convenient numbering scheme.

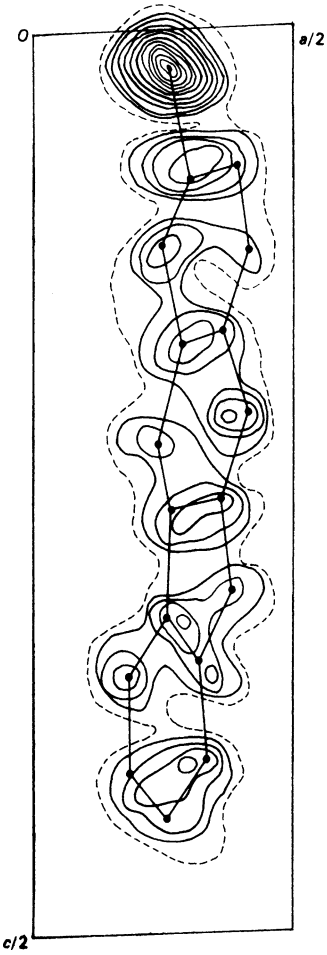


Fig. 9.12 Asymmetric unit of $\rho(xz)$ phased on the bromine atoms; the probable atomic positions are marked in

9.2.5 Refinement

During the final stages of refinement of the structure, the hydrogen atoms were included in the evaluation of the structure factors $|F_c|$, but no attempt was made to refine the parameters of the hydrogen atoms because the main interest in the problem lay in determining the molecular conformation. We note also that we have a favorable ratio of 10:1 for data to variables. Had we included the hydrogen atoms in the refinement, with isotropic temperature factors, the ratio would have been decreased to 9.4; see Sect. 8.4.3. The final adjustments of the structural parameters of the Br, O, and C atoms (x, y, z, and anisotropic temperature factors) and the scale factor were carried out by the method of least squares. The refinement converged with an R factor of 0.070, and a final difference-Fourier synthesis showed no fluctuations in density greater than about twice $\sigma(\rho_o)$, and then only around the position of the bromine atom. The analysis was considered to be satisfactory, and the refinement was terminated at this stage.

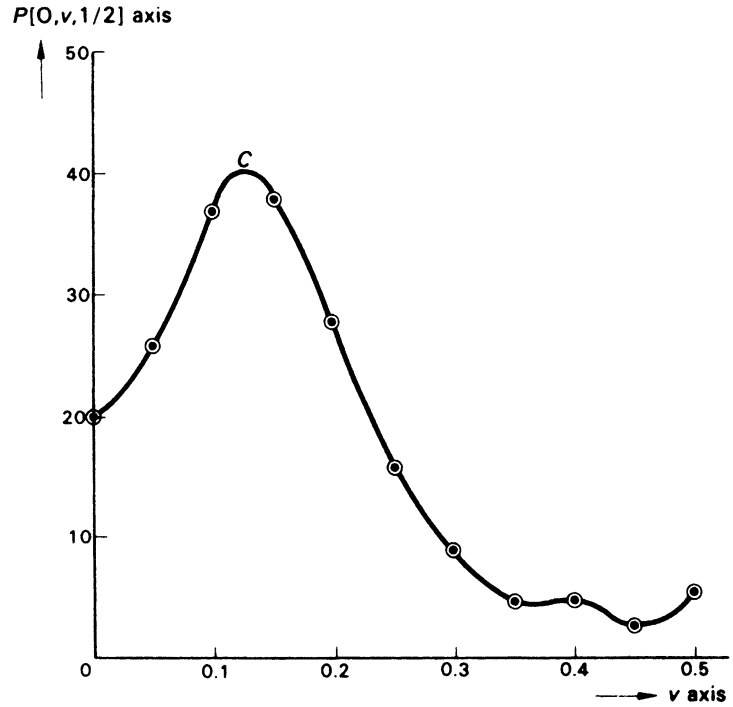


Fig. 9.13 Patterson function along the Harker line $[0, v, \frac{1}{2}]$, showing a double-weight Br–Br vector at C

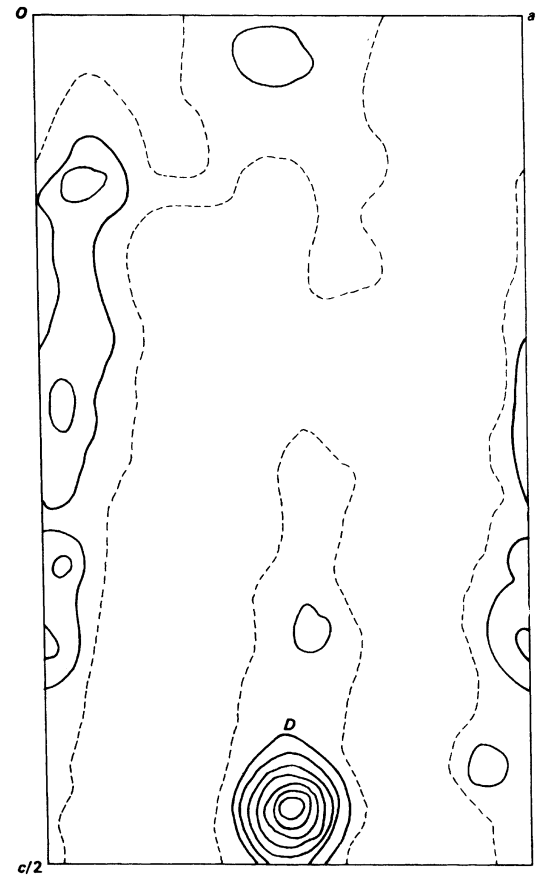


Fig. 9.14 Patterson section $(u, \frac{1}{2}, w)$ showing a double-weight Br–Br vector at D

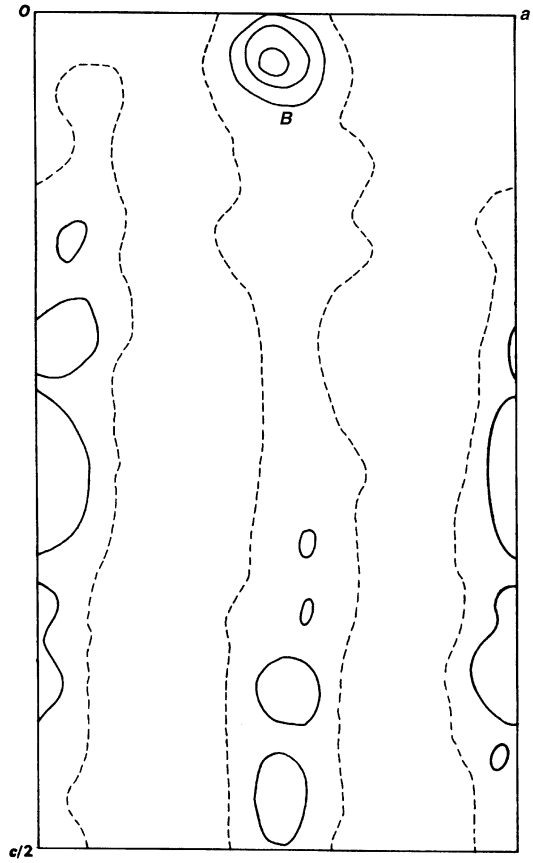


Fig. 9.15 Patterson section (u,0.375,w), showing a single-weight Br–Br vector B

9.2.6 Molecular Geometry

It remained to determine the bond lengths, bond angles, and other features of the geometry of the molecule and its relationship with other molecules in the unit cell.

From the coordinates of the atomic positions (Table 9.2) and using (7.119) and (7.121), bond lengths and angles were calculated. They are shown on the drawings of the molecule in Figs. 9.18 and 9.19. Figure 9.20 illustrates the packing of the molecules in the unit cell, as seen along b; the average intermolecular contact distance is 3.7 Å, a typical intermolecular contact distance in organic compounds, in which van der Waals forces link the molecules in the solid state.

In a molecule of this nature, the planarity of the ring system is of stereochemical interest. The equation of a plane, $Ax + By + Cz = D$, can be solved by three triplets x, y, z. Hence, the best molecular plane is obtained by a least-squares procedure that minimizes the sum of the squares of the deviations d of all of the atoms from the plane. The results are listed in Table 9.3. It can be seen that the deviations of the atoms from the best plane are not significant, and it is possible to conclude, therefore, that the introduction of the heteroatom has but little effect on the planarity of the benzofluorene moiety.

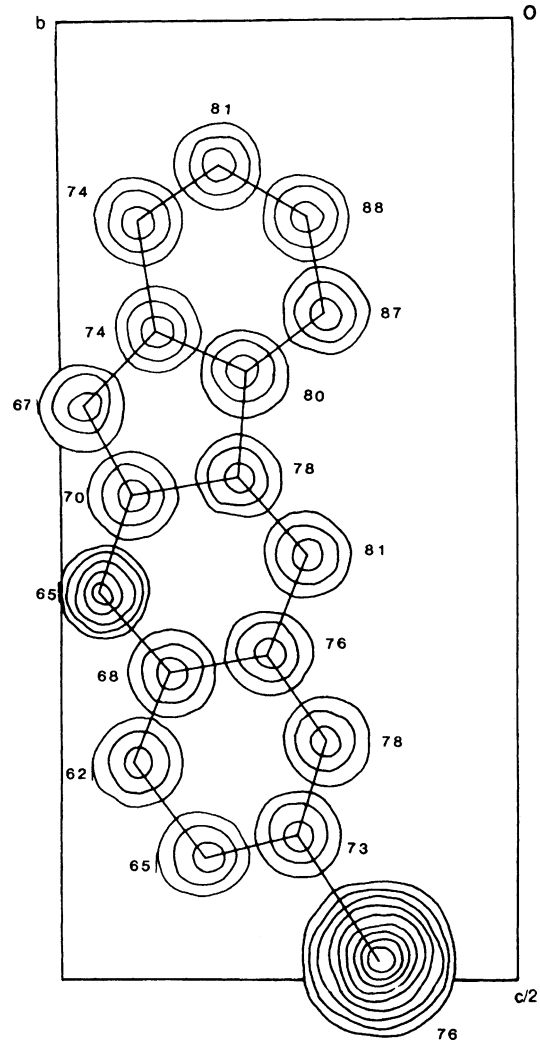


Fig. 9.16 Composite three-dimensional electron density map with the molecule (excluding H atoms) marked in, as seen along *a*. The contour of zero electron density is not shown, and the numbers represent 100*x* for each atom. A symmetry-related position to that chosen in Sect. 9.2.4 has been selected here for the Br atom (What is this symmetry operation?)

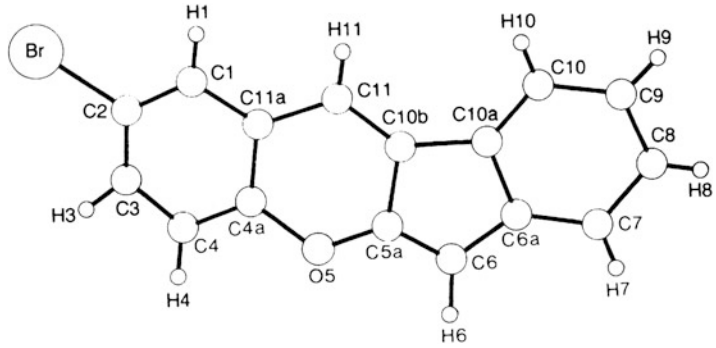


Fig. 9.17 Structural formula for BBIP

Table 9.2 Fractional atomic coordinates in BBIP, with *esd*s in parentheses[a]

Atom	*x*	*y*	*z*
Br	0.7602(2)	0.3152(3)	0.4848(0)
C(1)	0.7820(16)	0.4187(22)	0.3789(4)
C(2)	0.7310(16)	0.4951(24)	0.4252(4)
C(3)	0.6524(16)	0.7075(25)	0.4297(5)
C(4)	0.6214(16)	0.8413(23)	0.3871(5)
C(4a)	0.6794(16)	0.7619(22)	0.3406(4)
O(5)	0.6520(11)	0.9051(14)	0.2995(3)
C(5a)	0.6973(14)	0.8329(21)	0.2526(4)
C(6)	0.6714(14)	0.9397(19)	0.2077(5)
C(6a)	0.7384(15)	0.7990(19)	0.1678(4)
C(7)	0.7401(17)	0.8230(24)	0.1150(4)
C(8)	0.8078(18)	0.6526(24)	0.0858(4)
C(9)	0.8766(17)	0.4574(24)	0.1079(5)
C(10)	0.8731(16)	0.4268(21)	0.1605(4)
C(10a)	0.8035(16)	0.5954(20)	0.1908(4)
C(10b)	0.7767(14)	0.6076(21)	0.2454(5)
C(11)	0.8064(15)	0.4734(21)	0.2850(4)
C(11a)	0.7593(14)	0.5475(20)	0.3359(5)
H(1)	0.809	0.239	0.375
H(3)	0.622	0.789	0.460
H(4)	0.565	0.999	0.389
H(6)	0.630	0.121	0.208
H(7)	0.674	0.944	0.097
H(8)	0.804	0.667	0.043
H(9)	0.886	0.361	0.076
H(10)	0.809	0.253	0.375
H(11)	0.870	0.311	0.285

[a]There are no esd's for the hydrogen atom coordinates because these parameters were not included in the least-squares refinement

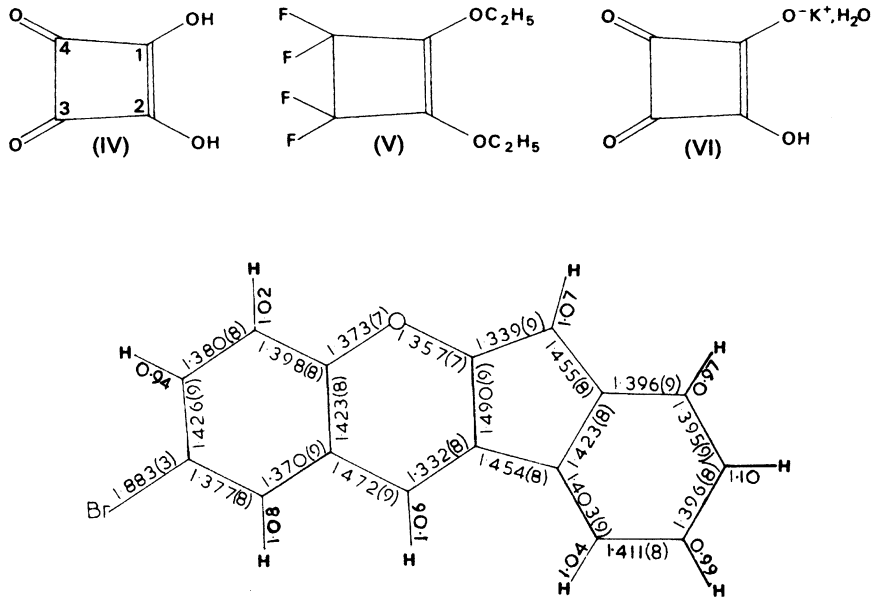


Fig. 9.18 Bond lengths in BBIP, with their estimated standard deviations in parentheses

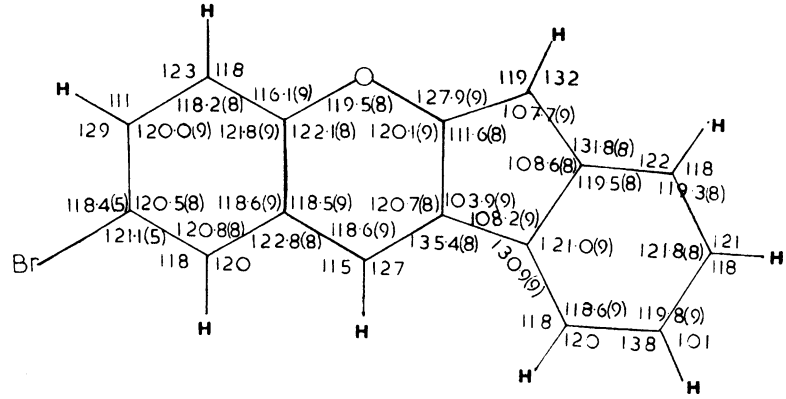


Fig. 9.19 Bond angles in BBIP, with their estimated standard deviations in parentheses

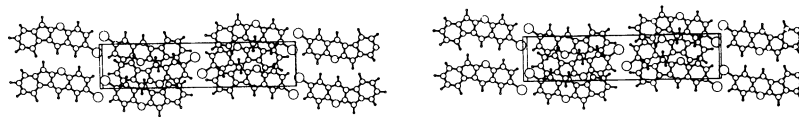


Fig. 9.20 Stereoview of the molecular packing in the structure of BBIP, as seen along *a*

Table 9.3 Deviations of atoms from the least-squares plane through the molecule[a]

Atom	Deviation (Å)
Br	0.03
C(1)	0.01
C(2)	−0.04
C(3)	−0.05
C(4)	−0.06
C(4a)	0.01
O(5)	0.05
C(5a)	0.06
C(6)	0.02
C(6a)	0.02
C(7)	−0.06
C(8)	−0.08
C(9)	−0.03
C(10)	0.03
C(10a)	0.04
C(10b)	0.04
C(11)	0.03
C(11a)	0.03

[a] The mean estimated standard deviation of the deviations is 0.02, so that hardly any atoms deviate significantly from the best plane at a 3σ level

9.3 Crystal Structure of Potassium 2-Hydroxy-3,4-Dioxocyclobut-1-ene-1-Olate Monohydrate (KHSQ) [3]

1,2-Dihydroxy-3,4-dioxocyclobut-1-ene (IV) may be prepared by the acid-catalyzed hydrolysis of 1,2-diethoxy-3,3,4,4-tetrafluorocyclobut-1-ene (V). On recrystallization from water, it has a melting point of 293°C, at which temperature it decomposes.

It has been called by the trivial name, squaric acid; the hydrogen atoms in the hydroxyl groups are acidic, and can be replaced by a metal. Potassium hydrogen squarate monohydrate (VI), which is the subject of this example, can be obtained by mixing hot, concentrated, equimolar aqueous solutions of potassium hydroxide and squaric acid and then cooling the reaction mixture.

9.3.1 Preliminary X-Ray and Physical Measurements

The compound was recrystallized from water as colorless, prismatic crystals with the forms {001}, {110}, and {100} predominant, Fig. 9.21. Under a polarizing microscope, straight extinction was observed on {001} and {100}, and an extinction angle of about 2° was obtained on a section cut normal to *b*. These results suggest strongly that the crystals belong to the monoclinic system. The density was measured in the manner indicated in Sect. 9.2.1.

The crystal specimen chosen for X-ray work had the dimensions 0.5, 0.5, and 0.3 mm parallel to *a*, *b*, and *c*, respectively. The details of the preliminary measurements are similar to those described for the previous example, and we list the crystal data immediately, Table 9.4. Copper $K\alpha$ radiation ($\lambda = 1.5418$ Å) was used throughout this work.

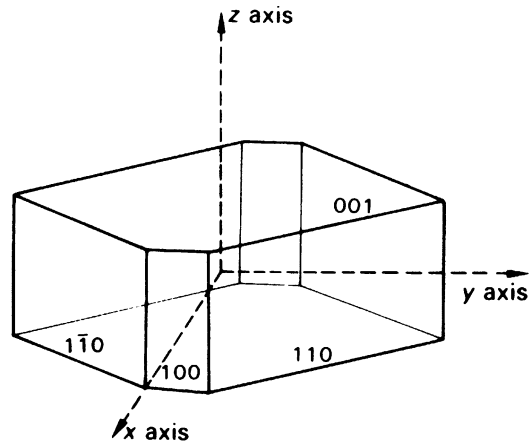


Fig. 9.21 Crystal habit of potassium hydrogen squarate monohydrate (KHSQ) with the crystallographic axes drawn in

Table 9.4 Crystal data for KHSQ at 20°C

Molecular formula	$C_4HO_4^-$, K^+, H_2O
M_r	170.17
Space group	$P2_1/c$
a (Å)	8.641(1)
b (Å)	10.909(1)
c (Å)	6.563(2)
β (°)	99.81(1)
V_c (Å^3)	609.6(2)
D_m (g cm^{-3})	1.839(7)
D_c (g cm^{-3})	1.854(1)
Z	4
F(000)	344

Table 9.5 Statistics of |E| values in KHSQ

	Acentric	Centric	This structure
$\overline{\lvert E\rvert^2}$	1.00	1.00	1.00
$\overline{\lvert E\rvert}$	0.89	0.80	0.81
$\overline{\lvert E\rvert^2 - 1}$	0.74	0.97	0.95
% $\geq$ 1.0	36.8	31.7	33.9
% $\geq$ 1.5	10.5	13.4	14.6
% $\geq$ 1.75	4.7	8.0	8.4
% $\geq$ 2.0	1.8	4.6	4.9
% $\geq$ 2.5	0.2	1.2	1.1

9.3.2 Intensity Measurement and Correction

Nine hundred symmetry-independent intensities with $\sin\theta/\lambda \leq 0.57$ A^{-1} were measured with a CAD-4 diffractometer. Corrections were applied for polarization and Lorentz effects, but not for absorption. Scale (K) and isotropic temperature (B) factors were deduced by Wilson's method, and the F_o data were converted to |E| values, see (8.1).

The structure analysis began with 142 |E| values $\geq$1.5, representing 15.8% of the experimental reflection data, and the |E| statistics are shown in Table 9.5. The agreement with the theoretical values for a centric distribution of |E| values is very close, in accord with the chosen centrosymmetric space group, Table 9.4.

9.3.3 $\sum_2$-Listing

The next stage was the preparation of a $\sum_2$ listing, Sect. 8.2.5. Symmetry-related reflections become very important in generating triplet relationships: 300 and 304 can lead to both 004 and 604, the latter

Table 9.6 Part of the $\sum_2$ listing for KHSQ[a]

h	|E(**h**)|	**k**	|E(**k**)|	**h** − **k**	|E(**h** − **k**)|	|E(**h**)|	|E(**k**)|	|E(**h** − **k**)|
531 (37)	2.6	010, 4	2.8	573	2.6	18.9		
		041	2.2	572	3.3	17.2		
		0041	2.0	570	2.7	14.0		
		114	2.3	625	1.7	10.2		
		032	1.7	563	2.0	8.8		
114 (45)	2.3	572	3.3	482	1.9	14.4		
		664	1.8	570	2.7	11.2		
		681	1.5	573	2.6	9.0		
		563	2.0	451	1.5	6.9		
		454	1.6	540	1.5	5.5		
032 (54)	1.7	531	2.6	563[b]	2.0	8.8		
		572	3.3	540	1.5	8.4		
		482	1.9	454	1.6	5.2		
		51	1.5	481[b]	2.0	5.1		
112 (39)	2.5	572	3.3	664	1.8	14.9		
		482	1.9	570	2.7	12.8		
		114	2.3	002	1.9	10.2		
		571	1.7	681	1.5	6.4		
010,4 (35)	2.8	332	2.2	372	1.9	11.7		
		625	1.7	681	1.5	7.1		
332 (46)	2.2	332	2.2	664	1.8	8.7		
		114	2.3	242	1.7	8.6		
		313	1.8	041	2.0	7.9		
		625	1.7	313	1.8	6.7		
002(25)	1.9	041	2.0	041	2.0	7.4		
		114	2.3	116	1.5	6.6		
		681	1.5	681	1.6	4.6		

[a]We use the notation **h** for hkl, **k** for $h'k'l'$, and **h** − **k** for $h - h'$, $k - k'$, $l - l'$
[b]Uses $03\bar{2}$

by replacing 304 by $\bar{3}04$, taking note of the phase symmetry. The relevant phase symmetry for space group $P2_1/c$ follows from (3.87), and may be summarized as follows:

$$s(hkl) = s(\bar{h}\,\bar{k}\,\bar{l}) \tag{9.2}$$

$$s(hkl) = s(h\bar{k}\bar{l})(-1)^{k+l} \tag{9.3}$$

A portion of the $\sum_2$ listing is shown in Table 9.6; full use of the symmetry relationships (9.2) and (9.3) has been made in setting up the listing. The numbers in parentheses under each **h** are the total numbers of $\sum_2$ triplets for each of these reflections; **k** and **h** − **k** represent those reflections forming a vector triplet with **h**.

9.3.4 Specifying the Origin

Following the procedure described in Sect. 8.2.2 and using the reflections in Table 9.6, three reflections were chosen and allocated positive signs, in order to fix the origin at 0,0,0. The symmetry relationships in the space group of this compound allowed, in all, 12 signs in the origin set, Table 9.7. The reader should check the signs, starting from the first one in each group of four, using (9.2) and (9.3).

Table 9.7 Origin-fixing reflections and their symmetry equivalents

Reflections	Sign	|E|	No. of $\sum_2$ triplets
531	+	2.6	37
531	+		
531	+		
531	+		
114	+	2.3	45
114	+		
114	–		
114	–		
032	+	1.7	54
032	+		
032	–		
032	–		

Table 9.8 Sign determination starting from the origin set

k	**h** − **k**	**h**	Indication for $s(\mathbf{h})$
531 +	114 –	625	–
531 +	$03\bar{2}$ –	563	–
$56\bar{3}$ –	114 +	451	–
$56\bar{3}$ –	032+	591	–
451–	$03\bar{2}$ –	481	+

9.3.5 Sign Determination

The $\sum_2$ listing was examined with a view to generating new signs, using (8.9), which may be given by

$$s[E(\mathbf{h})] \approx s\left[\sum_k E(\mathbf{k})E(\mathbf{h}-\mathbf{k})\right] \tag{9.4}$$

where the sum is taken over the several **k** triplets all involved with the given **h**. The probability of (9.4) is given by (8.10). If only a single $\sum_2$ interaction is considered, (9.4) becomes

$$s[E(\mathbf{h})] \approx s[E(\mathbf{k})sE(\mathbf{h}-\mathbf{k})]$$

and the probability calculation omits the summation given in (8.11).

Since the values of the probability $P_+(\mathbf{h})$ were sufficiently high, the signs were accepted as correct; very small or zero values of $P_+(\mathbf{h})$ indicated strongly a negative sign for **h**. Some examples of the application of (9.5) are given in Table 9.8. It does not matter which reflections in a triplet are labeled **h** and **k**.

Table 9.9 Symbolic signs

Reflection	Sign	|E|	No. of $\sum_2$ relationships
112	A	2.5	39
$\bar{1}\,\bar{1}2$	A		
$1\,\bar{1}\,\bar{2}$	$-A$		
$\bar{1}\bar{1}2$	$-A$		
010,4	B	2.8	35
$0\bar{1}0, \bar{4}$	B		
$\bar{1}0, 4$	B		
$0\bar{1}0,4$	B		
332	C	2.2	46
$\bar{3}\bar{3}2$	C		
$33\bar{2}$	$-C$		
$\bar{3}\bar{3}2$	$-C$		

Use of Sign Symbols

The above process of sign determination was applied to the entire $\sum_2$ listing which, although it contained 1276 triple products, was exhausted after only 24 signs had been found. To enable further progress to be made, three reflections were assigned the symbols A, B, and C, where each symbol represented either a plus or minus sign. Twelve symbolic signs, Table 9.9, were thus added to the set, and the sign determination was continued, now in terms of both signs and symbols. It may be noted that although the symbols are given to reflections with large |E| values and large numbers of $\sum_2$ interactions, there are not, necessarily, any restrictions on either parity groups or the use of structure seminvariants.

Some examples of this stage of the process are given in Table 9.10. The values of **h** and **k** are taken from either Tables 9.7 and 9.9, which constitute the "starting set," or as determined through (9.4). The reader is invited to follow through the stages in Table 9.10, working out the correct symmetry-equivalent signs from (9.2) and (9.3) as necessary.

From Table 9.10, we see that six more reflections have been allocated signs, and another 17 are determined in terms of A, B, and C. Multiple indications can now be seen. For example, there are two indications that $s(573) = B$, two indications that $s(570) = -$, and two indications that $s(540) = A$. Three indications for 041 suggest that both $s(041) = -$ and $A = -$.

Continuing in this manner, it was found possible to allot signs and symbols to all 142 |E| values greater than 1.5. The symbols A, B, and C were involved in 65, 72, and 55 relationships, respectively. Consistent indications, such as those mentioned above for $s(041)$, led finally to the sign relationships $A = AC = B = -$, from which it follows that $C = +$. It does not always turn out that the signs represented by symbols can be allocated from the analysis in that complete and satisfactory manner. If there are n undetermined symbols, then there will be, in general, 2^n sets of signs to be examined. In this case, figures of merit, such as those discussed in Sect. 8.2.13, can be used to indicate that set of signs most likely to be correct. It may not follow that the indicated set *is* correct, and some trials with Fourier syntheses may be needed at this stage in order to elicit the correct result.

9.3.6 The E Map

The signs of the 142 |E| values used in this procedure were obtained with a high probability, and an electron density map was computed using the signed |E| values as coefficients. The sections of this

Table 9.10 Further sign determinations[a]

k	$s(\mathbf{k})$	$(\mathbf{h}-\mathbf{k})$	$s(\mathbf{h}-\mathbf{k})$	**h**	Sign indication, $s\mathbf{h}$
010.4	B	625	$-$	681	$s(68\bar{1}) = B$
010.4	B	332	C	372	$s(372) = -BC$
112	A	681	B	571	$s(571) = AB$
531	$+$	010,4	B	573	$s(573) = B$
114	$+$	681	B	573	$s(573) = B$
332	C	332	C	664	$s(664) = CC = +$
114	$+$	664	$+$	570	$s(570) = -$
112	A	570	$-$	482	$s(482) = A$
112	A	664	$+$	572	$s(572) = -A$
114	$+$	572	$-A$	482	$s(482) = A$
032	$+$	572	$-A$	540	$s(540) = A$
032	$+$	482	A	454	$s(454) = -A$
114	$+$	454	$-A$	540	$s(540) = A$
332	C	114	$+$	242	$s(242) = -C$
112	A	114	$+$	002	$s(002) = A$
112	A	482	A	570	$s(570) = -AA = -$
570	$-$	531	$+$	041	$s(041) = -$
625	$-$	332	C	313	$s(313) = C$
313	C	332	C	041	$s(041) = -\mathrm{CC} = -$
531	$+$	041	$-$	572	$s(57\bar{2}) = +$
002	A	041	$-$	041	$s(041) = A$
002	A	681	B	681	$s(681) = AB$
002	A	114	$+$	116	$s(116) = A$

[a]Symmetry relations should be employed as necessary

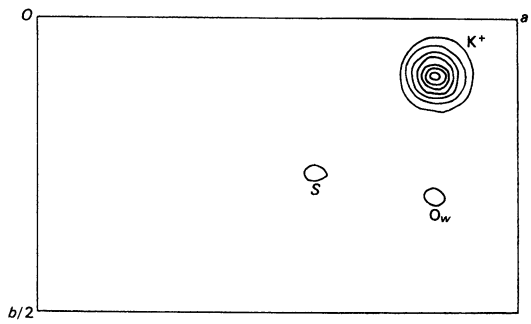


Fig. 9.22 E map for KHSQ at $z = 0.15$

map $\rho(xyz)$ at $z = 0.15, 0.20, 0.25$, and 0.30 are shown in Figs. 9.22, 9.23, 9.24, and 9.25. They reveal the K^+ ion and the $C_4O_4^-$ ring system clearly; the oxygen atom O_w of the water molecule was not indicated convincingly at this stage of the analysis. A tilt of the plane of the molecule with respect to (001) can be inferred from Figs. 9.23, 9.24, and 9.25. Some spurious peaks S may be seen. This is a common feature of E maps. We must remember that a limited data set (142 out of 900) is being used, and that the $|E|$ values are sharpened coefficients corresponding to an approximate point-atom model. The data set is therefore terminated while the coefficients for the Fourier series are relatively large, a procedure that can lead to spurious maxima; see Sect. 6.9.1. However, such peaks are often of smaller weight than those that correspond to true atomic positions.

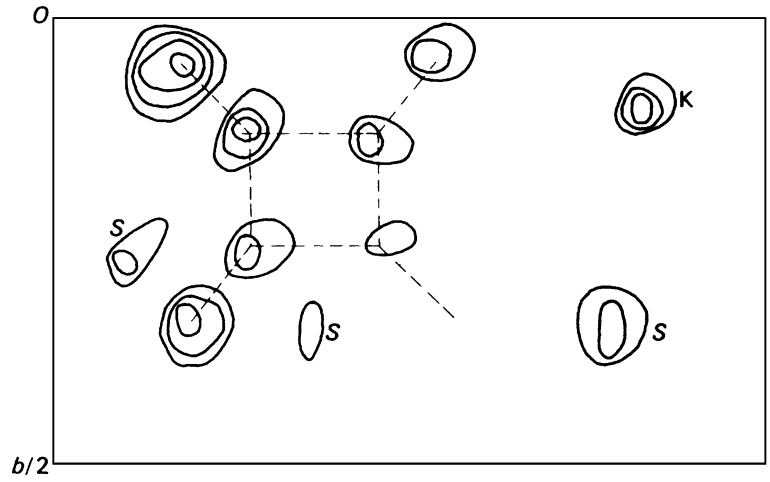


Fig. 9.23 E map for KHSQ at $z = 0.20$

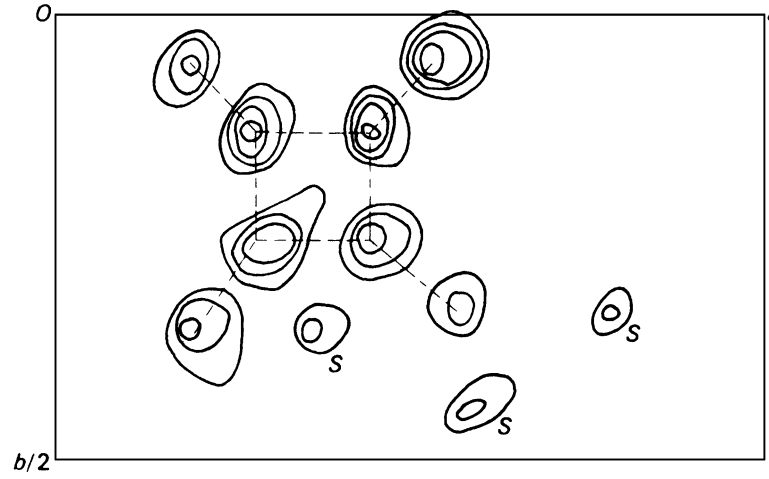


Fig. 9.24 E map for KHSQ at $z = 0.25$

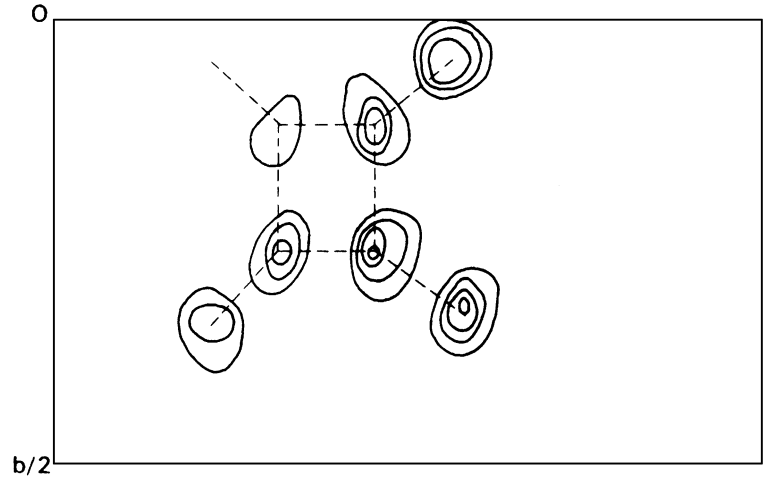


Fig. 9.25 E map for KHSQ at $z = 0.30$

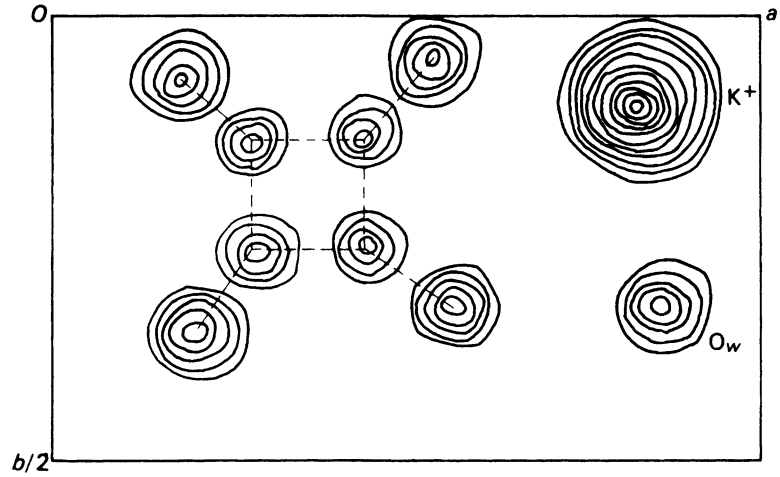


Fig. 9.26 Composite electron density map for KHSQ (excluding H atoms); the atomic coordinates are listed in Table 9.11

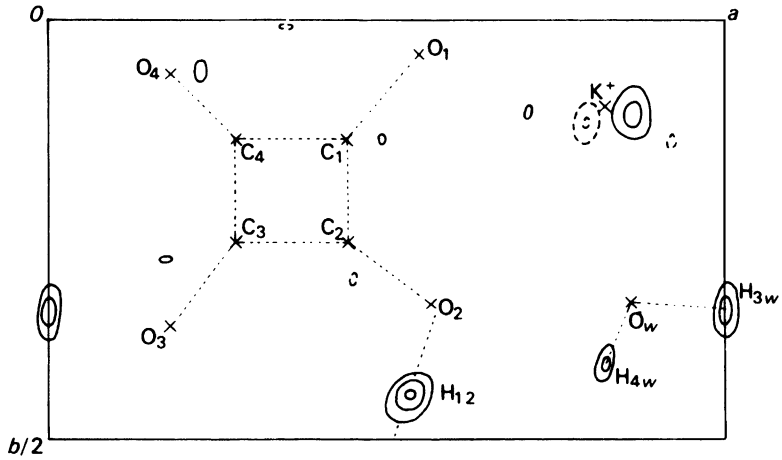


Fig. 9.27 Composite difference electron density map for KHSQ. Positive contours are solid lines and negative contours are broken lines. Bonds in the squarate ring and those involving hydrogen atoms are shown as dotted lines. Some spurious, small peaks (unlabeled) we shown by this synthesis

9.3.7 Completion and Refinement of the Structure

Sometimes, all atomic positions are not contained among the peaks in an E map. Those peaks that do correspond to chemically sensible atomic positions may be used to form a trial structure for calculation of structure factors and an F_o electron density map. A certain amount of subjective judgment may be required to decide upon the best peaks for the trial structure at such a stage.

This situation was obtained for KHSQ, although it was not difficult to pick out a good trial structure. Coordinates were obtained for all non-hydrogen atoms except the oxygen atom of the water molecule. The *R*-factor for this trial structure was 0.30, and the composite three-dimensional electron density map obtained is shown in Fig. 9.26, which now reveals O_w clearly. It may be noted in passing that the small peak labeled O_w in Fig. 9.22 corresponds to the position of this atom, but this fact could not be determined conclusively at that stage of the analysis.

Table 9.11 Fractional atomic coordinates for KHSQ

	x	*y*	*z*
K^+	0.8249(2)	0.1040(2)	0.1295(3)
C(1)	0.4353(9)	0.1295(7)	0.2572(12)
C(2)	0.4495(9)	0.2597(7)	0.2714(12)
C(3)	0.2795(9)	0.2714(8)	0.2462(11)
C(4)	0.2659(9)	0.1345(7)	0.2305(12)
O(1)	0.5399(6)	0.0450(5)	0.2649(10)
O(2)	0.5649(6)	0.3346(5)	0.2920(10)
O(3)	0.1874(7)	0.3582(6)	0.2386(10)
O(4)	0.1578(6)	0.0605(5)	0.2022(10)
O_w	0.8789(7)	0.3429(6)	0.0424(10)
H(12)	0.522	0.413	0.246
H(3*w*)	1.000	0.346	0.075
H(4*w*)	0.826	0.400	0.100

Further refinement was carried out by the method of least squares, and an *R*-factor of 0.078 was obtained. Figure 9.27 shows a composite three-dimensional difference-Fourier map for KHSQ. Peaks numerically greater than 0.5, representing about twice $\sigma(\rho_0)$, are significant, and have been contoured. Some of these peaks indicate areas of small disagreement between the true structure and the model. Three positive peaks, however, are in positions expected for hydrogen atoms. Inclusion of these atoms in the structure factor calculations in the final cycles of least-squares refinement had a small effect on the *R*-factor, bringing it to its final value of 0.077. The fractional atomic coordinates for the atoms in the asymmetric unit are listed in Table 9.11.

Interatomic distances and angles are shown in Fig. 9.28, and a molecular packing diagram, as seen along *c*, is given in Fig. 9.29. From the analysis, we find that intermolecular hydrogen bonds exist between O(2) and O(1)′ [2.47(1) Å], between O(3)″ and O_w [2.76(1) Å], and between O(4)′ and O_w [2.95(1) Å]; they are largely responsible for the cohesion between molecules in the solid state.[1]

Electron Counts on Atoms

We cannot determine from this analysis how the electronic charge on the individual species is distributed. With X-rays, a more precise analysis would be needed, followed by a peak integration of an electron density map. However, an ab initio calculation with Gaussian 94 on a free KHSQ species resulted in the following electron population parameters *p* on the individual atomic species.

	p		*p*		*p*		*p*
K	18.111	C_1	5.581	C_2	5.833	C_3	5.605
C_4	5.599	O_1	8.808	O_2	8.760	O_3	8.626
O_4	8.544	O_w	8.855	H_{12}	0.564	H_{3w}	0.560
H_{4w}	0.554						

The value of $\sum p$ is 86, which is also F(000)/4, as expected, the number of electrons in a single species. Although the above results would be modified slightly in the crystal environment, they are in excellent agreement with the formula weight and indicate the drawing of electron density from the less electronegative species, particularly the hydrogen atoms, towards the more electronegative

[1] Single and double primes indicate different neighboring molecules.

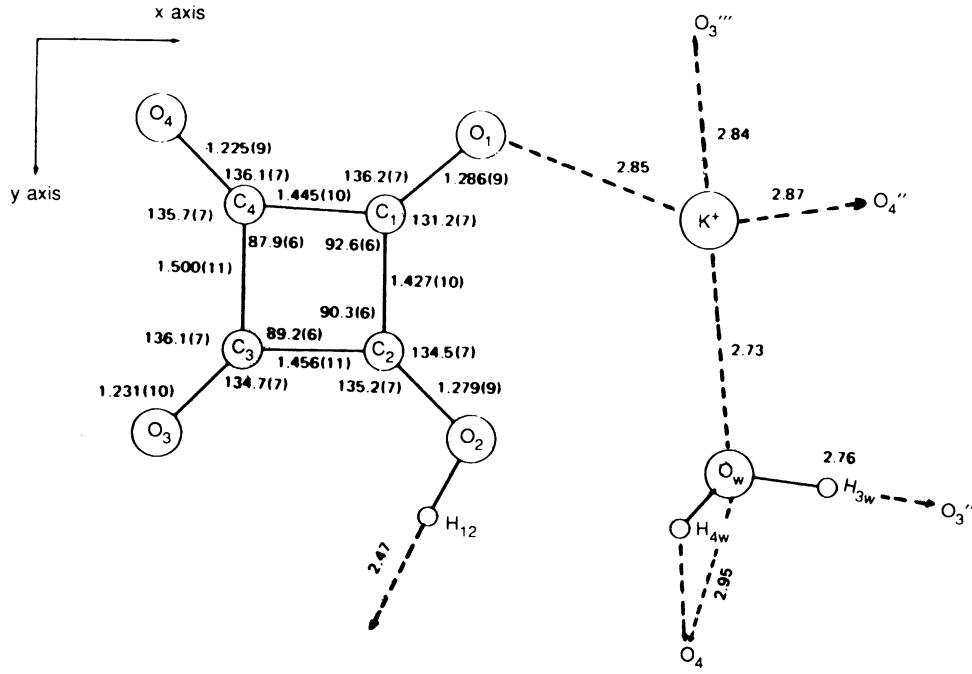


Fig. 9.28 Bond lengths and bond angles in the asymmetric unit of KHSQ; the OH. . .O distances refer to the overall O. . .O separations. Primes on atom symbols indicate neighboring asymmetric units; this diagram should be studied in conjunction with Fig. 9.29

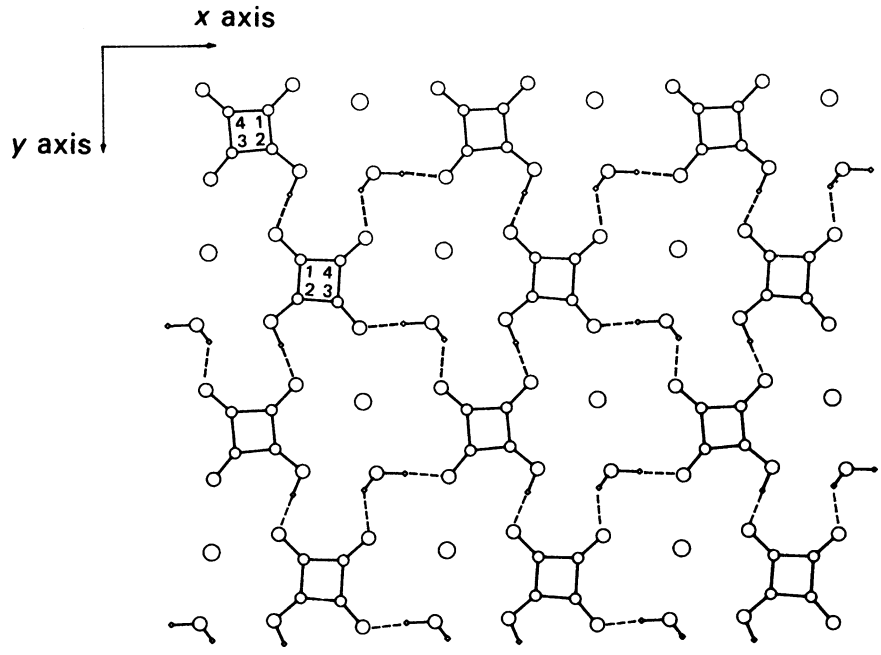
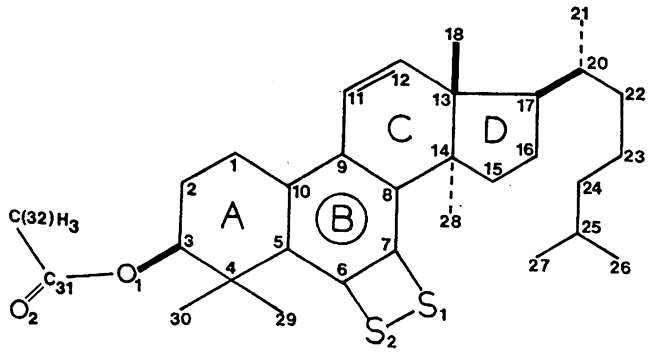


Fig. 9.29 Molecular packing diagram of one layer of the KHSQ structure as seen along *c*. The circles in order of decreasing size represent K, O, C, and H. The hydrogen-bond network is shown by dashed lines

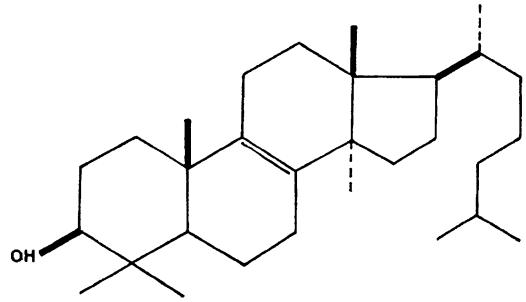
oxygen atoms. This effect leads to bond lengths involving hydrogen that measure shorter by X-ray diffraction than with neutrons, as we shall see in Chap. 11.

9.4 Crystal and Molecular Structure and Absolute Configuration of 3β-Acetoxy-6,7-Epidithio-19-Norlanosta-5,7,9,11-Tetraene [4]

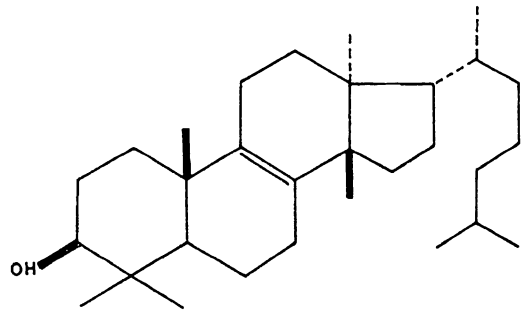
This interesting and stable dithiete compound (I) was obtained during an attempt to convert a derivative of lanostenol (II) into a derivative of euphenol (III), through the series of reactions (IV)–(VII). There are very few authenticated reports of 1,2-dithiones (VIII) or their tautomers, 1,2-dithietes (IX) [5]. It may be noted *en passant* that (III) is closely related to the compound in Fig. 1.7b, but a 3-β-hydroxyl group takes the place of the 3-β-iodoacetoxy group. At the time that this structure was reported, it was the first representative of its class. The stability of the dithiete arises most probably from a steric protection by the host molecule. From a crystallographic point of view, it has the interest of two heavy atoms, not related by symmetry, in the asymmetric unit, and of easily sufficient "weight" for a reliable determination of the absolute configuration of the molecule.



(I)

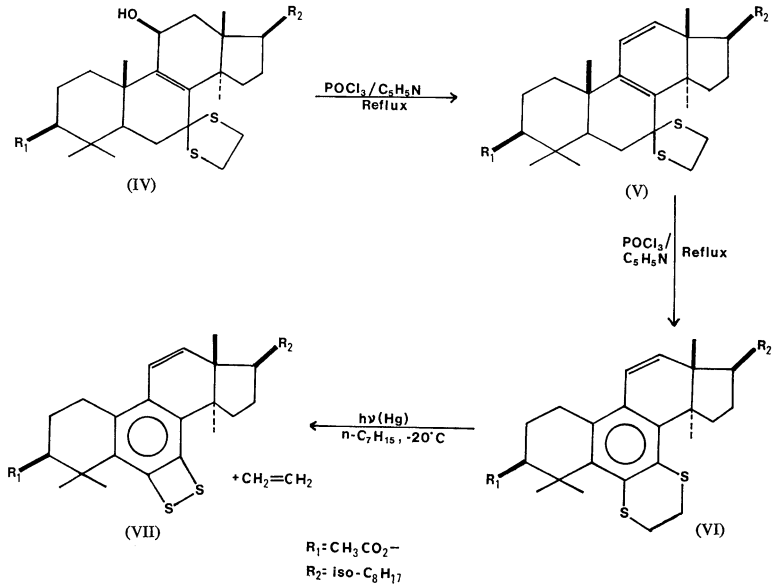


(II)



9.4.1 Preparation and Preliminary Optical and X-Ray Examinations

3-β-Acetoxy-7,7-ethlenedithio-5α-lanost-8-en-11-β-ol (IV) in pyridine heated under reflux with phosphorus oxychloride gave a colorless compound, the expected 8,11-diene (V). After a prolonged reaction time under reflux, colorless crystals corresponding to the formula $C_{33}H_{48}O_2S_2$, m.p. 161–163°C, $[a]_D = 45°$ separated. Examination by n.m.r and u.v. indicated a 2,3-dihydro-1,4-benzodithiine (VI). Photolysis of this compound in *n*-heptane at −20°C under medium-power mercury light produced a quantitative conversion to $C_{31}H_{44}O_2S_2$ (VII, I) and ethylene (C_2H_4) [6].



The compound (I) was recrystallized from a 1:1 mixture of dichloromethane and methanol as pale-yellow, lath-like parallelepipeds elongated in the direction of an axis, later to be described as *z*. The crystal forms present in decreasing order of development were {100}, {010}, and {001}. An optical examination indicated strongly that the crystals were monoclinic, with a β-angle close to 90°.

The density was measured by flotation in aqueous sodium bromide at 25°C. The result, 1.17 (1) g cm^{-3}, was obtained by measuring the refractive index of the flotation-equilibrium solution and then obtaining the density from a calibration chart.

X-ray photographs confirmed that the crystals were monoclinic and provided approximate unit-cell dimensions. The only systematic extinctions were $0k0$ for $k = 2n + 1$; hence, the space group was either $P2_1$ or $P2_1/m$. From the unit-cell and density measurements Z evaluated to 1.996, or 2 to the nearest integer. Thus, for crystals of the asymmetric molecule (I) under investigation, space group $P2_1/m$ was eliminated, and so we find two sulphur atoms per unit in space group $P2_1$.

9.4.2 X-Ray Measurement of the Unit-Cell Dimensions and Intensities

Accurate unit-cell dimensions were obtained from a least-squares fit to the θ values of 30 reflections with $\theta > 49°$, measured on a Siemens four-circle diffractometer. Resolution of the Cu $K\alpha_1$ peaks

Table 9.12 Crystal data for the title compound

Formula	$C_{31}H_{44}O_2S_2$
M_r	512.82
System/space group	Monoclinic/$P2_1$
Crystal dimensions/mm	0.2, 0.3, 0.5
a (Å)	20.1896(9)
b (Å)	11.0709(5)
c (Å)	6.4953(3)
β (°)	90.578(2)
V_c (Å^3)	1451.7(1)
D_m (g cm^{-3})	1.17(1)
D_c (g cm^{-3})	1.173(1)
Z	2
F(000)	556
μ(Cu $K\alpha$) (cm^{-1})	19
$\sum Z^2$(S)/$\sum Z^2$ (light atoms)	0.4

Table 9.13 Statistics of |E| values

	This structure	Acentric	Centric
$\overline{\lvert E^2 \rvert}$	1.00	1.00	1.00
$\overline{\lvert E \rvert}$	0.88	0.89	0.80
$\overline{\lvert E \rvert^2 - 1}$	0.76	0.74	0.97
% ≥ 1.5	11.2	10.5	13.4
% ≥ 1.75	4.7	4.7	8.0
% ≥ 2.0	2.4	1.8	4.6
% ≥ 2.5	0.4	0.2	1.2

($\lambda = 1.54056$ Å) and those for Cu $K\alpha_2$ ($\lambda = 1.54439$ Å) was obtained for these reflections, which were chosen so as to be evenly distributed in reciprocal space. The crystal data are summarized in Table 9.12. In the usual order of $a < b < c$, a and c have been reversed in order that the higher convention of $\beta > 90°$ was preserved.

The intensities of 2605 of a theoretically possible 2629 symmetry-independent reflections with $\sin\theta/\lambda \leq 0.60$ were measured on the diffractometer by a five-measurement technique [7]. Of the total measured reflections, 243 for which $I < 2.58\sigma(I)$, were very weak; the average standard deviations of the F_o data were 1.8% (2362 data) and 2.7% (all reflections). The very weak reflections were omitted from the structure analysis. A reference reflection was measured after every 20 reflections: the data were scaled, first within batches of 20 reflections and then to the average value of the reference reflection. The scaling factors varied over a range of less than ±2%, and there was no evidence of crystal decomposition.

Corrections were applied for Lorentz and polarization effects, but not for absorption ($\overline{\mu R} = 0.78$). |E| values were calculated by the K-curve method; their statistics, listed in Table 9.13, show very good agreement with those for the ideal acentric distribution.

9.4.3 Structure Determination and Refinement

Patterson and sharpened ($|E|^2 - 1$) syntheses revealed the expected heavy-atom peaks on the $(u, \frac{1}{2}, w)$ Harker section, and in addition four peaks of about double the weight of the Harker peaks, with a geometry in accord with that of two pairs of heavy (S) atoms, each pair being related by 2_1 symmetry. The y coordinate of S(1) was fixed at $\frac{3}{4}$, and that for S(2) was deduced to be 0.69. Their difference in y of 0.66 Å was sufficient to break the pseudo-symmetry which occurs in $P2_1$ where the asymmetric unit contains two heavy atoms. A similar situation has been reported, for example, with epilimonol iodoacetate [8].

Successive Fourier syntheses located all non-hydrogen atoms in the molecule, and the structure was refined on |F| by full-matrix anisotropic least squares. A Fourier difference synthesis enabled most of the hydrogen atoms to be located and they were refined isotropically. The six hydrogen atoms on the terminal methyl groups C(26) and C(27) were relatively poorly defined, and their coordinates were determined geometrically in terms of their minimum energy conformations. It is highly probable that some disorder exists in this region of the structure; see also Fig. 1.7 and accompanying text.

Corrections for anomalous dispersion were applied ($\Delta f_S'' = 0.60$) and the refinement converged at $R = 3.4\%$ ($R_w = 4.7\%$). All computations were performed by both local programs and the XRAY-72 system [9]. A list of structure factors has been deposited with the British Library Lending Division as Supplementary Publication No. SUP 31423 (23 pp. 1 microfiche).

9.4.4 Absolute Configuration

The absolute configuration was examined by carrying out the least-squares refinement again, but with the signs of the imaginary components $\Delta f''$ of the structure factors reversed. Convergence was attained at $R = 3.6\%$ ($R_w = 4.9\%$). The ratio of the weighted R values is 1.04, and by Hamilton's ratio test [10], the result is significant at a level better than 0.01. Thus, the absolute configuration corresponds, not surprisingly, to the acetoxy group in the β-configuration, with the configurations at C(13), C(14), and C(17) unchanged from those of the starting material (IV). Although the configuration was quite clear from this result, the Flack parameter, Sect. 7.6.1, would now constitute the preferred test.

9.5 Discussion of the Structure

Figure 9.30 is a stereoscopic illustration of the molecule (I) as seen along $[0\bar{2}1]$. The central ring system shows a high degree of planarity: the average deviation of the atoms forming the plane C(5)–C(10) is 0.01 Å, and that for the atoms S(1), S(2), C(1), C(4), C(11), and C(14) from the same plane is 0.07 Å.

Fig. 9.30 Stereoscopic illustration of the title compound (I), as seen along $[0\bar{2}1]$

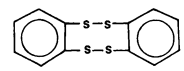
The slightly different environments of the two sulphur atoms are shown by their deviations from the plane (0.00 and 0.05 Å) and by the differing S–C–C angles (135.7(3)° and 100.5(3)°).

The question of the existence of the two sulphur atoms as a dithione (VIII) or a dithiete (IX) has been clearly resolved by the geometry of the epidithio ring system, as shown by the data below:

Bond lengths (Å)	Bond angles (°)
S(1)–S(2) = 2.119(1)	C(5)–C(6)–C(7) = 123.8(4)
S(1)–C(7) = 1.767(4)	C(6)–C(7)–C(8) = 122.1(3)
S(2)–C(6) = 1.782(4)	C(5)–C(6)–S(2) = 135.7(3)
C(6)–C(7) = 1.387(5)	C(6)–C(7)–S(1) = 103.3(3)
	C(6)–S(2)–S(1) = 80.3(3)
	C(7)–C(6)–S(2) = 100.5(3)
	C(8)–C(7)–S(1) = 134.5(3)
	C(7)–S(1)–S(2) = 79.3(3)

Additionally, the Raman spectrum showed a band at 486 cm^{-1}, which corresponds to a –S–S– stretch [11]; the >C=S frequency is in the range 1171–1180 cm^{-1}.

(VIII) (IX)



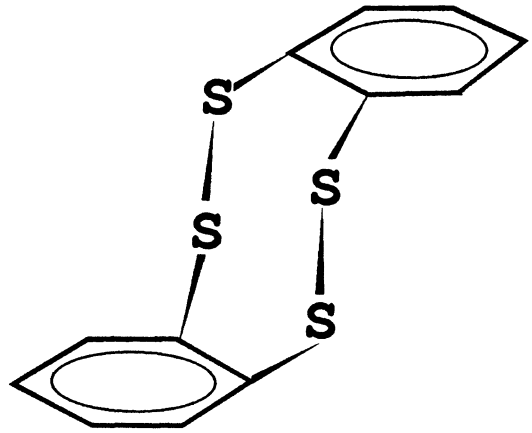
(X)

A theoretical study was carried out in connection with this unusual dithiete compound (I) [12, 13]. CNDO/2 calculations were applied to a model consisting of ring *B* + S(1) and S(2), a hypothetical C_6S_2 molecule, for a range of C(7)–S(1), C(6)–S(2), C(7)–S(1)–S(2), and C(6)–S(2)–S(1) parameters, with the following encouraging results for a minimum energy conformation:

	Calculated	Experimental
C(7)–S(1)	1.74 Å	1.767(4) Å
C(6)–S(2)	1.76 Å	1.782(4) Å
C(7)–S(1)–S(2)	100°	103.3(3)°
C(6)–S(2)–S(1)	99°	100.5(3)°

A compound $C_6H_4S_2$, corresponding to the [C_6S_2] fragment, cannot be isolated, as it dimerizes through S–S bonding to the compound dibenzo[*c*,*g*]tetrathiocin, $C_{12}H_8S_4$, which could exist in a rigid *trans* form (XI), or in a mobile form with a range of possible conformations: CNDO/2 calculations as a function of the dihedral angle between the two rings of the dimer showed the *trans* form to have the most stable (minimum energy) conformation.

Fig. 9.31 Stereoview of the packing of the title compound (I), as seen along $[02\bar{1}]$



The cohesion of the S–S bond is most probably occasioned by *pd*-orbital overlap. This mechanism is not possible for the smaller oxygen species and *o*-quinones are well established as >C=O species. Figure 9.31 illustrates the molecular packing of the structure. The intermolecular S. . .S distances are approximately 10 Å, and there is no tendency to dimerization. Steric factors probably prevent a close approach of the two molecules.

In the reactions of (IV) to (VII, I), stage (IV) to (V) gave the expected elimination of water and formation of the 11,12 double bond. Stage (V) to (VI) was accompanied by the loss of a $-CH_3$ group and three other hydrogen atoms. The final stage (VI) to (VII, I) was quantitative, with loss of ethylene. The probable mechanism is a disruption at the spiro-C(7) carbon atom, followed by migration of a sulphur atom to C(6) and subsequent aromatization of ring *B*. The aromatization provides a conformational driving force for the elimination of the methyl group at C(19), the methyl group that is attached to C(10) in II. Under u.v. activation, the structure is further stabilized by the loss of ethylene from the dithian ring.

9.6 Some Remarks on X-Ray Structure Determination

No description of the process of X-ray crystal structure analysis can be as complete or as satisfying as a practical involvement with the subject. In teaching crystallography and structural chemistry, projects that include crystal structure determinations have become increasingly important. However, in order to attempt to replicate the practical side of structure analysis, insofar as is possible in isolation from the laboratory, problems on this topic additional to those at the end of this chapter are given in Chap. 13; they involve the program system XRAY and the several sets of data that accompany it.

They have been designed to give practice, albeit in two dimensions, with the fundamental techniques of solving crystal structures, given the crystal and reflection data. The programs can be executed on any IBM-type PC. We encourage the reader strongly to tackle these problems and so

engage in a practical way with the techniques of crystal structure analysis. We refer the reader also to Appendix D, wherein are listed the many program systems now freely available to the structural crystallographer.

9.7 Biomolecular Modeling: Bioinformatics

We have seen that the technique of X-ray crystallography is a powerful means for the determination of accurate crystal structures, and that this in turn provides a source of information on molecular geometry in the form of bond lengths, bond angles, conformational parameters, planarity, and intermolecular interactions. There are many reasons that prompt such studies. For example, we may wish to design new industrial materials, a process which requires access to such experimental results. We may wish to design a new drug molecule with modified properties, such as increased or decreased potency, improved specificity, or lacking unwanted side effects. Again, detailed knowledge of the molecular properties of the starting material, including its molecular geometry and flexibility as well as those of the biological target are a necessary prerequisite for such research. One approach where the crystallographic data can be applied effectively for such studies involves the use of intensive computational techniques in which structural data from one or more starting materials are employed. Obviously, there are many parameters and variables involved in this type of work and we have selected an example which demonstrates many of the techniques employed.

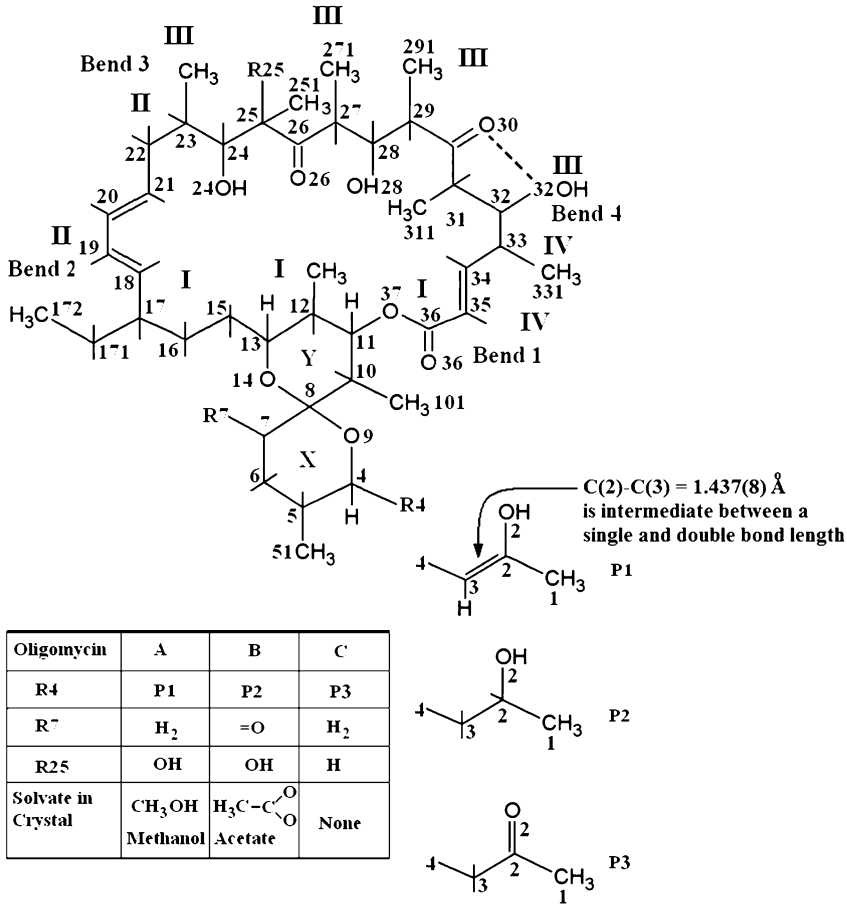
We introduce in this section, as an important by-product of X-ray structure determination, the technique known as *biomolecular modeling*. The method described here depends on the availability of a known macromolecular structure, usually a protein or a nucleic acid, such as DNA or RNA, and another usually much smaller molecule or ligand that is known to bind either chemically or more usually to interact non-covalently with the macromolecule. It is usually an advantage, when embarking on studies of this kind, for both structures to be known from previous X-ray or neutron diffraction analysis. Frequently, it is possible to determine the structure of the macromolecule-ligand assembly as such through a further diffraction study, as discussed later in Chaps. 10 and 11. However, when this is not possible or practicable, biomolecular modeling may provide plausible answers about the mode and location of the interactions between the two molecules.

9.8 Docking Oligomycin into ATP Synthase: Ligand and Receptor

The macrocyclic naturally occurring antibiotic oligomycin (Fig. 9.32) binds to and thereby inhibits the biological activity of the multi-subunit protein enzyme ATP synthase (ATPase). The ATP synthase complex is constructed of numerous protein subunits that work together (Fig. 9.33). It is of interest to locate the site of interaction between the ligand oligomycin and the receptor ATPase. The detailed X-ray structures of the three oligomycin structures [14] A, B, and C described below are available for modeling studies designed to locate and characterize this site of interaction.

9.8.1 Why Modeling Studies?

Ideally the best way to investigate ligand/receptor binding would be to carry out a new X-ray or neutron diffraction study on the complex of the two components. However, the protein (ATPase) is so large and complex that it has not been possible to determine the complete structure as such. Instead, a possible structure has been assembled from several independent studies, Sect. 9.10. Thus, it has been



Oligomycin	A	B	C
R4	P1	P2	P3
R7	H_2	=O	H_2
R25	OH	OH	H
Solvate in Crystal	CH_3OH Methanol	$H_3C-C(=O)-O$ Acetate	None

Oligomycins A, B and C: Crystallographic numbering, R groups and their locations, and solvation characteristics.

The approximate locations of the chain stretches I to IV and the bends 1 to 4 are indicated.

Fig. 9.32 Chemical structures of oligomycins A, B, and C. The crystallographic numbering scheme is shown and major features of the structures as discussed in the text are labeled

necessary to model the binding of oligomycin to ATPase, at least until a better way is devised for studying the structure of the whole complex by diffraction methods.

9.9 X-Ray Structures and Absolute Configurations of the Antibiotics Oligomycins A, B, and C: Inhibitors of ATP Synthase

Oligomycin is a naturally occurring antibiotic that occurs in the gram-positive bacterium *Streptomyces diastatochromogenes*. Knowledge of the site on the FO (O = oligomycin) domain (Fig. 9.33) where oligomycin binds to ATPase may enable the development of new improved ATPase inhibitors that could be produced synthetically. As the structure of oligomycin (Fig. 9.32) contains a 26-membered ring, and therefore has numerous conformational possibilities, a detailed X-ray structure

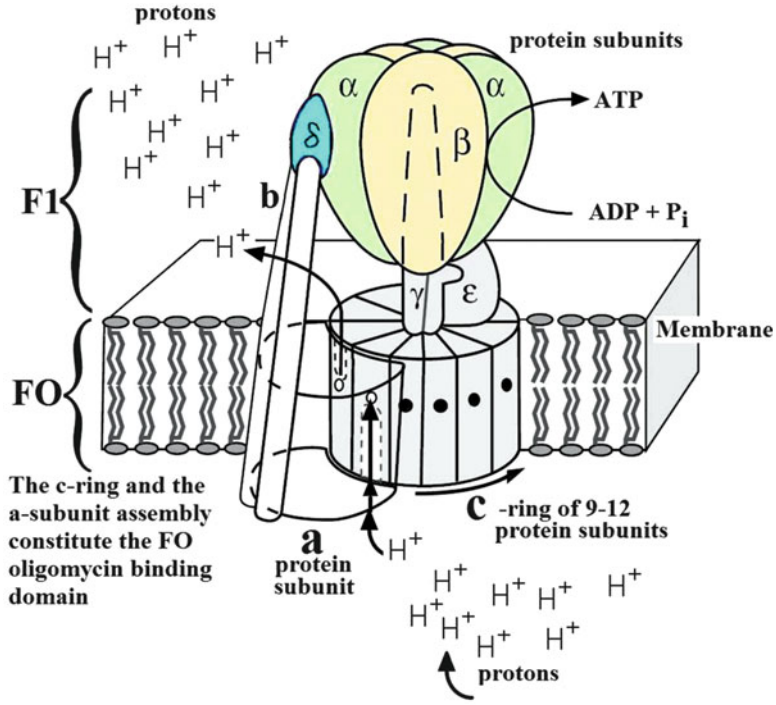


Fig. 9.33 Subunit arrangement for *Escherichia coli* ATP synthase

is an essential prerequisite for the modeling studies that are aimed at locating and characterizing the FO–ATPase binding site. In fact, recently published X-ray structures [14] of the three common forms, of oligomycin, A, B, and C, including their absolute configurations, are available for these modeling studies. The X-ray studies reveal regions with differences in the three-dimensional structures and hydrogen-bonding propensity between the oligomycins, which may be associated with their potential to bind to ATPase. The experimentally determined X-ray structures of the three forms of oligomycin were used as the starting point for docking studies.

9.9.1 Summary

All three oligomycins crystallized in space group $P2_12_12_1$ with four molecules per unit cell. Oligomycin A crystallized as the methanol solvate $C_{45}H_{72}O_{11}{\cdot}CH_3OH$ with unit-cell parameters $a = 10.476(3)$ Å, $b = 17.342(1)$ Å, $c = 26.825(5)$ Å; oligomycin B as an acetic acid solvate $C_{45}H_{71}O_{12}{\cdot}CH_3CO_2$ with unit-cell parameters $a = 10.351(3)$ Å, $b = 17.305(1)$ Å, $c = 26.929(5)$ Å; and oligomycin C, $C_{45}H_{74}O_{10}$, with unit-cell parameters $a = 10.385(2)$ Å, $b = 11.9510(9)$ Å, $c = 38.007(4)$ Å. Oligomycin A refined with final R indices $[I > 2\sigma(I)]$, $R_1 = 0.0734$, $wR_2 = 0.1940$; R (all data): $R_1 = 0.1106$, $wR_2 = 0.2100$, and absolute structure parameter $= -0.7(4)$; for oligomycin B, final R indices $[I > 2\sigma(I)]$ are $R_1 = 0.0479$, $wR_2 = 0.1388$; R (all data): $R_1 = 0.0581$, $wR_2 = 0.1435$, and absolute structure parameter $= -0.2(2)$; and for oligomycin C, final R indices $[I > 2\sigma(I)]$ are $R_1 = 0.0454$, $wR_2 = 0.1130$; R (all data): $R_1 = 0.1061$, $wR_2 = 0.1221$, and absolute structure parameter $= 0.1(3)$.

The present study has provided:

1. Corrections to the previously published chemical structures of the oligomycins
2. Full descriptions of the absolute configurations
3. Information on regions of the structures with minor but important differences in their three-dimensional structures that may create differences between the oligomycins in their potential to bind to sites on the ATP synthase molecule. These results are all of major importance for future studies designed to establish details of the actual binding of oligomycins to ATP synthase.

9.9.2 Background

Oligomycin is an antibiotic that occurs in the Gram-positive bacterium *S. diastatochromogenes*. It is known [15] to act as a potent inhibitor of oxidative phosphorylation, and as a specific inhibitor acting at the site of coupling between respiration and phosphorylation. Oligomycin is used extensively in experiments designed, for example, to investigate the mechanisms of processes in mitochondria [16] and studies involving oligomycin-sensitive ATP synthase [17, 18].

The three distinct forms or types studied here, known as Oligomycin A, B, and C, differ from each other chemically only in one or two substituents, as discussed below. Experiments using oligomycin as a probe do not usually attempt to distinguish between possible differences between the effects of using a particular type of oligomycin, and often use the unrefined mixture of all three. Enzymes known as ATP synthases are complex multi-subunited protein structures that can synthesize adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate. In mitochondria, the ATP synthase molecule can be visualized as having a major domain, F1, outside the cell membrane, and a minor domain, FO, embedded within the membrane. FO derives its name from being the oligomycin binding domain, and is also known as OSCP (the oligomycin sensitivity conferral protein). The F1 domain, diameter 9 nm, can be seen in the transmission electron microscope by negative staining, first observed, at least partially, by Fernandez-Moran [19]. The antibiotic oligomycin binds to and inhibits ATP synthase. The crystal structure of the F1 catalytic-domain of ATP synthase is known [20] and is largely consistent with Boyer's rotary-catalysis model [21].

9.9.3 Experimental

Samples of Oligomycin A, B, and C were purchased from Reanal Finechemical company. Small seed crystals were produced by controlled evaporation of solutions at room temperature over several weeks in (a) 50/50 acetone and acetic acid (for A); (b) 50/50 acetone/ethanol (for B); and acetone only (for C). Recrystallization at 4°C was induced by seeding and produced X-ray quality crystals after 20 days.

The forms A and B are almost colorless and prismatic or platy crystals of typical dimensions 0.25, 0.15, 0.15 mm. Form C crystals are also colorless but more plate-like with typical dimension 0.3, 0.25, 0.10 mm. In all cases, crystal quality was tested by means of Mar IP images taken in-house using copper X-radiation prior to data collection.

Crystals were mounted for room temperature data collection on glass fibers, using epoxy resin (araldite), and having dimensions 0.20, 0.15, 0.15 mm^3 (A), 0.20, 0.20, 0.15 mm (B), and 0.3, 0.25, 0.10 mm (C) and then mounted for X-ray intensity measurement on an Enraf-Nonius CAD-4 automated 4-circle diffractometer equipped with a graphite monochromator for room temperature

recording of the diffraction pattern. Intensity data were recorded using monochromatic Cu $K\alpha$ radiation; CAD-4 Express Software [22] was used for cell determination and refinement and data reduction. In all cases, accurate cell parameters were determined from 25 reflections ($25 < \theta < 28°$). For data collection, $\omega - 2\theta$ scans were used under computer control. The intensities 4680 (A), 5266 (B), and 5498 (C) of reflections were measured for $\theta < 70°$. The crystals showed no significant variations in the intensities of the three standard reflections during the course of data collection. Lorenz and polarization corrections were applied and a semi-empirical correction for absorption [23] was made for each set of data. All three oligomycins were found to have space group $P2_12_12_1$, the A and B unit cells being very similar but bearing no relationship to that of C.

9.9.4 Structure Determination and Refinement

The oligomycin structures were solved using SHELXS-86 [24] and refined with SHELXL-97 [12]. Non-hydrogen atoms were refined anisotropically by full-matrix least-squares techniques. Apart from the exceptions discussed below, hydrogen-atom positions were calculated geometrically and refined in riding mode with isotropic displacement parameters fixed by the program and common for all hydrogen atoms attached to the same carbon atom. In the riding mode, hydrogen atoms are first positioned according to standard molecular geometry. Then, in subsequent refinements, these atoms are adjusted in such a way as to preserve the chosen geometry, and the isotropic temperature factors U_{iso} of these atoms can be refined in the usual manner. A very high value of U_{iso} may indicate a misplaced hydrogen atom, and the difference-Fourier map should be carefully examined. Calculations were made with SHELXL-97 [25] as implemented in WinGX [26] and geometrical calculations were made with the programs PARST and PLATON [27] as implemented in WinGX.

9.9.5 Results

Figure 9.32 summarizes the chemical constitutions of the three forms as determined in these structure analyses and the atom numbering used: ORTEP [28]/Raster3D [29] generated corresponding views of the three molecular conformations which are shown in Fig. 9.34a–c. Figure 9.35a–c are surface representations prepared with the program VMD [30].

9.9.6 Discussion

General

Apart from the exceptions discussed below, corresponding bond lengths and bond angles between the three forms of oligomycin are in good agreement with one another and the values are within the expected ranges for this type of molecule (see Tables 8.21 and 8.22).

Side Chain R4

According to the literature available [31, 32] when the X-ray analyses presented here were initiated, the R4 side chains (Fig. 9.32) in all three forms were designated as being propanone (P3, Fig. 9.32). As the analyses progressed it soon became clear that this was not the correct assignment for any of the forms.

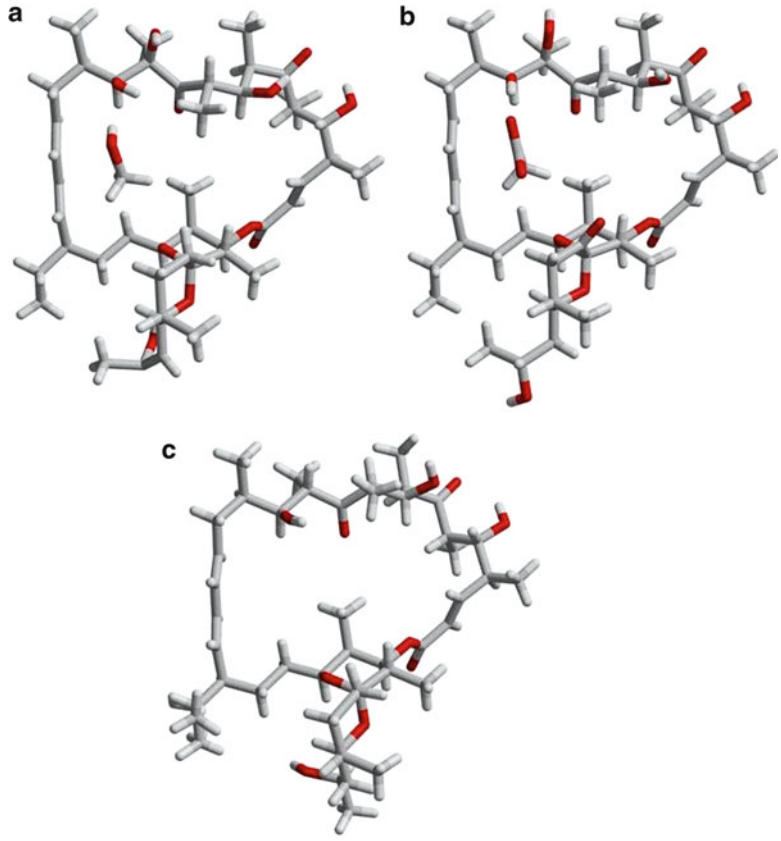


Fig. 9.34 Molecular conformations of Oligomycins A, B, and C respectively. The view selected displays the most open projection of the molecule (ORTEP/Raster)

Oligomycin A

In the refined Oligomycin A structure, the geometry for side chain R4 is as follows: C(1)–C(2) = 1.536(10) Å, O(2)–C(2) = 1.487(9) Å, C(2)–C(3) = 1.437(8) Å, C(3)–C(4) = 1.516 (7) Å; C(1)–C(2)–C(3) = 111.0(6)°, O(2)–C(2)–C(3) = 104.0(5)°, C(2)–C(3)–C(4) = 123. 1(5)°. The bond angle C(2)–C(3)–C(4) is aromatic rather than tetrahedral, as is the associated short C (2)–C(3) bond length. In difference electron density maps prior to assignment of hydrogen-atom positions, there was only one significant peak close to where hydrogen atoms were expected to be attached to C(3). Several different models involving the geometrical assignment of hydrogen atoms on C(3) were set up. Only one hydrogen atom, H(3A), was left with an acceptably small temperature factor, U_{iso} = 0.2054 Å^2, after refinement of a starting model in which two tetrahedrally positioned hydrogen atoms on C(3) were both refined with respect to their positions and isotropic thermal parameters. In the very final model, for consistency with the overall model, H(3) was geometrically fixed using the aromatic option and tied to C(3), H(3) being close to the former refined H(3A) position. These results are consistent with the R4 side chain in Oligomycin A as being propenol (R4 = P1, Fig. 9.32), not propanone (P3) as was previously thought [31]. In view of the unusual geometry revealed in this analysis for side chain R4, several complete intensity data sets were measured for Oligomycin A using fresh crystals; the results were always consistent with those reported above.

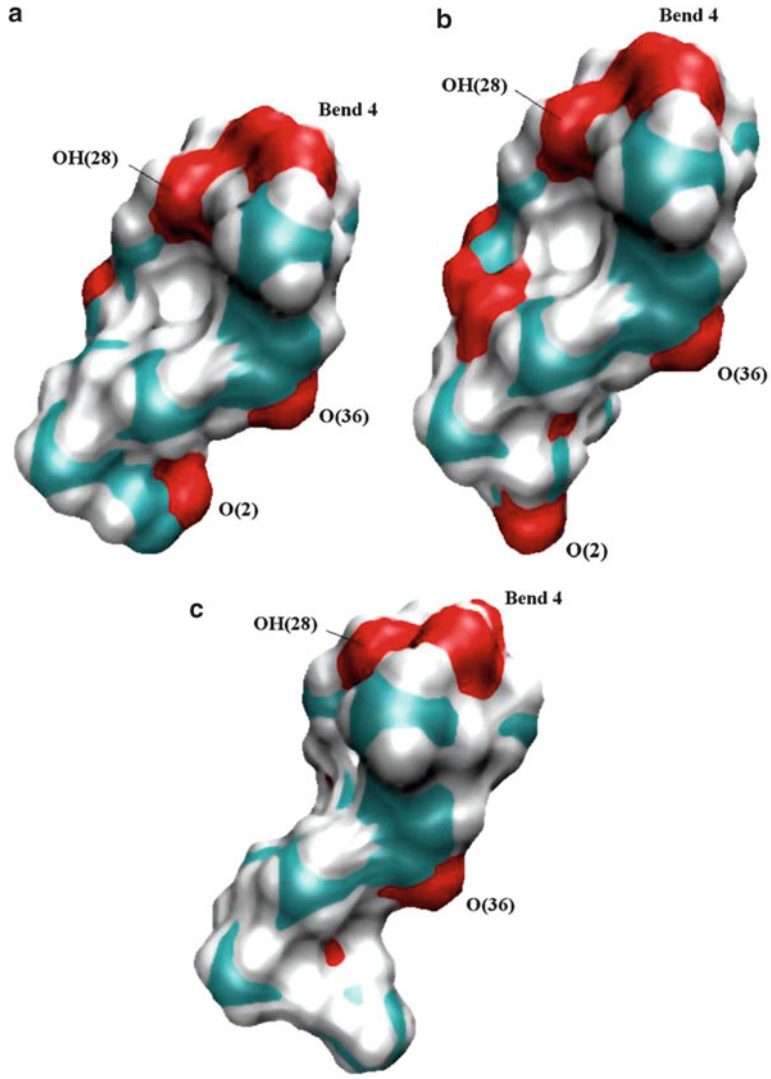


Fig. 9.35 Surface representations of Oligomycins A, B, and C (VMD SURFACE). Red = oxygen, blue = carbon, white = hydrogen. (**a**) Oligomycin A: OH(25) forms an intramolecular hydrogen bond with O(24) which could exclude both from receptor binding. (**b**) Oligomycin B: there are five active groups in this surface view of oligomycin B, compared to two in oligomycin A and zero in oligomycin C. This suggests that oligomycin B has a greater potential for receptor binding than A or C. (**c**) Oligomycin C: this shows three potential receptor binding sites. OH (24) appears to be partially buried but does in fact participate in an intermolecular hydrogen bond. In addition OH(2) is quite exposed in the structure but out of this view

Oligomycins B and C

The R4 side chains for both Oligomycins B and C were also quite clearly not propanone (P3) but were consistent with P2 (Fig. 9.32), that is, with propanol. Both side chains refined with perfectly normal geometry and well-behaved thermal displacement parameters including H(3A) and H(3B) in both structures.

Oligomycins A and B have very similar unit cells and belong to the same space group. Many of the atoms are in similar positions in the two structures. Chemical differences can be noted from the formulae (Fig. 9.32). Those noted in the R4 side chains in the above discussion were quite unexpected and may be important for future activity studies. As discussed below, the R4 side chain in Oligomycin A differs in conformation from the corresponding side chains in Oligomycins B and C and forms an internal hydrogen bond between O(2) and O(7) in pyranose ring X (Figs. 9.32 and 9.34a). The correct assignment of the chemical features of any biologically active molecule is of course an essential prerequisite to an understanding of its function and mode of action.

Pyranose Rings X and Y

Both of the *spiro* linked pyranose rings X and Y (Figs. 9.32 and 9.34) are distorted chairs in all three forms of oligomycin, with corresponding rings having the same type of chair conformation in all three structures. The ring torsion angles range from about ±45 to ±65° in both rings, and corresponding torsion angles differ by as much as 9° but usually agree within 4°. The eight *spiro* linkage torsion angles are either about ±170 or ±70°, with similar variations between Oligomycins as exhibited by the rings.

9.9.7 Conformational Variations in the Macrocyclic Structures

Figure 9.32 shows the common macrocyclic structure of the oligomycins. The 26 atom chain runs from C(11) to the ether oxygen O(37) which is the only non-carbon atom in the chain. Close inspection of Fig. 9.34a–c reveals small differences in the overall macrocycle ring conformations. Structures A and B are generally more similar in conformation to each other than to structure C which has differences in torsion angle of up to 18° as discussed below.

Chain 1: C(36) to C(17) is Predominantly Fully Extended

The chain stretching from O(37) to C(17) and including side chain atom C(171) is fully extended in all three structures (torsion angles ca. ±180°). The ether oxygen O(37) is associated with a pronounced kink in the chain (torsion angle about bond O(37)–C(11) is ca. −80°) in all structures.

Chain 2: C(17) to C(23)

This chain is also in an extended conformation but is slightly convex due to opposite sense twists at either end.

Chain 3: C(21) to C(34)

This long stretch forms an irregular but gradual clockwise coil and is the most conformationally variable part of the structure. Correspondence between the chain torsion angles in the three oligomycins in this region is generally the most variable with differences as great as 17° about C(24)–C(25) between Oligomycin A and C, and differences frequently around 10° elsewhere. Conformational variability in this region can be seen from close examination of Fig. 9.33a–c.

Chain 4: C(33) to C(36)

This short stretch of chain is fully extended in all three structures.

Bends in the Macrocyclic Chains

1. The bend between chains 1 and 4:
   This bend is created by a *cis* twist about bond C(35)–C(36) of about −3°.
2. The bend between chains 1 and 2:
   This bend is associated with changes in the torsion angles about the bonds C(16)–C(17) and C(17)–C(18) from the extended mode on either side (ca. 180°) to values of approximately −60 and 100° in all three structures.
3. The bend between chains 2 and 3:
   This bend similarly involves transitions in two torsion angles, namely about C(21)–C(22) and C(22)–C(23) from the extended mode, their values being approximately −130 and −60° respectively.

4. The bend between chains 3 and 4:
   This bend also involves two torsion angles, namely, C(32)–C(33) and C(33)–C(34) changing from extended mode on either side. C(32)–C(33) has values around 55° in all three structures whereas C(33)–C(34) is highly variable with values of 134.3, 135.7, and 117.7° in Oligomycins A, B, and C respectively.

Side Group Orientations

Close inspection of Fig. 9.33a–c reveals small differences in some side group orientations resulting from variations in the ring conformation discussed above. These differences may be important factors in determining the binding characteristics essential for understanding the biological activities of the oligomycins. In the following discussion the term "external" applies to groups facing in a general direction outwards, that is, away from the macrocyclic ring and "internal" facing groups are pointing.

CH_3 Groups

In Oligomycin A, B, and C, two CH_3 groups are internal, that is, those on C(12) and C(27) (Fig. 9.32). All other CH_3 groups are external.

OH Groups

In Oligomycin A, the five OH groups are all external and four of them participate in hydrogen bonds, namely, O(2)H...O(7) (in ring X, see Fig. 9.32), stabilizing the R4 side chain; O(25)H...O(24)H, stabilizing bend 3 (see above); and O(32)H...O(30), both stabilizing bend 4 (see above). Group O(28)H forms an intermolecular hydrogen bond with O(36) (see below).

In Oligomycins B and C, the situation is quite different, possibly due in part to the lack of the R25 OH group (Fig. 9.32). The four OH groups are all external but with O(32)H...O(30), stabilizing bend 4, being the only intramolecular hydrogen bond formed. This provides Oligomycins B and C with more opportunities for hydrogen-bonding to protein receptors, and may explain differences which could exist in their binding characteristics compared to Oligomycin A. Owing to the presence of three more intramolecular hydrogen bonds, Oligomycin A may be deemed to be a much more stable conformation than Oligomycin B or C.

=O Groups

In Oligomycin A, B, and C, =O(26) is internal, and =O(30) and =O(36) are external. Oligomycin B has an additional =O(7) on pyranose ring X and this side group may be described as external and does in fact participate in intermolecular hydrogen bonds.

Summary of the Hydrogen Bonds in Oligomycins A, B, and C

Only one intermolecular hydrogen bond, between OH(28) and O(36), is a consistent feature of these structures. It may be concluded therefore that both OH(28) and =O(36) are highly exposed on the molecular surface: OH(2) is also exposed in Oligomycins B and C forming hydrogen bonds either to solvent, in B, or to symmetry-related molecules in C. In Oligomycin A, OH(2) forms an intramolecular hydrogen bond to O(9) on the pyranose ring X: OH(24) is also exposed forming hydrogen bonds to solvate molecules in A and B and to symmetry-related OH(2) groups in structure C. The OH(32) group forms consistently a stabilizing intramolecular hydrogen bond with =O(30), and both oxygen atoms can be discounted from binding studies. Apart from the intramolecular hydrogen bond between O(2) and O(9) in Oligomycin A, the pyranose oxygen atoms O(9) and O(14) do not participate in hydrogen-bonding interactions and can be taken to be inaccessible, as can =O(26) and –O(37)–. It may therefore be concluded from these studies that the key groups which are likely to be involved in intermolecular interactions with protein subunits are: OH(2) (probably only in Oligomycins B and C), OH(24), OH(28), and O(36). The conclusions about OH(28) and =O(36)

are clearly borne out by the VMD [22] SURFACE plots of the three molecules shown in Fig. 9.35. Other features and differences of the molecular structures discussed above can be clearly seen by careful inspection of these plots.

Absolute Configuration of Oligomycins A, B, and C
Oligomycins A, B, and C all include a number of asymmetric centers which, for the purposes of modeling studies, need to be defined in the correct absolute configuration. The X-ray studies reported in the present article have in all three cases been of sufficiently high quality to enable the correct assignment of configuration in all three cases. The refined Flack [33] parameter, Sect. 7.6.1, and its standard deviations for the three structures are: $-0.7(4)$, $-0.2(2)$, and $0.1(3)$, respectively. These three values are all equal to 0.0 within 3 standard deviations or better and it may be concluded that each structure has been assigned the correct absolute configuration, which can be seen by inspection to be the same in each case. All figures depicting the oligomycin structures show them in their correct absolute configurations.

Conclusions

While Oligomycins A, B, and C are chemically very similar, and A and B are crystallographically pseudo-isomorphous, the small differences that do exist between them have significant effects on how they take part in intermolecular interactions. This information is important for the assessment of the receptor binding potential of oligomycin to sites in the ATP synthase system and, in particular, when such assessments are being carried out by graphics-aided computer simulations.

It has been shown that in the case of Oligomycin A two surface-located sites are available for intermolecular binding, while there are four in Oligomycin B and three in Oligomycin C. They are highlighted in Table 9.1, counting interactions with solvent as possible intermolecular binding sites. This result indicates that Oligomycin B, with an extra oxygen atom, O(7), available on the surface, may be better favored than the A or C form for taking part in interactions with protein subunits.

9.10 Structure of ATP Synthase (ATPase): The Receptor

The three-dimensional structure of ATPase shown schematically in Fig. 9.33 has been constructed from the following components:

1. The X-ray structure of the catalytic-domain F1 [20]
2. Collected information about domain FO from biochemical studies

The crystal structure of F1 is largely consistent with Boyer's [22] rotary-catalysis model. The absence of an X-ray structure for domain FO provides the motive for carrying out modeling studies of the oligomycin binding. ATPase is a large subunited protein complex found in the mitochondria of cells, and its function is to extract energy via the conversion of ADP into adenosine triphosphate (ATP). The general reaction scheme is written as:

$$\text{ADP} + \text{Phosphate} \rightleftharpoons \text{ATP}$$

The mechanism by which this conversion takes place is unique as it employs a biological rotor driven by the passage of protons through a membrane. Various studies have shown that the ATPase complex is constructed of numerous protein subunits that work together. Figure 9.33 shows schematically the various protein subunits (a, b, c, α, β, γ and δ) and how they are arranged to form the

ATPase complex. The ATPase complex straddles a cell membrane, which is itself a complex protein assembly. Further details of the ATPase structure are given below.

The F1 Domain

The α and β subunits of ATPase (Fig. 9.33) contain the catalytic sites that carry out the reaction between ADP and ATP; they are arranged cyclically in an $\alpha\beta\alpha\beta\alpha\beta$ assembly. In the center of this construct is the "stalk" which consists of the γ subunit with, in the case of mitochondrial ATP synthase, δ and ε subunits at the "foot," the arrangement being slightly different in the case of *Escherichia coli* ATP synthase. This foot is in contact with the *c*-ring of protein subunits which is made up of 9–12 identical *c*-subunits. The *c*-ring and central stalk rotate within the $\alpha\beta\alpha\beta\alpha\beta$ assembly in order to cause the conformational changes required for catalytic activity. This rotation is a result of proton transport facilitated by the α-subunit that allows access to a proton (H^+) which binds to a hydrophilic channel between the *c*-ring and the α-subunit, the ring moves around exposing an empty site for the next proton. The bound proton stays attached to the *c*-ring until it has traveled nearly full circle back to the α-subunit where it enters a second hydrophilic channel that opens inside, whereupon it is released leaving an empty site for the next proton. The top of the structure ($\alpha\beta$ and γ complex) is known as the F1 domain and is external to the cell membrane. The crystal structure of the F1 catalytic-domain of ATP synthase is known [20] and is largely consistent with Boyer's [34] rotary-catalysis model.

The FO Domain

The *c*-ring, the α-unit, and other transmembrane protein units (Fig. 9.33) are known as the FO domain and is a cell membrane embedded component of ATPase. The FO domain is thought to be where the antibiotic oligomycin binds, thereby inhibiting the production of ATP. There are numerous examples of mutational studies that reveal a possible binding site for oligomycin at the interface between the α-subunit and the *c*-ring [35, 36].

9.11 Docking Oligomycin into ATPase

The study reported here locates and characterizes the oligomycin binding mode and site within the FO subunit of an homology model of the ATP synthase FO subunit from *E. coli*, employing the three oligomycin molecular structures independently.

9.11.1 ATP Synthase FO Model

Inspection of the European Bioinformatics Institute (EBI) database revealed a model [37] of the FO structure derived from NMR measurements. This is a solution study as opposed to a single crystal study. The PDB entry code for this structure is 1c17. However, this structure does not represent the complete FO complex as it has omitted the vital section with residues numbered 167–198 in the α-subunit. Since a complete model of the α-subunit is needed for oligomycin binding studies, it was necessary to first build this missing section into the structure, and for this purpose the full sequence of the α-subunit was submitted to the SWISSMODEL (now INTERPRO) fully automated protein structure homology-modeling facility. This server is accessible either via the ExPASy web server, or from the program DeepView (Swiss PDB-Viewer) [38, 39]. The resulting model included the stretch missing in 1c17.

9.11.2 Homology Modeling

The homology modeling procedure consists of four sequential steps:
1. Template selection
2. Target-template alignment
3. Model construction
4. Model assessment

Template selection and target-template alignment may be carried out together. The target is the unknown structure being investigated and the template is a sought-for structure with a highly similar amino acid sequence (homology) which it is hoped to find in the Protein Data Base. Usually a number of possible templates will be found. The identification of templates relies on the production of sequence alignments.

The simplest method of template identification relies on serial pairwise sequence alignments aided by database search techniques. Choosing the best template from among the candidates is a key step, and can affect the final accuracy of the structure significantly. Sometimes several homology models are produced for a single query sequence, with the most likely candidate chosen only in the final step.

9.11.3 Refining the Model: Energy Minimization

The model provided by SWISSMODEL [40] was superimposed onto the coordinates of the corresponding subunits in the PDB structure 1c17 using the program MOE [41] and an energy minimization of the three chains involved in ligand binding was then carried out.

Energy Minimization
Molecules are not static structures, and many biological processes are mediated by molecular movement. However, a static molecular model, where the atoms are located at an energy minimum, is a useful simplification; ideally this energy minimum should be the global minimum for the given collection of atoms.

The interactions between neighboring atoms include spring-like forces (representing chemical bonds), and steric and van der Waals forces between nonbonded atoms. The Lennard-Jones potential is commonly used to describe nonbonded interactions, and electrostatic interactions are computed from Coulomb's law. The complete potential energy function E for a biomolecule also includes terms that constrain the deviation of bond lengths, bond angles, and torsion angles from their equilibrium values, in addition to the nonbonded interactions. Thus, we write

$$E = E_{\text{bonds}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{non-bonded}} + E_{\text{electrostatic}}$$

The set of parameters characterizing the potential function consists of equilibrium bond lengths, angles, partial charges, force constants, and van der Waals parameters, collectively known as a force field. Different implementations of molecular mechanics use differing mathematical expressions and different parameters for the potential function. The force fields in use today have been developed by a combination of high-level quantum-mechanical calculations and fitting to experimental data.

Given a model for the potential energy of the asymmetric unit, an equilibrium (zero force) set of atomic positions can be found by energy minimization. The method of *conjugate gradients* is suitable for finding the local energy minimum. Starting with an initial intelligent guess for the atomic positions, the method of conjugate gradients will minimize the residual forces until local equilibrium

is achieved. However, this is not necessarily the minimum energy structure: as with the refinement that we have discussed in Sect. 8.4ff, a true, or global, minimum will not be reached unless we have a reasonably accurate set of coordinates with which to begin. Lower energy states are more thermodynamically stable and play the most important role in chemical and biological processes.

After the FO model, provided by SWISSMODEL, had been superimposed onto the coordinates of the corresponding subunits in the PDB structure 1c17, using the program MOE [41], an energy minimization of the three chains involved in ligand binding was then carried out until all van der Waals contacts had been eliminated. Two models which were used for the docking were created from the resultant model, one containing the full 12 *c*-subunits (Model 1) and the other including only the three subunits involved in binding with the ligand (Model 2).

9.11.4 Creation of a Pocket for Docking Oligomycin into the ATP Synthase FO

Inspection of the completed FO model indicated that in its present form there was insufficient space between the subunits, thought to be involved in the binding of oligomycin, in which to fit this ligand. It was therefore necessary to create an artificial pocket in FO for this purpose. This is a reasonable requirement because, when actually docking in vivo, it is quite feasible that the ligand would have to force its way into the receptor.

Consequently a small space at the desired location was found using the facility Site Finder in MOE. Dummy atoms were then placed into this pocket and Oligomycin B was manually overlaid onto these dummy atoms and an energy minimization [42] was carried out using the program AMBER99 on the three chains involved in binding (two *c*-subunits and the α-subunit). The backbone atoms of these chains were fixed in order to retain the overall integrity of the structure. To develop further the binding pocket, Oligomycin A was overlaid on to the coordinates of the energy minimized ligand within the pocket, and energy minimization was again performed. This procedure was followed during the creation of the pocket in both Model 1 (12 *c*-subunits and 1 α-subunit) and Model 2 (subunits involved in ligand binding) which was then capable of accommodating any of the oligomycins.

Docking Oligomycin into the ATP Synthase FO

The docking studies performed have highlighted contacts to the following residues that are known to be essential for proton transport through the membrane: Asp61, Arg210, Asn214, and Gln252 [43]. Figure 9.36 demonstrates Oligomycin C bound in place, contacting these integral residues. According to the docking study, these residues were contacted preferentially by specific ligating groups on the antibiotic. The ligating group that appears to be most essential to binding is O(24) with a preference for Asn214; this is a consistent theme for all forms of oligomycin and both FO models. According to the results the ligand has a tendency to arrange itself in such a way that the outer ligating groups O(24), O(28), O(32), and O(14) are positioned such that they will coordinate with Asn214. This allows for a certain amount of mobility within the pocket permitting the ligating groups to contact other residues, for example O(28)–Gln252 and O(2)–Asp61.

Conclusions

Docking of all oligomycins to the FO pocket indicated that the B and C forms would bind more tightly than the A form. Consideration of the single crystal X-ray data alone indicated the B form to be the best inhibitor and that O(24) was the most important ligating group for binding. This was supported

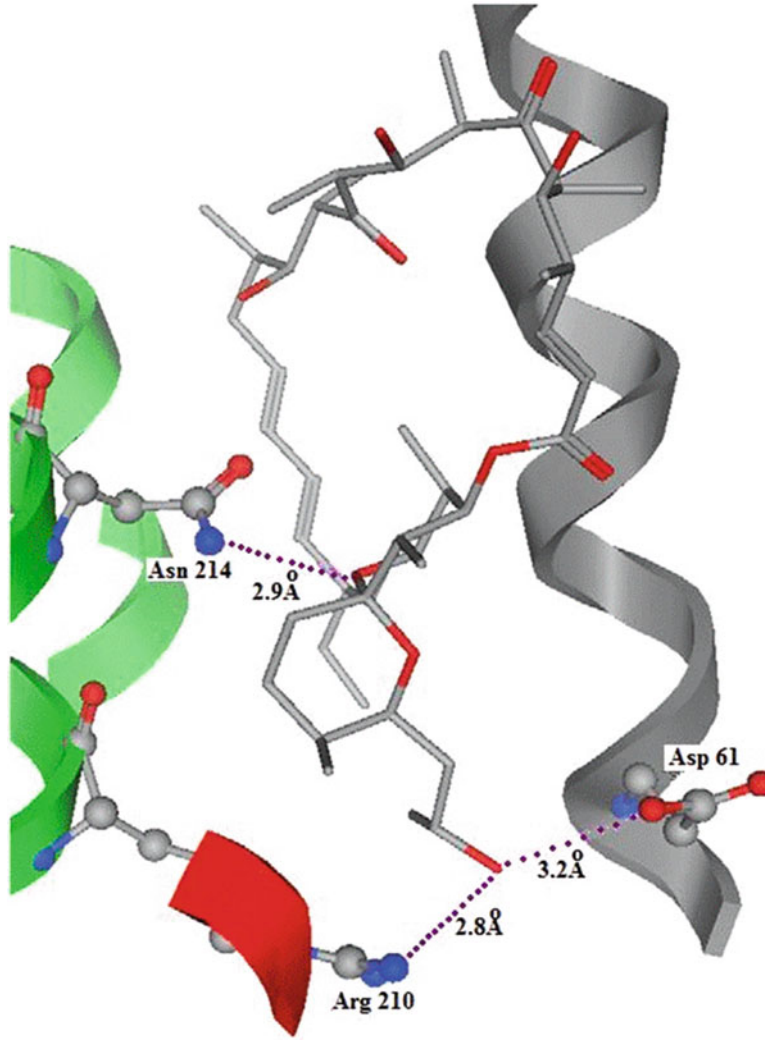


Fig. 9.36 Example of the docking results: oligomycin C docked, showing the hydrogen-bonding to Asn 214, Arg 210, and Asp 61

by the docking data. The latter revealed Asn214 and other key proton translocating residues to be the main residues contacted by the inhibitor.
These data have allowed the binding modes of different forms of oligomycin to be deduced from X-ray single crystal data supported by molecular modeling and computational docking studies.

9.12 Problems

9.1. The unit cell of euphenyl iodoacetate, $C_{32}H_{53}O_2I$, has the dimensions: $a = 7.26$ Å, $b = 11.55$ Å, $c = 19.22$ Å, and $\beta = 94.07°$. The space group is $P2_1$ and $Z = 2$. Figure P9.1 is the sharpened and "selected" Harker section $(u, \frac{1}{2}, w)$.

Fig. P9.1 Sharpened Harker section $(u, \frac{1}{2}, w)$ for euphenyl iodoacetate

$(\sin\theta)/\lambda$	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40
f_1	53.0	51.7	48.6	45.0	41.6	38.7	36.1	33.7	31.5

(a) Determine the x and z coordinates for the iodine atoms in the unit cell.
(b) Atomic scattering factor data for iodine are tabulated below; temperature factor corrections may be ignored.
Determine probable signs for the reflections 001 ($F_o = 40$), 0014 ($F_o = 37$), 106 ($F_o = 33$), and 300 ($F_o = 35$). Comment upon the likelihood of the correctness of the signs which you have determined.
(c) Calculate the length of the shortest iodine-iodine vector in the structure.

9.2. The following |E| values were determined for the [100] zone of a crystal of space group $P2_1/a$. Prepare a $\sum_2$ listing, assign an origin, and determine signs for as many reflections as possible, and give reasons for each step that you carry out. In this projection, two reflections for which the indices are not both even may be used to specify the origin.

$0kl$	|E|	$0kl$	|E|
0018	2.4	0310	1.9
011	1.0	0312	0.1
021	0.1	059	1.9
024	2.8	081	2.2
026	0.3	0817	1.8
035	1.8	011,7	1.3
038	2.1	011,9	2.2

9.3. The chart in Fig. P9.2 shows |E| values taken from the $hk0$ data for potassium hydrogen squarate. Take an origin at the center of a sheet of centimeter graph paper and copy the |E| values on to it, using the top left portion of each appropriate square. For each |E| value plotted, add the corresponding symmetry-related |E| values to the other three portions of the graphical reciprocal space representation. Remember to change the signs of |E| in accordance with the space group symmetry. Next, draw an identical chart on transparent paper, but with the |E| values in the bottom right portion of each appropriate square.

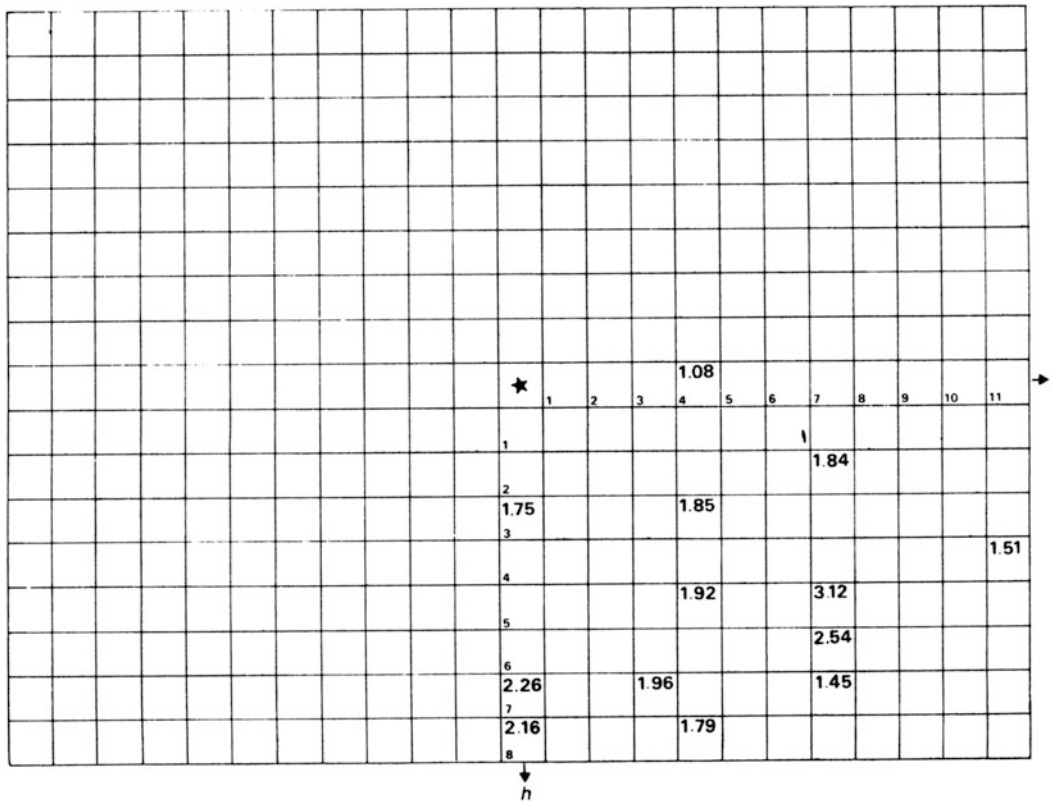


Fig. P9.2 Chart of the |E(*hk*0)| data for the KHSQ structure

(a) Obtain a $\sum_2$ listing: take each plotted |E| value in turn on the original chart and superimpose the transparency, with the origin of the transparency over the chosen |E| value and keeping the two sets of *h*, *k* axes in register. Look for any superimposed |E| values. A $\sum_2$ triplet is given by the |E| value on the original chart under the origin of the transparency, together with the superimposed values, with the *hk* indices read, one from the original and the other from the transparency. Thus, with the origin of the transparency on the original |E(300)|, we read 840 on the original and 540 on the transparency. Set up the $\sum_2$ listing as follows:

$\mathbf{h}$	$\lvert E_{\mathbf{h}}\rvert$	$\mathbf{k}$	$\lvert E_{\mathbf{k}}\rvert$	$\mathbf{h}-\mathbf{k}$	$\lvert E_{\mathbf{h}-\mathbf{k}}\rvert$	$\lvert E_{\mathbf{h}}\rvert\lvert E_{\mathbf{k}}\rvert\lvert E_{\mathbf{h}-\mathbf{k}}\rvert$
300	1.75	840	1.79	$\bar{5}\bar{4}0$	1.92	6.01
.						
.						
.						
700	2.26 · · ·					

The rationale for the graphical procedure may be seen from Fig. P9.3.

(b) Assign an origin in accordance with the rules discussed in the previous chapter and allocate signs to as many reflections as possible; use symbols if necessary. It may be assumed that the products $|E_{\mathbf{h}}||E_{\mathbf{k}}||E_{\mathbf{h}-\mathbf{k}}|$ are all sufficiently large for the indications to be accepted.

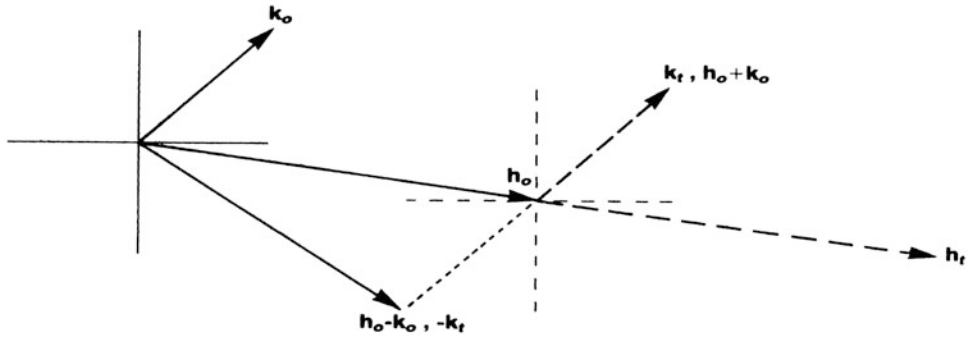


Fig. P9.3 $\sum_2$ relationships: subscript o refers to the original chart, and subscript t refers to the transparency. If the origin of the transparency is placed over a chosen $\mathbf{h}_o$, it may be seen that coincidences of |E| values given by $+\mathbf{k}_t$, $\mathbf{h}_o + \mathbf{k}_o$ or by $-k_t$, $\mathbf{h}_o - \mathbf{k}_o$ will represent $\sum_2$ triplets. It should be noted that this technique applies only to centrosymmetric projections of space groups

9.4. The dithiete compound discussed in Sect. 9.4 has the sulphur atoms placed as follows:

	x	*y*	*z*
S(1)	0.2092(2)	$\frac{3}{4}$	−0.1522(2)
S(2)	0.1095(1)	0.6933(1)	−0.1216(2)

(a) Calculate the S(1)–S(2) bond distance and its esd.

(b) Draw a diagram to show the positions and weights of the Patterson vector disposition around the origin of the unit cell, as seen in projection on to the *xz* plane. For the purpose of the drawing, a β-angle of 90° can be assumed.

References

1. Ladd MFC, Povey DC (1972) J Cryst Mol Struct 2:243
2. International critical tables 3:80 (1928), 7:73 (1930); Mellor's treatise on inorganic and theoretical chemistry 2:941 (supplement, 1961)
3. Bull RJ et al (1973) Cryst Struct Commun 2:625
4. Ladd MFC, Povey DC (1976) Acta Crystallogr B32:1311
5. Kusters W, de Mayo P (1974) J Am Chem Soc 96:3502
6. Boar R et al (1975) J Chem Soc Chem Commun 756
7. Hoppe W (1965) Angew Chem 74:484
8. Arnott S et al (1961) J Chem Soc 4183
9. Stewart JM et al (1972) The X-ray system—version. University of Maryland, College Park
10. Hamilton WC (1965) Acta Crystallogr 18:502
11. Cymerman J, Willis JB (1951) J Chem Soc 1332
12. Ladd MFC (1976) Unpublished results
13. Boar R (1975) loc. cit.
14. Palmer RA, Potter BS (2008) J Chem Crystallogr 38:243
15. Lardy HA (1980) Pharmacol Ther 11:649
16. Fossati G et al (2003) J Immunol 170:1964
17. Altamura N et al (1996) FEBS Lett 382:111
18. von Glehn M et al (1972) FEBS Lett 15:267
19. Fernandez-Moran H (1962) Circulation 26:1039
20. Abrahams JP et al (1994) Nature 370:621
21. Boyer PD (1993) Biochim Biophys Acta 1140:215
22. Enraf-Nonius (1988) CAD-4 software. Enraf-Nonius, Delft, Holland

23. North ACT et al (1968) Acta Crystallogr A24:351
24. Sheldrick GM (1986) SHELXS-86. University of Göttingen, Germany
25. Sheldrick GM (1993) SHELXL-97. University of Göttingen, Germany
26. Farrugia LJ (1998) J Appl Crystallogr 32:837
27. Spek AL (1990) Acta Crystallogr A46:C34
28. Farrugia LJ (1997) J Appl Crystallogr 30:565 (based on ORTEP-III (v 1.0.3) by C. K. Johnson and M. N. Burnett)
29. Merrit EA, Bacon DJ (1997) Methods Enzymol 277:505 (implemented in WinGX and generated by ORTEP-III for Windows)
30. Humphrey W et al (1996) J Mol Graphics 14:33
31. Berdy J et al (1987) Handbook of antibiotic compounds, vol IV. CRC Press, Boca Raton, FL
32. Ramirez F et al (1982) Eur J Biochem 121:275
33. Flack HD (1983) Acta Crystallogr A39:876
34. Boyer PD (1994) Biochim Biophys Acta 1140:215
35. Devenish RJ et al (2000) J Bioenerg Biomembr 32:507
36. Galanis M et al (1989) FEBS Lett 249:333
37. Rastogi VK, Girvin ME (1999) Nature 402:263
38. Arnold K et al (2006) Bioinformatics 22:195
39. http://www.expasy.org
40. http://www.ebi.ac.uk/interpro/
41. http://www.chemcomp.com/
42. Wang J et al (2000) J Comp Chem 21:1049
43. de Jonge MR et al (2007) Proteins Struct Funct Bioinform 67:971

Bibliography: Published Structure Analyses

Acta Crystallographica (the early issues (1948–1968) are most suitable for the beginner)
Journal of Crystallographic and Spectroscopic Research
Journal of Chemical Crystallography (formerly Journal of Crystal and Molecular Structure)
Zeitschrift für Kristallographie

General Structural Data

Kennard O et al. Cambridge Structural Database (CSD)
Wyckoff RWG (1963–1966, 1968, 1971) Crystal structures, vols 1–6. Wiley, New York

Molecular Mechanics

Frenkel D, Smit B (1996) Understanding molecular simulation. Academic, New York