AUTOMATION OF THE SYMBOLIC ADDITION METHOD

VERHANDELINGEN DER KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN, AFD. NATUURKUNDE
EERSTE REEKS – DEEL XXV, No. 5

AUTOMATION OF THE SYMBOLIC ADDITION METHOD

APPLICATION TO SOME COMPOUNDS FROM THE VITAMIN-A CHEMISTRY

H. SCHENK

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM-LONDON — 1969

LIBRARY OF CONGRESS
CATALOGUE CARD NUMBER 72-94060

AANGEBODEN MAART 1969
AANVAARD APRIL 1969
GEPUBLICEERD: NOVEMBER 1969

CONTENTS

CHAPTER 1

OBJECTS, METHODS AND RESULTS; A CHRONOLOGICAL REPORT

1.1 *Introduction*

Chapter 1 contains a report about the realization of the objects of the investigations.

The motives of the investigations, the usefulness of the employed methods and the discussion of the main aspects of the determined crystal structures are emphasized.

1.2 *First object: the structure determination of retro vitamin-A acid*

In the last fifteen years some crystal structures of vitamin-A-related compounds and carotenoids have been determined in the Laboratory for Crystallography, University of Amsterdam and elsewhere, namely:

1) all-*trans*-β-ionylidene-γ-crotonic acid (Eichhorn and MacGillavry, 1959).
2) *cis*-β-ionylidene-γ-crotonic acid (Koch and MacGillavry, 1963).
3) vitamin-A acid (Stam and MacGillavry, 1963).
4) 15,15′-dehydro-β-carotene (Sly, 1964).
5) β-carotene (Sterling, 1964).
6) 15–15′-dehydrocanthaxanthin (Bart and MacGillavry, 1968a).
7) canthaxanthin (Bart and MacGillavry, 1968b).

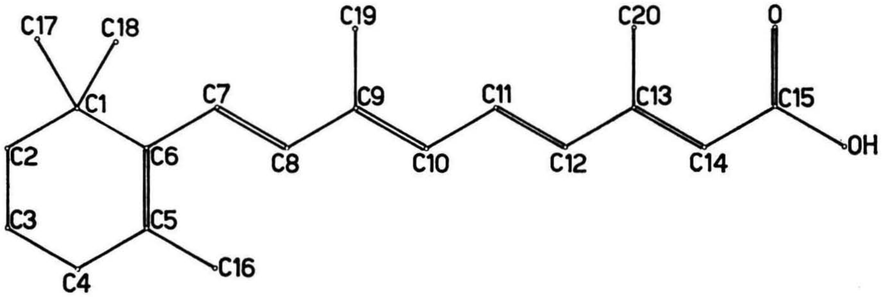


Fig. 1.2.1. The numbering of the atoms in vitamin-A acid.

In Fig. 1.2.1 the structural formula of vitamin-A acid is shown; all other compounds have related structures. The structures have some details in common, as has been reviewed thoroughly by Bart and MacGillavry (1968b). We mention two remarkable features:

a) The nearly planar polyene chain is appreciably curved in its plane. The angles opposite the methyl groups attached to the conjugated chain are smaller than the expected 125° and the angles opposite the double bond are larger than 110°. In double bond systems free from steric hindrance one usually finds angles of 125, 125 and 110°, as is shown in fig. 1.2.2.

b) The attachment of the ring and chain in all structures except in (1) and (2) is approximately single-*cis*. The deviations from the single-*trans* configuration about the bond C(6)–C(7) for the different structures are: (1) 10°, (2) 100°, (3) 145°, (4) 136°, (5) 145°, (6) 152° and (7) 137°.

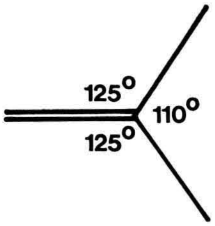


Fig. 1.2.2. Angles around a carbon atom in a double bond system free from steric hindrance

The effects a) and b) have been interpreted as being the result of intramolecular steric hindrance between the various methyl groups and atoms not bound to the same atom as the methyl group considered. The corresponding distances have a minimum length of 2.9 to 3.0 Å (Stam and MacGillavry, 1963).

The first object of this thesis was the determination of the crystal structure of *retro* vitamin-A acid. The molecular formula is $C_{20}H_{28}O_2$ and is schematically shown in fig. 1.2.3. The only difference with vitamin-A acid is a shift in the conjugated double bond system, as is evident from fig. 1.2.1 and fig. 1.2.3.

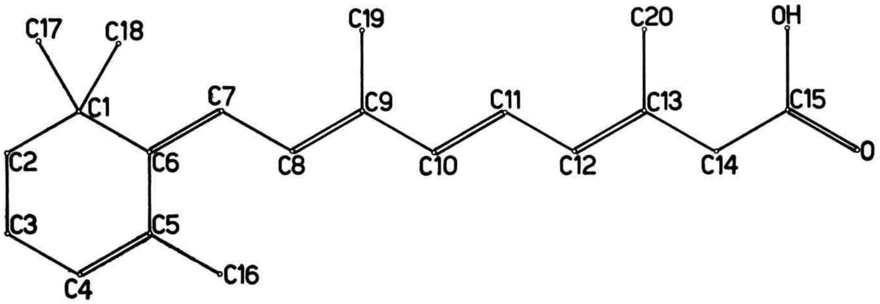


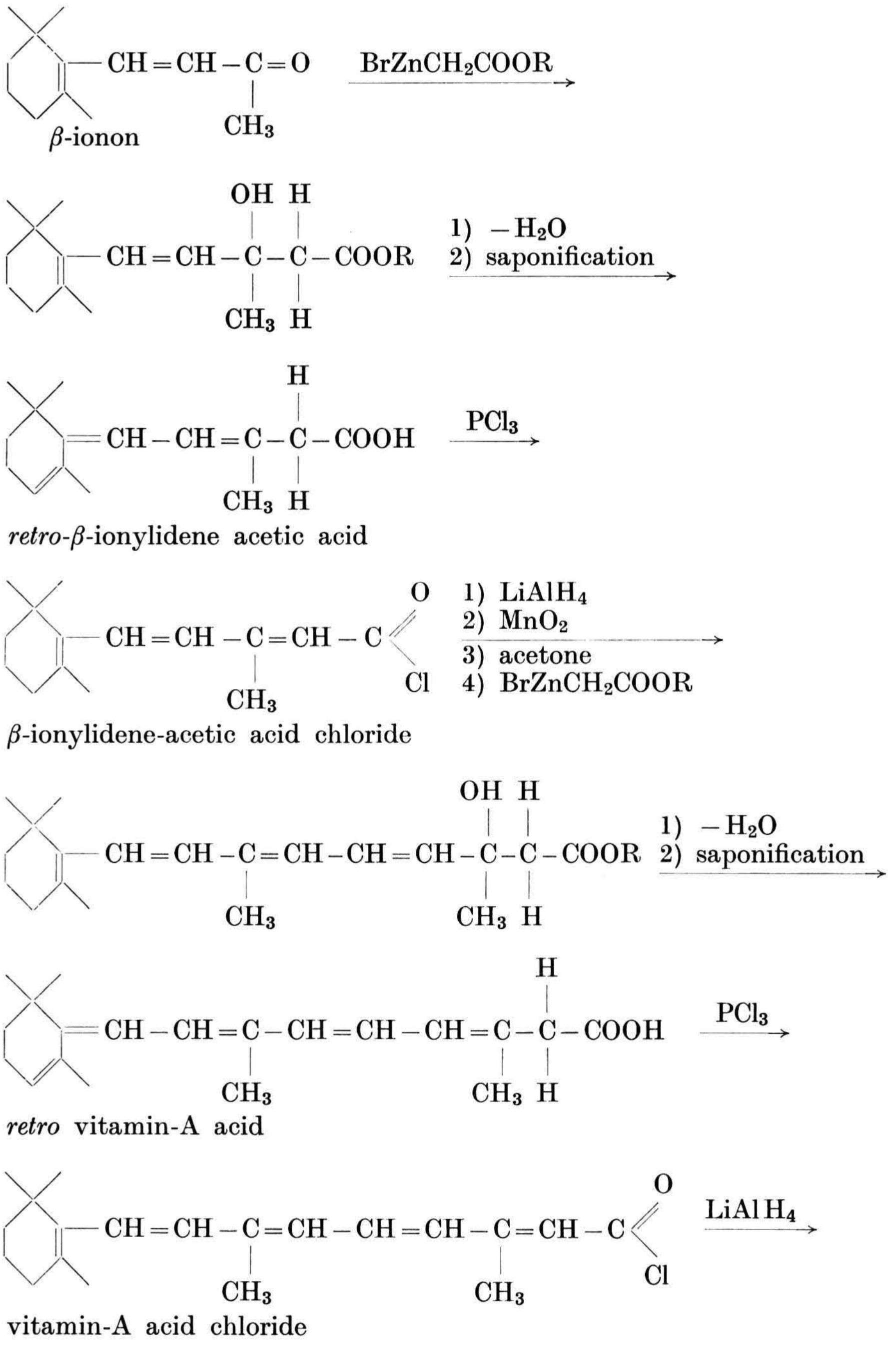
Fig. 1.2.3. Numbering of the atoms in *retro* vitamin-A acid.

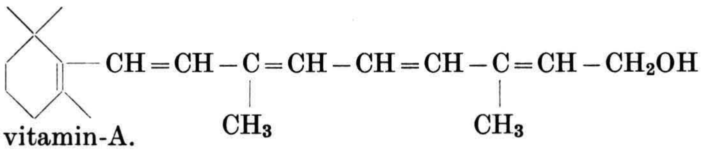
The bond C(6)–C(7) in the retro compound is a double one. It is expected, that the configuration about this bond is either pure *cis* or pure *trans*. The main goal of the investigation of *retro* vitamin-A acid is to determine the stereochemical aspects of this ring-chain attachment.

1.3 *Some chemical information about retro vitamin-A acid*

Retro vitamin-A acid, as well as *retro-β*-ionylidene acetic acid are intermediate stages in the vitamin-A synthesis of HUISMAN, SMIT, VAN LEEUWEN and VAN RIJ (1956).

The reaction scheme of this synthesis is as follows:





$$\mathrm{-CH{=}CH{-}\underset{\underset{CH_3}{|}}{C}{=}CH{-}CH{=}CH{-}\underset{\underset{CH_3}{|}}{C}{=}CH{-}CH_2OH}$$

vitamin-A.

According to Huisman c.s. it is likely that *retro*-β-ionylidene acetic acid (*retro*-C15-acid) and *retro* vitamin-A acid (*retro*-C20-acid) are the all-*trans* isomers.

No other isomers were obtained in a pure crystalline form.

1.4 *Determination of retro Vitamin-A acid, a first attempt*

Our first attempt to determine the structure of *retro* vitamin-A acid had been carried out with the usual Patterson method. The interpretation of the Patterson synthesis was based on an all-*trans* model of the conjugated system. We found easily the plane of this system and the direction of the zig zag.

In order to find the positional parameters of the atoms of our model we needed additional information about the carboxylic group. We did not succeed in finding even the plane of the carboxylic group and therefore we used the model, shown in fig. 1.4.1, and also packing considerations. All trials were unsuccessful.

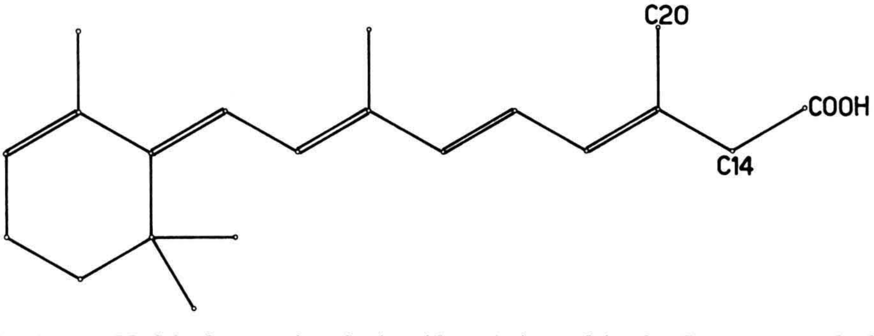


Fig. 1.4.1. Model of *retro* vitamin-A acid, as it is used in the Patterson method.

1.5 *Retro-β-ionylidene-acetyl-p.Br-anilide*

In the mean time in our laboratory mrs S. Paul–Roy had the same difficulties in unraveling the Patterson synthesis of *retro*-β-ionylidene-acetic acid.

The determination of the structure of a heavy atom derivative of this acid, the p.Br-anilide (PAUL–ROY, SCHENK and MACGILLAVRY, 1969) explained the failure of the Patterson method in the determination of the acid as such. The configuration of the chain with respect to the double bond C(8) = C(9) is *cis*, whereas the *trans*-configuration was expected (see fig. 1.5.1).

It was assumed that *retro* vitamin-A acid probably has the same *cis*-configuration with respect to the double bond C(12) = C(13).

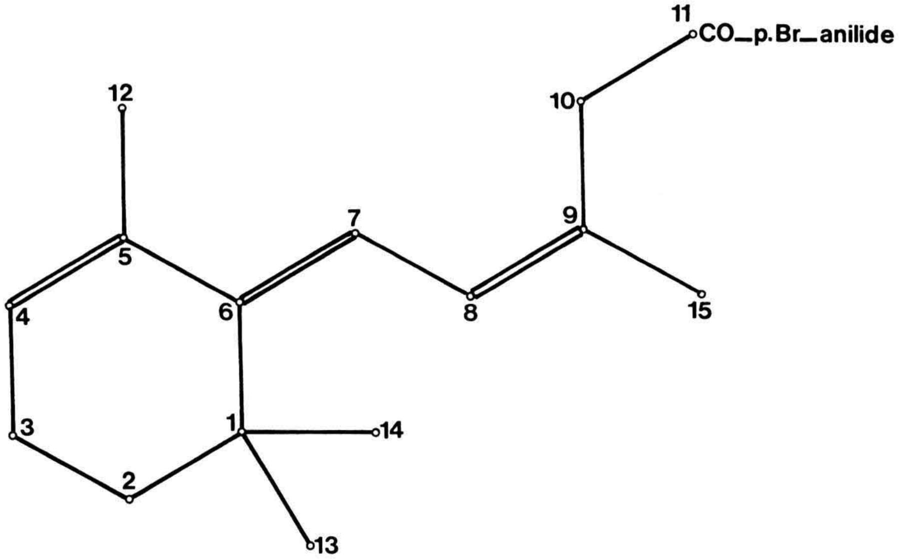


Fig. 1.5.1. A schematical drawing of the structure of *retro*-β-ionylidene-acetyl-p.Br.-anilide.

1.6 *Object: automation of the symbolic addition method for centrosymmetric structures*

It was clear that the conventional Patterson technique had failed. Therefore we tried to solve the structure of *retro* vitamin-A acid with a direct method, namely the Symbolic Addition Method of KARLE and KARLE (1966), which is briefly reviewed in chapter 2.

The possibility to employ this method in our laboratory had been opened up by Mr. H. v. d. Meer. He wrote a computer program which prints out the Σ_2-listing, a very essential tool in the Symbolic Addition procedure.

The determination of the signs of the various reflections by hand, however, proved to be very laborious (chapter 3).

The next object was the semi-automation or better the full automation of the method for centrosymmetric structures.

1.7 *Result: semi-automation of the Symbolic Addition Method*

With two computer programs it was possible to speed up the Symbolic Addition procedure in the centrosymmetric case with a factor ten with respect to the hand application of the method (1.6).

The first program computes symbolic signs for the reflections. The second program computes criteria for the quality of the possible solutions for the various unknown symbols.

The two programs are described in chapter 4 and have been tested in the structure determination of 1,1,4-trimethyl-5-carboxyl-cyclopentene-4.

1.8 *The crystal structure of* 1,1,4-*trimethyl*-5-*carboxyl-cyclopentene*-4.

This structure, also referred to as C9-acid (see fig. 1.8.1) was found with the Symbolic Addition Method without any ambiguity.

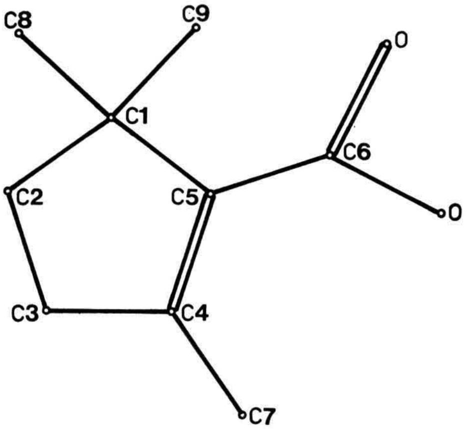


Fig. 1.8.1. Numbering of the atoms in C9-acid.

The main features are:

1) The conjugated system, including the bonds C(4) = C(5) and C(6) = O(1) is *trans*.
2) The plane through the carboxylic group is rotated approximately 20° about the bond C(5)–C(6) out of the plane through the double bond C(4) = C(5). This is caused by steric interaction of the methyl group C(9) with O(1).
3) The bond lengths in the cyclopentene-ring are in general agreement with the corresponding lengths in cyclohexene-rings, but the endo-angles of the ring are decreased about 5 to 15°.

Full details of the structure analysis are given in chapter 6.

1.9 *Result*: *The crystal structure of retro-vitamin-A acid*

The structure of *retro* vitamin-A acid was determined in the same way (see 1.8). There were no difficulties of importance, from which one may conclude that in this case the symbolic addition method was a much stronger tool than the Patterson method in solving the phase-problem.

The main aspects of the structure are:

1) The attachment of ring and chain is *trans*.
2) The conjugated chain is *cis* with respect to the double bond C(12) = C(13).

Full details of the crystal structure and its determination are given in chapter 7.

1.10 *Result: automation of the symbolic addition method for centrosymmetric structures*

With one completely new computer program and self-iterative versions of the semi-automatical programs (see 1.8) the Symbolic Addition Method for centrosymmetric structures has been completely automatized. These programs are described in chapter 5.

The structure determination of *retro* β-ionylidene acetic acid was used as a test case.

1.11 *The crystal structure of retro β-ionylidene acetic acid*

The structure was solved by the automatic symbolic addition procedure, as described in chapter 5.

The main features of the structure are:

1) The ring and chain are attached to each other in the *trans*-configuration.
2) With respect to the double bond C(8) = C(9) the chain is *cis*.
3) The two C–O-lengths are of the same magnitude as a result of conformational disorder in the COOH-group.

Full details of the structure determination are given in chapter 8.

CHAPTER 2

SHORT REVIEW OF THE SYMBOLIC ADDITION METHOD

2.1 *Introduction*

The structure determinations, described in this thesis, were carried out with the Symbolic Addition Method of KARLE and KARLE (1966). In their paper Karle and Karle present the theoretical background and practical procedures for phase determination for both centrosymmetric and non-centro-symmetric crystalstructures. Here we shall discuss the method for centrosymmetric crystals only.

2.1.1 Notation

H stands for hkl
$H-H'$ stands for $h-h'$ $k-k'$ $l-l'$
S(H) stands for the sign of F(hkl)

2.2 *Formulae*

For a theoretical discussion of the method we refer to KARLE and KARLE (1966).

However, we want to mention some important expressions:

1) normalized structure factors E_H are:

$$|E_H|^2 = \frac{|F_H|^2}{\varepsilon \sum_{j=1}^{N} f_j^2(H)} \tag{2.2.1}$$

in which F_H = structure factor on absolute scale and corrected for overall temperature movement.
ε = a factor, which corrects for space group extinctions (if half the number of possible reflections is systematically absent in a special group of reflections then $\varepsilon = 2$)
N = the number of atoms in the unit cell.
$f_j(H)$ = the atomic scattering factor

2) for centrosymmetric crystals the following relation holds:

$$\text{sign } (E_H) = \text{sign } (\sum_{H'} E_{H'} \, E_{H-H'}) \tag{2.2.2}$$

with the probability for a positive sign

$$p_+(H) = \tfrac{1}{2} + \tfrac{1}{2} \tanh \sigma_3 \sigma_2^{-3/2} |E_H| \sum_{H'} E_{H'} \, E_{H-H'} \tag{2.2.3}$$

in which $\sigma_n = \sum_{j=1}^{N} Z_j^n$ with Zj = atomic number of atom j.

2.3 *The Σ_2 list*

Essential for the Symbolic Addition method is a list of all combinations of reflections H′ and H–H′ for a given reflection H.
Because this list has to be made for each reflection H, the procedure must be restricted to sufficiently large values of $|E|$.
In the list for each triplet interaction H, H′, H–H′ the probability that $S(H)=S(H')\times S(H-H')$ is given by the correspondingly reduced version of (2.2.3):

$$p(H)=\tfrac{1}{2}+\tfrac{1}{2}\tanh \sigma_3\sigma_2^{-3/2}|E_H||E_{H'}||E_{H-H'}| \qquad (2.3.1)$$

2.4 *Practical application*

KARLE and KARLE (1966) describe a practical procedure to determine the signs of the $|E|$-values. A brief review of the procedure recommended by Karle and Karle follows:

1) Calculation of a Σ_2 list with about ten to fifteen reflections per atom in the asymmetric unit.
2) Selection of five to ten of the largest $|E_H|$-values in each of the eight subgroups, defined by the parity of the h-, k- and l-indices.
3) The signs which specify the origin and one unknown symbol are assigned, bearing in mind that the particular reflections chosen should have many combinations of reflections H′ and H–H′. (See 2.3).
4) Equation 2.2.2 is employed to define as many signs of the largest $|E_H|$ as possible in terms of the specified ones and others that have been newly determined. According to Karle and Karle the probability, given by (2.2.3), has to satisfy the working rule of $p>.97$. If the sign of a reflection is known, the signs of all symmetry-related reflections are known and can be used.
5) One new unknown symbol is assigned to a strong $|E_H|$ with many triplet interactions, followed by step 4. This is necessary, if there are subsets of reflections for which the phase indications (2.2.2) and (2.2.3) in terms of already known signs S(H′) and S(H–H′) are either too weak ($p<.97$) or absent altogether.
6) The next five to ten largest $|E|$-values in each parity subgroup are added. Eventually full use is made of the Σ_2 listing. This is followed by step 4 and if necessary by step 5.
7) In the next step the unknown symbols are eliminated as far as possible, making use of the following considerations:
   a) Often strong relationships between the unknown symbols exist.
   b) Other phase-determing formulae may be employed:
      Σ_1-formula (HAUPTMAN and KARLE (1953)).
      $B_{3,0}$-formula (KARLE and HAUPTMAN (1959)).
      $I_{3,0}$-formula (KARLE and HAUPTMAN (1959)).
   c) Absence of an atom at the origin eliminates solutions for which the signs are predominantly plus.

d) Some sign combinations may lead to a relatively large number of inconsistencies among the contributors to expression (2.2.2). E-Fouriers for internally consistent sets should be computed first (see 8).

e) Known features of the crystal structure can be used to fix certain phases.

8) Calculation of Fourier-series with the signed E-values as coefficients, the so called E-maps.
If an E-map is incorrect, the peaks do not make good chemical sense.

2.5 *Discussion*

2.5.1 Necessity of a Σ_2-computer program

To apply the Symbolic Addition method a computer program to produce the Σ_2-list is needed.

For instance, for a crystal structure with 20 atoms in the asymmetric unit the signs of 200–300 $|E|$-values have to be determined and this requires a rather long Σ_2-list. It is not feasible to prepare such a list by hand.

2.5.2 Some remarks about the practical application of Karle & Karle's procedure, reviewed in 2.4

a) It is not necessary to select 5 to 10 reflections in each parity subgroup (2.4,2). One can take the 50 strongest reflections, provided that the signed reflections belong to all parity groups and do not have systematic absences caused by pseudo translations.

b) In my experience, it is better to take a stronger working rule, for instance $p > .99$ (2.4,4). Then additional unknown symbols must be introduced, but in the last stage these symbols can be easily eliminated because their relationships with other symbols must be very strong. If this should not be the case the more reason for choosing the stronger working rule.

CHAPTER 3

SYMBOLIC ADDITION BY HAND

3.1 *Introduction*

This chapter deals with the description of an Algol-program, written by Mr. H. v. d. Meer of our Laboratory, which produces a Σ_2-list from F^2-values on an arbitrary scale.

Further the hand application of the Symbolic Addition method is discussed.

3.1.1 Notation

H stands for hkl
H − H′ stands for h − h′ k − k′ l − l′

3.2 *Σ_2-listing program (code: SA 1)*

An Algol-program which produces a Σ_2-list has been written by Mr. H. v. d. Meer.

It consists essentially of three parts:

a) calculation of a Wilson plot
b) calculation of E_H-values and statistical quantities
c) printing of a Σ_2-list with the strongest E_H-values.

The input data consist of specific data and a reflection tape containing $|F_H|^2$-values on an arbitrary scale. Equation 2.2.2

$$|E_H|^2 = \frac{|F_H|^2}{\epsilon \sum_{j=1}^{N} f_j \left(\frac{\sin^2\theta}{\lambda^2}\right)} \tag{3.2.1}$$

requires $|F_H|^2$-values on absolute scale corrected for temperature movement.

These corrections are realized in the program by carrying out a Wilson-plot, which gives a scalefactor K and a overall temperature factor B.

In the second part of the program $|E_H|$-values are calculated with a variant of expression (3.2.1):

$$|E_H|^2 = \frac{K^2 \exp\left(2B\frac{\sin^2\theta}{\lambda^2}\right)|F_H|^2}{\epsilon \sum_{j=1}^{N} f_j \left(\frac{\sin^2\theta}{\lambda^2}\right)} \tag{3.2.2}$$

where B and K are the estimated values from the Wilson plot and the $|F_H|$ are the uncorrected observed values. Then various statistical quantities for the normalized structure factor amplitudes $|E_H|$ are calculated.

In the third part of the program a Σ_2-list is printed out for $|E_H|$-values above a given limit. The reflections are sorted after l and k to simplify the searching operations. For each reflection H all possible combinations of H′ and H–H′ are printed out together with the associated probabilities, calculated with expression (2.3.1).

$$p(H) = \tfrac{1}{2} + \tfrac{1}{2} \tanh \sigma_3 \sigma_2^{-3/2} |E_H| \, |E_{H'}| \, |E_{H-H'}|. \qquad (3.2.3)$$

All three reflections H, H′ and H–H′ must belong to the group with $|E_H|$ above the given limit.

3.3 *Discussion*

With the help of the Σ_2-list and the directions for use, reviewed in the preceding chapter, section 2.4, we started trying to solve the phase problem for *retro* vitamin-A acid, space group $P2_1/a$. To determine the structure of this compound with 22 nonhydrogen atoms, we had to find the signs of about 250 reflections. The corresponding Σ_2-list consisted of about 7000 triplet interactions H, H′, H–H′.

Since it is essential to find as many triplet sign relations as possible (see 5.1.1) one has to go many times through all the triplets.

In the case of *retro* vitamin-A acid, at the stage at which the signs of about 90 reflections had been determined in terms of seven unknown symbols and the three origin defining signs, a full cycle through the 7000 triplet interactions took about one week. Proceeding in this way was not very attractive and it was decided to turn over the determination of the symbolic signs to a computer.

In all probability the elimination of the unknown symbols would be a very laborious process too, as for each reflection the relations between the symbols must be taken into account. This elimination process is a problem typically suited for a computer.

CHAPTER 4

HALF-AUTOMATIC SYMBOLIC ADDITION

4.1 *Introduction*

Two computer programs written in Algol 60 are described in this chapter, to speed up the Symbolic Addition procedure for centrosymmetric crystals.

The first program searches for the symbolic signs of unknown reflections in terms of those of the known reflections (code: SA 2).

The second program systematically allots signs to the unknown symbols and computes criteria for the consistency of the various solutions, and computes the densities on the centres of symmetry in the unit cell (code: SA 3).

Further the way to use the programs in the sign determination is discussed.

4.1.1 Notation

H stands for hkl.

H − H′ stands for h − h′ k − k′ l − l′.

S(H) stands for the sign of reflection H.

4.2 *Symbolic Sign Searching Program (SA 2)*

4.2.1 Organisation of the program

The program requires as input data two reflection tapes. Reflection tape 1 consists of all reflections with known symbolic signs. A symbolic sign in this context is either a sign in which unknown symbols are present or a plus or minus sign.

Reflection tape 2 consists of all reflections with $|E|$ above a given limit value. These reflections have not yet been signed.

The program computes first a Σ_2-list with all reflections of reflection tape 1. For any reflection H all possible pairs H′, H–H′ are collected and the symbolic sign S(H′) × S(H–H′) is printed out together with the probability p, calculated with expression 3.2.3. By means of this Σ_2-list all possible signs S(H′) × S(H–H′) can be compared with the sign S(H) to eliminate possible sign errors.

The second part of the program computes a restricted Σ_2-list for the reflections of reflection tape 2. The restriction is that the reflections H′ and H–H′ of a triplet interaction must belong to reflection tape 1, so that for each pair the symbolic sign S(H′) × S(H–H′) can be computed. These symbolic signs are printed out together with their probabilities p. Table 4.2.1 gives an example of the output.

TABLE 4.2.1

Part of a Σ_2-list, calculated by program SA 2.

H	H′	H−H′	argument of tanh	p	symbolic sign
−7−1+1	+2+5−1	−9−6+2	3.44	.999	+ad
	+3+5−2	−10−6+3	1.78	.972	+ad
	+0+10−2	−7−11+3	2.66	.995	+ad
−13−1+1	+4+5−1	+9−6+2	1.60	.961	+bd
	+8+6−3	+5−7+4	1.48	.950	+bd
	+10+11−4	+3−12+5	1.51	.954	+bd
+7−1+1	−1+5+0	+8−6+1	1.46	.949	+a
	+7+11−3	+0−12+4	3.33	.999	+a

4.2.2 How to work with program SA 2

Input

A good criterion for the number of reflections on reflection tape 1 is twice the number of atoms in the asymmetric unit.

Output

The symbolic signs of reflection tape 1 can be checked in the first computed Σ_2-list.

From the second Σ_2-list the symbolic signs of new reflections are accepted if their probabilities are high enough ($p > .99$, see 5.1.1).

It is possible to find relations between the unknown symbols from both the Σ_2-lists. In counting the number of times a definite relation occurs in the first and second Σ_2-lists one has to remember, that each triplet H, H′, H–H′ appears three times in the first list and only once in the second list.

Iterative procedure

The newly determined reflections from the second Σ_2-list are used to extend the number of reflections on reflection tape 1 in a new cycle of computations with SA 2.

Relations between the symbols are only accepted if they appear many times with high probabilities (see 5.4. XI). During the iterative application of program SA 2 the limit value of $|E|$ of reflection tape 2 is lowered in successive steps until the tape contains a number of ten to twelve reflections per atom in the asymmetric unit.

The iterative procedure can be stopped when five to seven reflections per atom in the asymmetric unit have been provided with a symbolic sign. At this stage an analysis of the unknown symbols is possible with program SA 3.

4.3 *Program: Searching for the best signs for the unknown symbols* (*code: SA* 3)

4.3.1 Theoretical considerations

In section 2.4.7 the elimination of the unknown symbols has been reviewed as recommended by KARLE and KARLE (1966). The steps, which are handled best in a computer program, are:

1) Looking for strong relations between the unknown symbols (2.4,7 a).
2) Checking the +/− distribution of the signs (2.4,7 c).
3) Checking if an assignment of a sign to the unknown symbols leads to inconsistencies (2.4,7 d).

However 1) is contained in 3); a strong relation would cause relatively few inconsistencies.

It is necessary to consider item 2) in some detail. At any of the centres of symmetry in the unit cell no electron density is permitted if none of the corresponding special positions is occupied. This requires the sum of the signed structure factor magnitudes to cancel in some of the parity groups. Example:

Consider the spacegroup $P2_1/n$. The formula for the electron density, written as a Fourier series, is

$$\varrho(xyz) = \frac{4}{V}\Big[\sum_h \sum_k \sum_l^{h+k+l=2n} (F_{hkl} \cos 2\pi hx \cos 2\pi ky \cos 2\pi lz + \\ - F_{hkl} \sin 2\pi hx \cos 2\pi ky \sin 2\pi lz) + \\ + \sum_h \sum_k \sum_l^{h+k+l=2n+1} (-F_{hkl} \sin 2\pi hx \sin 2\pi ky \cos 2\pi lz + \\ - F_{hkl} \cos 2\pi hx \sin 2\pi ky \sin 2\pi lz)\Big].$$

The eight centres of symmetry in the cell have the following coordinates:

$0\ 0\ 0$, $0\ 0\ \tfrac{1}{2}$, $0\ \tfrac{1}{2}\ 0$, $\tfrac{1}{2}\ 0\ 0$,
$\tfrac{1}{2}\ \tfrac{1}{2}\ \tfrac{1}{2}$, $\tfrac{1}{2}\ \tfrac{1}{2}\ 0$, $\tfrac{1}{2}\ 0\ \tfrac{1}{2}$, $0\ \tfrac{1}{2}\ \tfrac{1}{2}$.

For the coordinates $x=\tfrac{1}{2}$ and $x=0$ $\sin 2\pi hx = 0$. The same applies to the corresponding y- and z-values. So for the centres of symmetry the electron density is given by

$$\varrho(xyz) = \frac{4}{V}\Big[\sum_h \sum_k \sum_l^{h+k+l=2n} F_{hkl} \cos 2\pi hx \cos 2\pi ky \cos 2\pi lz\Big] = 0.$$

For each of the centres of symmetry we can write the electron density in terms of four subsummations involving four parity groups, as follows:

$$\varrho(000) = \frac{4}{V}[\text{sum eee} + \text{sum eoo} + \text{sum ooe} + \text{sum oeo}] = 0$$

$$\varrho(00\tfrac{1}{2}) = \frac{4}{V}[\text{sum eee} - \text{sum eoo} + \text{sum ooe} - \text{sum oeo}] = 0$$

$$\varrho(0\tfrac{1}{2}0) = \frac{4}{V}[\text{sum eee} - \text{sum eoo} - \text{sum ooe} + \text{sum oeo}] = 0$$

$$\varrho(\tfrac{1}{2}00) = \frac{4}{V}\,[\text{sum eee} + \text{sum eoo} - \text{sum ooe} - \text{sum oeo}] = 0$$

$$\varrho(\tfrac{1}{2}\tfrac{1}{2}0) = \frac{4}{V}\,[\text{sum eee} - \text{sum eoo} + \text{sum ooe} - \text{sum oeo}] = 0$$

$$\varrho(\tfrac{1}{2}0\tfrac{1}{2}) = \frac{4}{V}\,[\text{sum eee} - \text{sum eoo} - \text{sum ooe} + \text{sum oeo}] = 0$$

$$\varrho(0\tfrac{1}{2}\tfrac{1}{2}) = \frac{4}{V}\,[\text{sum eee} + \text{sum eoo} - \text{sum ooe} - \text{sum oeo}] = 0$$

$$\varrho(\tfrac{1}{2}\tfrac{1}{2}\tfrac{1}{2}) = \frac{4}{V}\,[\text{sum eee} + \text{sum eoo} + \text{sum ooe} + \text{sum oeo}] = 0$$

in which

$$\text{sum eee} = \sum\sum\sum^{h,\,k,\,l,\,=2n} F_{hkl}$$

$$\text{sum eoo} = \sum\sum\sum^{h=2n,\,k,\,l=2n+1} F_{hkl}$$

$$\text{sum ooe} = \sum\sum\sum^{h,\,k=2n+1,\,l=2n} F_{hkl}$$

and

$$\text{sum oeo} = \sum\sum\sum^{h,\,l=2n+1,\,k=2n} F_{hkl}$$

These eight equations hold true only if in each of the parity groups eee, eoo, ooe and oeo the sum of the structure factors cancels, where e stands for even and o for odd in order of h, k and l respectively.

Thus the criterion for the +/− distribution in spacegroup $P2_1/n$ is: In each of the parity groups eee, eoo, oeo and ooe the sum of signed E-values must approximately cancel.

In the same way one can find the relevant criterion for each space group.

4.3.2 Organisation of program SA 3

Input data

The program requires as input data two reflection tapes. Reflection tape 1 consists of all reflections with known symbolic sign. Reflection tape 2 consists of all reflections above a certain limit value of $|E|$. These reflections are not signed.

Signing the reflections

Each of the unknown symbols is taken as − and + in a systematic way, so that if there are n unknown symbols this results in 2^n different solutions. To each of these solutions corresponds a set of signs for the reflections of tape 1.

Σ_2-lists

To check the consistency of the different solutions two Σ_2-lists are calculated, one with the reflections of tape 1 and a restricted one for the reflections of tape 2.

The restriction is, that of each triplet H, H′, H–H′ the last two must belong to reflection tape 1. For each triplet the sign S(H′) × S(H–H′) for the reflection H is printed out for each of the 2^n solutions. This is illustrated for one reflection H in table 4.3.1.

TABLE 4.3.1

Part of a Σ_2-list, calculated by program SA 3. All reflections H′ and H−H′ for H=12 12 2 are given with the signs s(H′) × s(H−H′) for eight solutions.

H′	H–H′	argument of tanh	p	S(H′) × S(H−H′) for solution							
				1	2	3	4	5	6	7	8
0 18 0	12 $\bar{6}$ 2	1.28	.93	−	+	+	−	−	+	−	−
$\bar{2}$ 12 $\bar{1}$	14 0 3	1.52	.95	−	+	−	+	−	+	−	+
8 2 $\bar{2}$	4 10 4	1.96	.98	−	+	−	+	−	+	−	+
4 8 $\bar{2}$	8 4 4	1.81	.97	−	+	+	−	−	+	+	−
14 0 $\bar{3}$	$\bar{2}$ 12 5	.96	.87	−	+	−	+	−	+	−	+
12 4 $\bar{3}$	0 8 5	.83	.84	+	−	−	+	+	+	−	−
6 10 $\bar{3}$	6 2 5	1.66	.97	−	+	−	+	−	+	−	+

The quality of a reflection H

For each reflection H we have a number of pairs H′, H–H′ which give a column of signs S(H) = S(H′) × S(H–H′) for each solution (see table 4.3.1). A measure for the quality of the sign of a reflection H in a definite solution is the agreement of the signs S(H′) × S(H–H′) in the corresponding column.

However a better quality criterion is found in the following way: In a column the arguments of the tanh from expression 2.3.1. are summed for all S(H′) × S(H–H′) = +. This sum is called sumplus:

$$\text{sumplus} = \sum_{\substack{\text{H}'\text{with} \\ \text{S(H}')\times\text{S(H}-\text{H}')=+}} \sigma_3\,\sigma_2^{-3/2}|E_H|\;|E_{H'}|\;|E_{H-H'}|$$

The same is done for all S(H′) × S(H–H′) = − in the column:

$$\text{summinus} = \sum_{\substack{\text{H}'\text{with} \\ \text{S(H}')\times\text{S(h}-\text{H}')=-}} \sigma_3\,\sigma_2^{-3/2}|E_H|\;|E_{H'}|\;|E_{H-H'}|.$$

With this notation expression 2.2.3 becomes:

$$p_+(H) = \tfrac{1}{2} + \tfrac{1}{2}\tanh\,(\text{sumplus-summinus})$$

A value p = .997 which corresponds to an argument of the tanh of 3.00 is used as a lower limit for accepting a sign.

This leads to the following classification for the quality of the sign of a reflection H:

a) A reflection H either with sumplus $>$ 3 and summinus $=0$ ($S(H)=+$) or with sumplus $=0$ and summinus $>$ 3 ($S(H)=-$) is called very good.

b) A reflection H either with sumplus $>$ 4.5 and sumplus $>$ 3 × summinus ($S(H)=+$) or with summinus $>$ 4.5 and summinus $>$ 3 × sumplus ($S(H)=-$) is called good.

c) A reflection not falling in the groups mentioned in a) and b) is called bad.

Consistency-criteria for the solutions

For each solution the number of very good reflections, the number of good reflections and the number of bad reflections is counted.

Further in the group of very good reflections the absolute values of the arguments of tanh are summed.

In the group of good reflections two sums are made. The first sum consists of the absolute values of the arguments in favour of the signs allotted to the reflections:

$$A=\sum_{\substack{\text{H with}\\ S(H)=+}} \text{sumplus}+\sum_{\substack{\text{H with}\\ S(H)=-}} \text{summinus}$$

The second sum consists of the absolute values of the arguments against these signs:

$$B=\sum_{\substack{\text{H with}\\ S(H)=+}} \text{summinus}+\sum_{\substack{\text{H with}\\ S(H)=-}} \text{sumplus}$$

In the group of bad reflections each reflection is also awarded a sign. If sumplus$>$summinus then this sign is plus and minus otherwise. For the bad reflections the same sums are calculated as for the good reflections (A' resp. B').

So for each solution there are 8 consistency criteria:

1) The number of very good reflections.
2) The sum total of the arguments of these reflections.
3) The number of good reflections.
4) The sum total of the arguments in favour of the signs of the good reflections (A).
5) The sum total of the arguments against the signs of the good reflections (B).
6) The number of bad reflections.
7) The sum total of the arguments in favour of the signs of the bad reflections (A').
8) The sum total of the arguments against the signs of the bad reflections. (B').

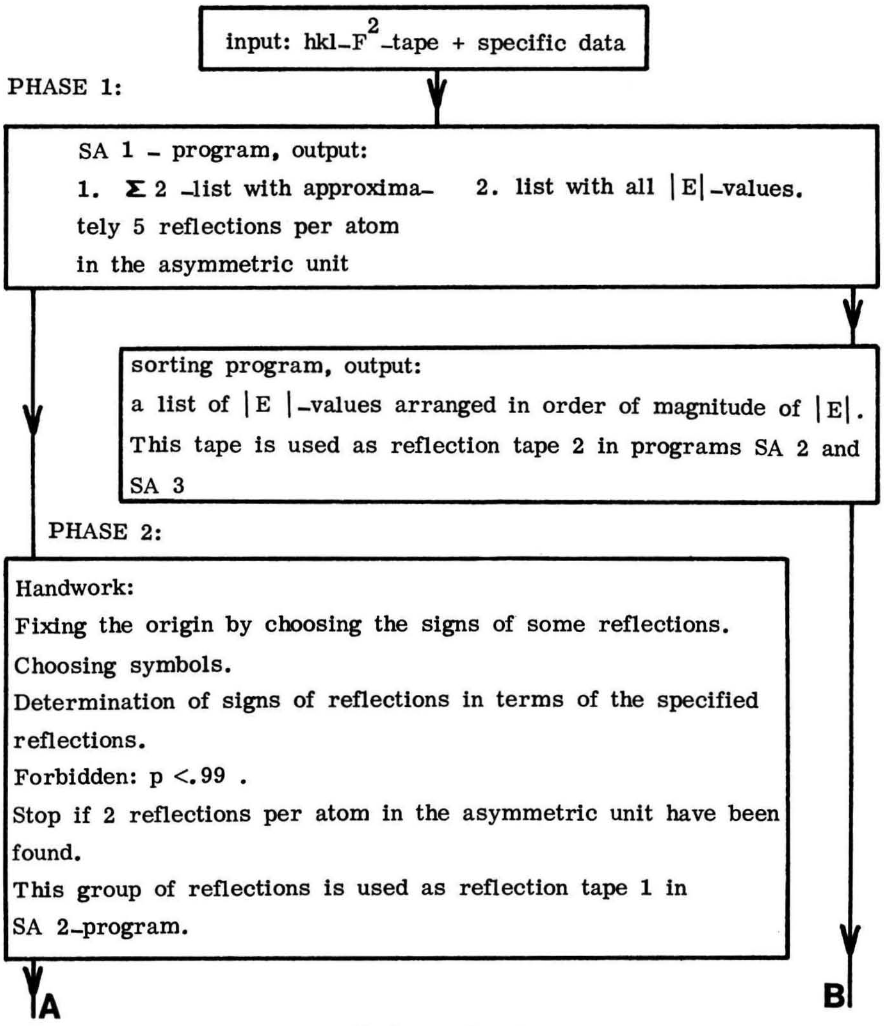
Plus-minus distribution criterion for the solutions

Another criterion is the +/− distribution of signs in the eight parity groups. Therefore in each parity group the sum $\Sigma S(H)|E_H|$ is taken over all reflections which are good or very good.
It is not permitted for some of these sums to blow up too much, as specified by the space group (for an example see 4.3.1).

Fourier-coefficient tapes

If required, Fourier-coefficient tapes which consist of the good and very good reflections can be produced.

Fig. 4.4.1 Flow chart of a structure determination



input: hkl-F^2-tape + specific data

PHASE 1:

SA 1 - program, output:
1. $\Sigma 2$-list with approximately 5 reflections per atom in the asymmetric unit
2. list with all $|E|$-values.

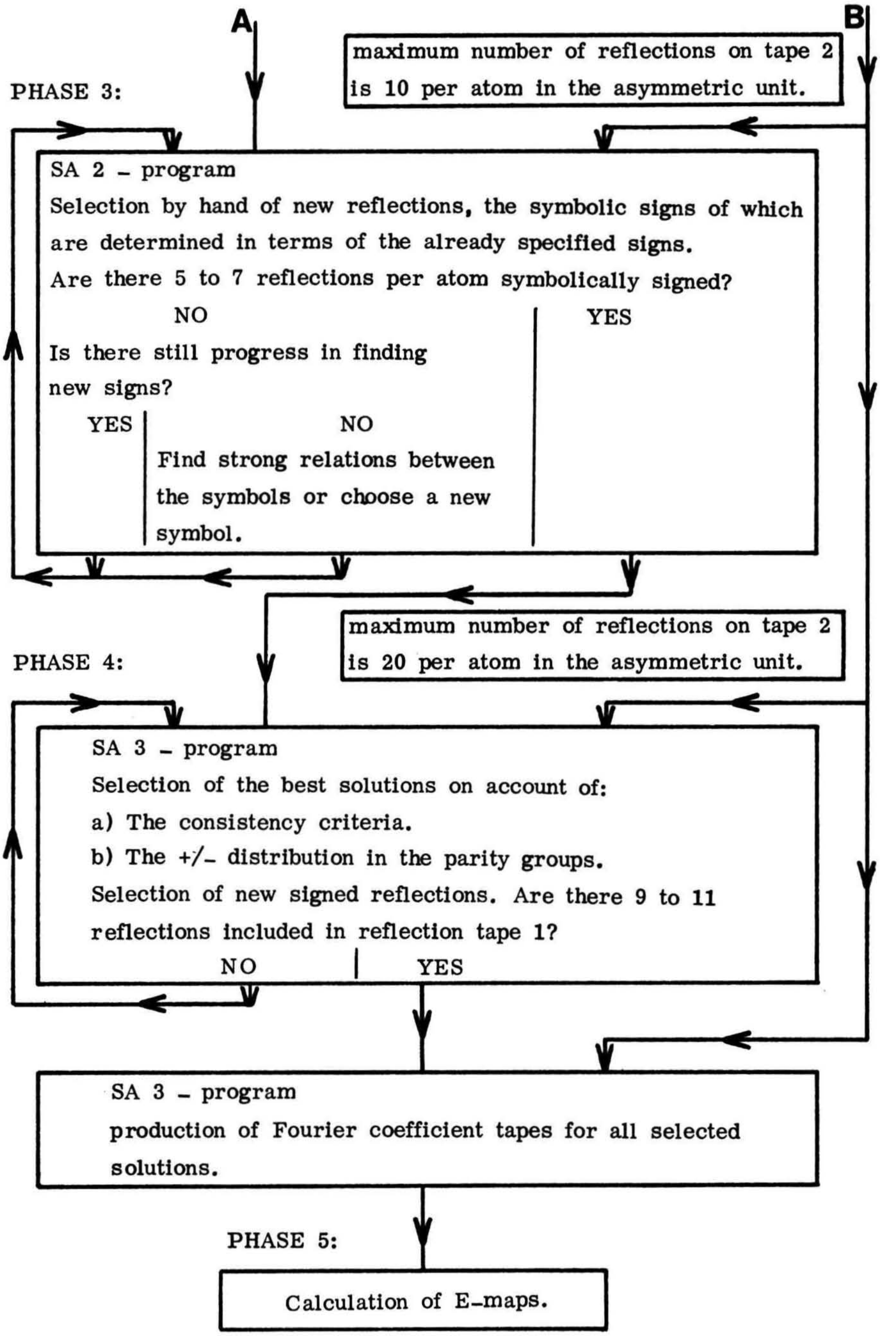
sorting program, output:
a list of $|E|$-values arranged in order of magnitude of $|E|$.
This tape is used as reflection tape 2 in programs SA 2 and SA 3

PHASE 2:

Handwork:
Fixing the origin by choosing the signs of some reflections.
Choosing symbols.
Determination of signs of reflections in terms of the specified reflections.
Forbidden: $p < .99$.
Stop if 2 reflections per atom in the asymmetric unit have been found.
This group of reflections is used as reflection tape 1 in SA 2-program.

(*To be continued*)

Fig. 4.4.1 Continued



4.3.3 How to work with SA 3

Input

It is possible to reduce the number of solutions by taking into account very strong relations between the symbols.

Output

In both the Σ_2-lists the qualifications good, very good or bad are printed out for each reflection H in each solution. The consistency criteria are given together with the sums of E_H in the eight parity groups. With the help of these criteria unreliable solutions can be eliminated.

Iterative procedure

This is carried out only with the selected most reliable solutions. The number of reflections on tape 1 is extended by including those reflections from tape 2, which are good or very good for all reliable solutions. Then a new cycle of SA 3 is started.

The limit value of $|E|$ must be lowered in successive steps, such that in the last cycle of SA 3 the number of reflections on tape 2 is 20 per atom in the asymmetric unit.

Fourier-coefficient tapes

In our experience Fourier-coefficient tapes can be computed if reflection tape 1 consists of 9 to 11 reflections per atom in the asymmetric unit.

4.4 *Summary: Flow chart of a structure determination*

The flow of a structure determination is summarized in fig. 4.4.1.

4.5 *Discussion*

The two computer programs SA 2 and SA 3 enable a fast sign determination. The time required for the determination of a structure of 20 atoms of nearly the same atomic number is about 1 hour of computing time with the X8 Electrologica computer and 30 hours of handwork.

Altogether 5 structure determinations have been carried out successfully with the help of these computer programs, two of which are presented in this report (chapters 6 and 7).

Some essential conditions for a successful sign determination are given in 5.1.1.

CHAPTER 5

AUTOMATIC SYMBOLIC ADDITION

5.1 *Introduction*

The flow chart of a structure determination given in 4.4 suggest the possibility of a full automation of the process. In order to achieve this, it is necessary:
a) to write a computer program for phase 2.
b) to make the programs SA 2 and SA 3 self iterative.
c) to couple the five programs.

It was possible to realize a) and b) so that the flow chart of a structure determination was simplified to that of fig. 5.1.1.

Item c) was not realized because the memory capacity of the computer was unsufficient.

The flow chart (fig. 5.1.1) shows that all essential work is done by the computer and that hand work is reduced to a minimum.
In this chapter the programs ASA 1, ASA 4, ASA 2 and ASA 3 are described and discussed. In 5.1.1 the essential conditions for a successful symbolic addition are given. The automatic procedure is compared with other automatic computer programs in 5.7.

A scheme is given for automatic programs in the noncentrosymmetric case (5.8).

5.1.1 Essential conditions for a successful sign determination

1) Throughout the symbolic addition procedure a high minimum value for the probability of a sign indication has to be maintained (for instance $p > .99$). As a result many symbols must be chosen (ASA 4).
2) Pseudo translations have to be found from the distribution of the reflections in the parity groups (ASA 4). If necessary, additional symbols have to be chosen in the underpopulated parity groups.
3) A reflection the symbolic sign of which is determined by two or more different combinations of symbols may be accepted with the symbol combination of highest probability if this probability is very high, for instance $p \geqslant .9995$ or $p \geqslant .9999$ (ASA 4, ASA 2).
4) Symbolic signs have to be used as long as possible. Two hand-rules for starting the analysis of the unknown symbols are:
    a) Most of the very strong reflections (two per atom in the asymmetric unit) must have a symbolic sign (ASA 4, ASA 2).
    b) In the larger group of strongest reflections (ten per atom in the asymmetric unit) at least half the number of reflections must have a symbolic sign (ASA 4, ASA 2).

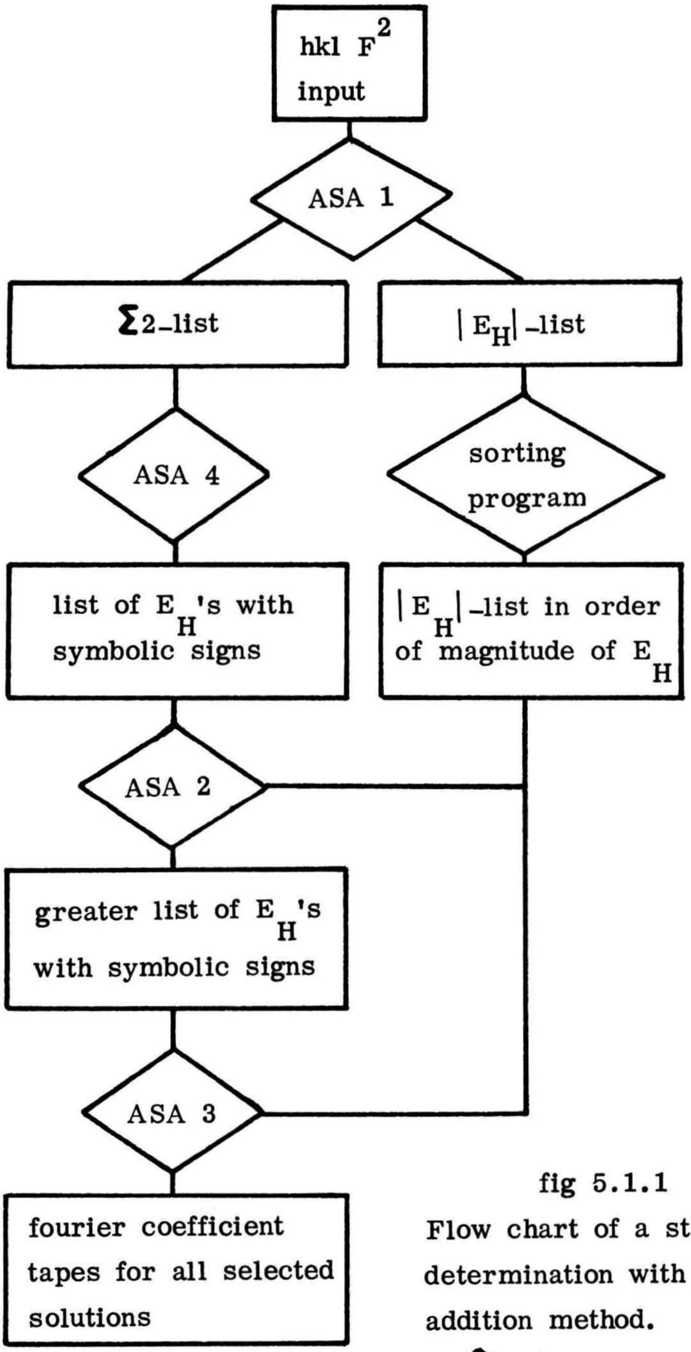


fig 5.1.1

Flow chart of a structure determination with the symbolic addition method.

◇ is a program

▭ is an input–output tape

5) Only very strong relations between symbols are accepted before the analysis of the remaining unknown symbols takes place on the basis of the +/− distribution of signs in the relevant parity groups and the consistency criteria. This is done in order to avoid calculating solutions with very bad consistency criteria. If weaker relations are accepted one runs the risk of introducing one or more wrong relations. Therefore the remaining unknown symbols have to be analysed simultaneously (ASA 3).
6) In our experience the best consistency criterion, calculated by SA 3 and ASA 3 is the sum total of the arguments against the signs of the bad reflections plus the sum total of the arguments against the signs of the good reflections. In all sign determinations thus far carried out the lowest value of this sum gave the correct solution.
7) In order to avoid solutions with high density on the centres of symmetry, relevant criteria for the +/− distribution in some of the parity groups as specified by the space group are essential (ASA 3).

5.1.2 Notation

H stands for h k l
H−H′ stands for h−h′ k−k′ l−l′
S(H) stands for the sign of reflection H.

5.2 *ASA 1*

The program ASA 1 is essentially identical to the program SA 1, written by Mr. H. v. d. Meer and described in 3.2.

5.3 *ASA 4*

The flow chart of this program is given in fig. 5.3.1. Some additional remarks about the various stages, indicated in the figure with Roman numerals, are given below:

ad I: The input data consist of an Σ_2 list in a special code with the symmetry relations incorporated. This list consists of approximately 5 reflections per atom in the asymmetric unit.

ad II: The priority group consists of the strongest reflections. The number must be approximately 1/2 reflection per atom in the asymmetric unit.
The strong E-group is a larger group of the strongest reflections; the number must be approximately 2 reflections per atom in the asymmetric unit.

ad IV: The searching procedure is organized such, that when a sign of a new reflection has been found, all the unsigned reflections which can be specified in terms of this sign are stored in a searching list. All these reflections are checked consecutively for their probabilities.

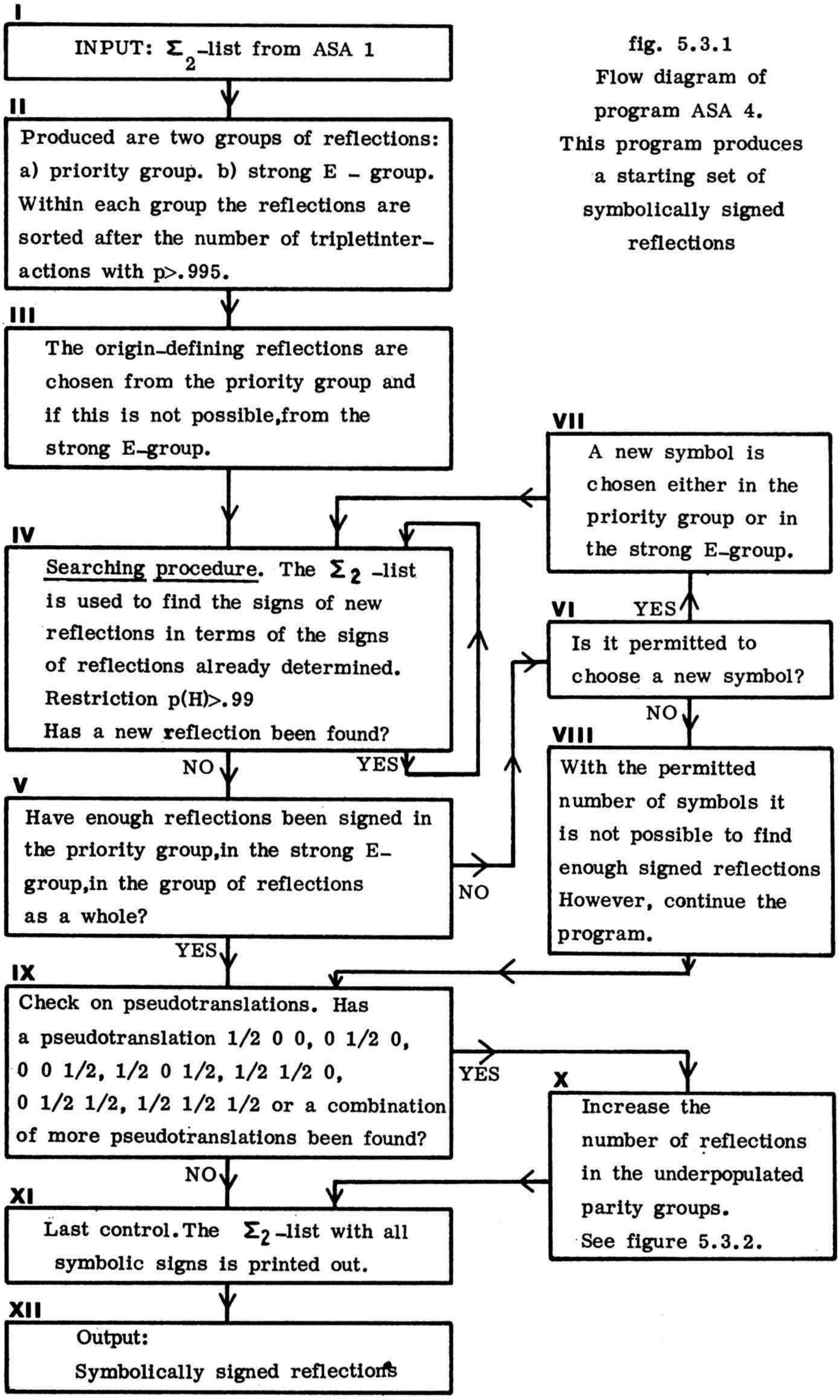


fig. 5.3.1
Flow diagram of
program ASA 4.
This program produces
a starting set of
symbolically signed
reflections

I INPUT: Σ_2-list from ASA 1

II Produced are two groups of reflections: a) priority group. b) strong E - group. Within each group the reflections are sorted after the number of tripletinteractions with p>.995.

III The origin-defining reflections are chosen from the priority group and if this is not possible, from the strong E-group.

IV Searching procedure. The Σ_2-list is used to find the signs of new reflections in terms of the signs of reflections already determined. Restriction p(H)>.99 Has a new reflection been found?

V Have enough reflections been signed in the priority group, in the strong E-group, in the group of reflections as a whole?

VI Is it permitted to choose a new symbol?

VII A new symbol is chosen either in the priority group or in the strong E-group.

VIII With the permitted number of symbols it is not possible to find enough signed reflections However, continue the program.

IX Check on pseudotranslations. Has a pseudotranslation 1/2 0 0, 0 1/2 0, 0 0 1/2, 1/2 0 1/2, 1/2 1/2 0, 0 1/2 1/2, 1/2 1/2 1/2 or a combination of more pseudotranslations been found?

X Increase the number of reflections in the underpopulated parity groups. See figure 5.3.2.

XI Last control. The Σ_2-list with all symbolic signs is printed out.

XII Output: Symbolically signed reflections

The probability of a symbolic sign from one triplet relation is calculated with

$$p(H) = p(H') \times p(H - H') \times p(\text{triplet})$$

where $p(\text{triplet}) = \frac{1}{2} + \frac{1}{2} \tanh \sigma_3 \sigma_2^{-3/2} |E_H| \, |E_{H'}| \, |E_{H-H'}|$. If the same symbol combination is found from two triplet relations the overall probability of this symbolic sign is:

$$p(H) = \frac{p_1(H)\, p_2(H)}{1 - p_1(H) - p_2(H) + 2p_1(H)p_2(H)}$$

(GERMAIN and WOOLFSON, 1968).

In order to exclude relations between the symbols in the starting set of signed reflections as far as possible, the following procedure is used:

a) The symbolic sign of a reflection H from the searching list is accepted if this symbol combination is the only one in the set of triplets belonging to this reflection and if its $p(H) > .99$.

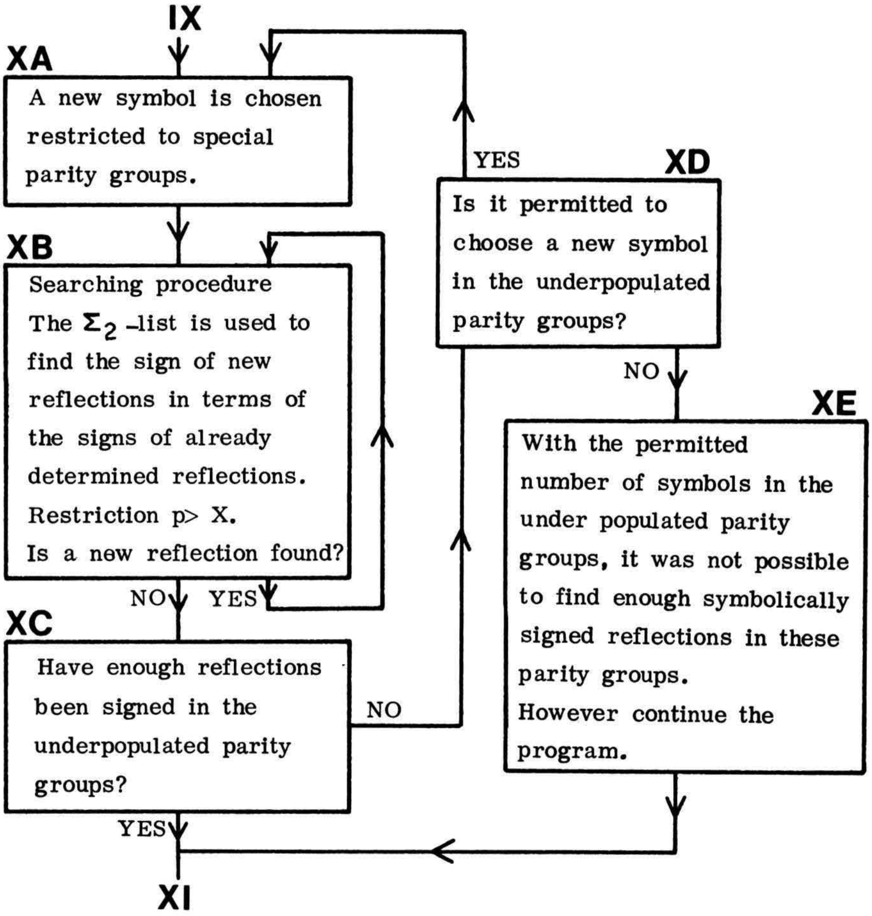


fig. 5.3.2 Flow diagram of part X of program ASA 4 (see fig. 5.3,1)

b) In the case of two different symbol combinations for H, only one of the combinations is accepted if the corresponding $p > .9995$.

c) When three or more different symbol combinations are found reflection H is not accepted.

ad V: The criteria used in this test are given in the specific data. We use 80 %, 70 % and 50 % respectively.

ad VI: The maximum permissible number of symbols is given in the specific data. A suitable number is $\frac{1}{2}$ symbol per atom in the asymmetric unit.

ad VII: A priority rule for choosing symbols can be given explicitly in the specific data.

ad VIII: In a new cycle of ASA 4 other criteria have to be used.

ad IX: Pseudo translations are found from the distribution of the signed reflections over the eight parity groups. The following possibilities of pseudo lattices and pseudo translations are checked: F, I, A, B, C, halving the x-axis, halving the y-axis, halving the z-axis, halving both x- and y-axis, halving both x- and z-axis, halving both y- and z-axis, C+ halving the z-axis, A+ halving the x-axis, B+ halving the y-axis, F+ halving the x-axis. A criterion for a pseudotranslation is the number of signed reflections in special parity groups in relation to the number of signed reflections in the remaining parity groups.
There is no test on pseudo translations of 1/300 etc. but this can easily be incorporated if necessary.

ad X A: A new symbol is chosen in one of the underpopulated parity groups.

ad X B: This searching procedure is identical to that of IV, with one exception. In order to increase the number of signed reflections in the underpopulated parity groups a less stringent probability criterion has to be used. In the case of *retro-β*-ionylidene acetic acid a suitable value has been found to be $x = .985$ (8.4.3).

ad X C: The minimum number is given in the specific data (A suitable value is about 3/4 of the number of atoms in the asymmetric unit).

ad X D: The maximum number is given in the specific data (A good value is approximately 1/4 of the number of atoms in the asymmetric unit).

ad X E: Perhaps a new cycle of ASA 4 is necessary with other criteria.

ad XII: The symbolically signed reflections can be used in ASA 2 (reflection tape 1).

5.4 *ASA 2*

The flow chart of program ASA 2 is given in fig. 5.4.1. This program is based on program SA 2, described in 4.2. Two modifications of SA 2 are incorporated in ASA 2:

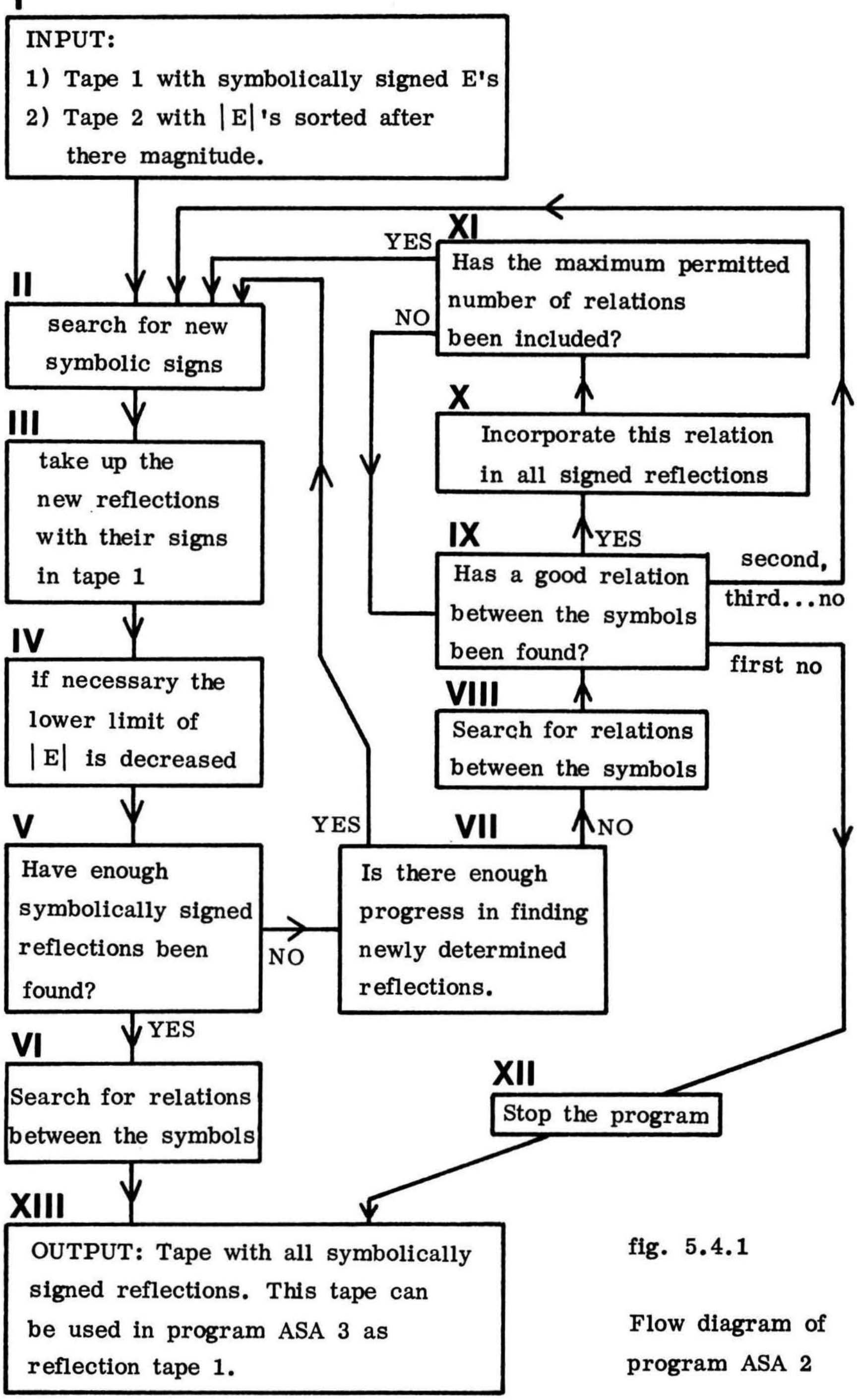


fig. 5.4.1

Flow diagram of program ASA 2

1) SA 2 with the possibility of selection of new symbolic signs. The minimum probabilities for a sign are given in the specific data of ASA 2. Recommended criteria are $p > .99$ for reflections with only one symbol combination and for reflections signed with two different symbol combinations $p > .9999$ for accepting one of these. This modification is used in II.
2) SA 2 with the possibility of searching for relations between the symbols. This modification is used in VI and VIII.

Some remarks on the various steps in program ASA 2:

ad I: The lower limit of $|E|$ used in the first run through II is given explicitly.

ad II: Here only the unsigned reflections above the lower limit of $|E|$ are considered.

ad IV: The lower limit of E must be decreased in successive steps, such that in step VI 10 to 12 $|E|$'s per atom in the asymmetric unit are considered.

ad V: A good criterion is 5 to 6 reflections per atom in the asymmetric unit.

ad VI: This searching procedure enables the inclusion of very strong relations in the specific data of ASA 3 (see 5.6.2.5).

ad VII: A criterion is 1/2 reflection per atom in the asymmetric unit.

ad IX: A criterion is: ten independent identical relations with individual probabilities of at least .97 and no relation against.

ad XI: Criterion: one or two relations.

ad XII: The program has been stopped because there is no progress and there are no relations with sufficiently high probabilities. The following possibilities are open to remove the dead lock:
1) Choose a new symbol.
2) Slacken the criteria.
3) If the number of symbolically signed reflections is not too far below the criterion of V, continue with ASA 3.

5.5 *ASA 3*

Program ASA 3 is based on program SA 3, described in 4.3. The flow chart of ASA 3 is given in fig. 5.5.1.

Some remarks about the various steps of ASA 3:

ad II: In this step the procedure of SA 3 is carried out. Reflections which are good or very good for many solutions are stored with their signs.

ad III: A criterion for this elimination may vary for different spacegroups (see 4.3.1).

ad IV: The selection in this step is based on the sum of two consistency criteria (see 4.3.2):
a) The total argument against the signs of the good reflections.

b) The total argument against the signs of the bad reflections. The solutions with the lowest sum of a and b are selected on the ground of criteria given in the specific data.

ad V: From the stored reflections (see ad II) those reflections are selected, which are good or very good for all selected solutions (see 4.3.3). These reflections are added to tape 1.

ad VI: A suitable criterion is nine to eleven reflections per atom in the asymmetric unit.

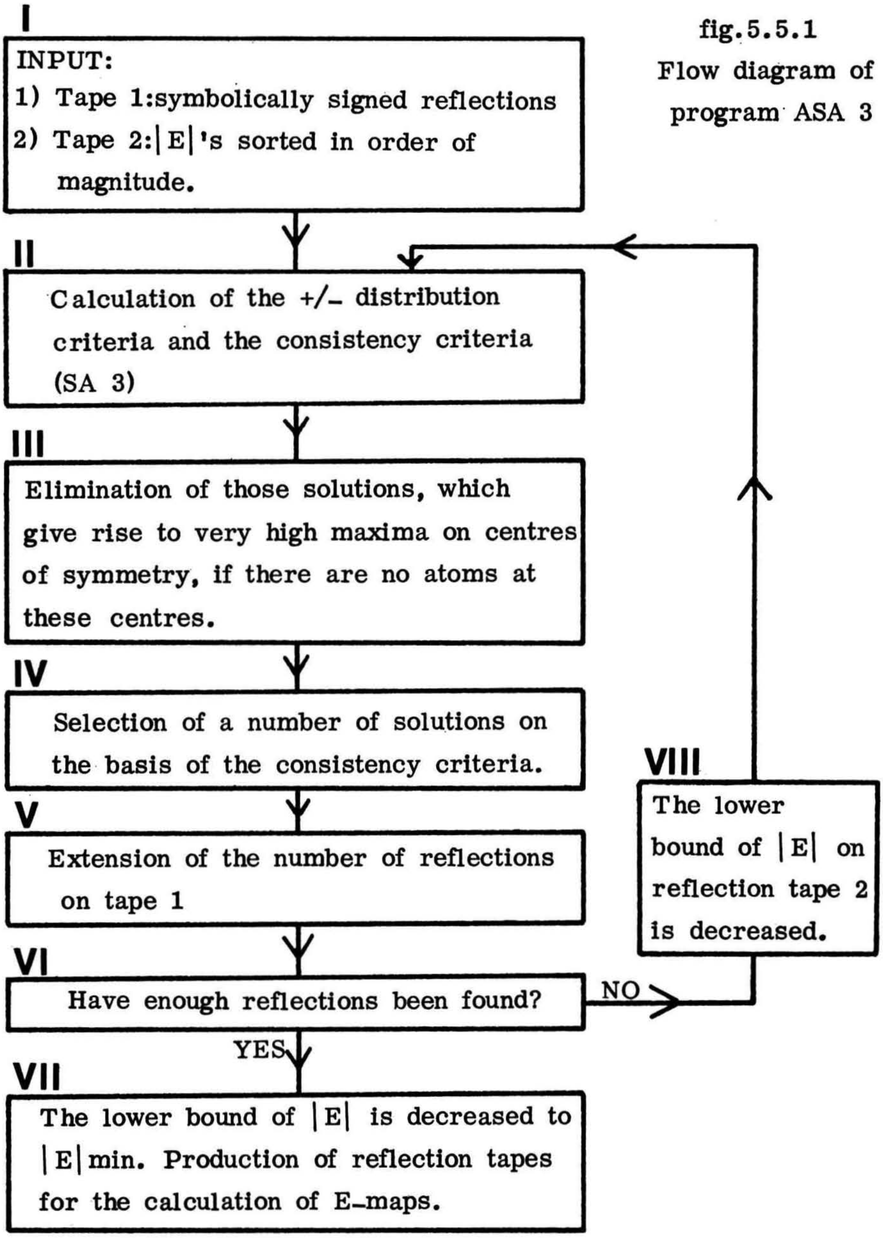


fig.5.5.1
Flow diagram of program ASA 3

ad VIII: The lower limit of $|E|$ is decreased gradually to $|E|_{min}$. $|E|_{min}$ is such that tape 2 contains approximately 20 reflections per atom in the asymmetric unit.

5.6 *Results*

The structure determination of retro-β-ionylidene acetic acid has been carried out with the automatic programs. The best solution given by program ASA 3 proved to be the correct one. Although a pseudo I centered unit cell was present all 299 signed reflections proved to be correct (see chapter 8).

The computing time was one hour on the X8 Electrologica computer. Preparation of various specific data tapes took about two hours.

5.7 *Comparison of some other automatic computer procedures for the symbolic addition with the one presented in this report.*

5.7.1 The method of Beurskens (1965).

Beurskens introduces symbolic signs for ten to fifteen very strong E's. With these signs he starts a sign generation, which is interrupted now and then to find relations between the symbols. If a relation has been found, one of the symbols is eliminated. This process is continued until a sufficient number of reflections has been signed. At the best all symbols have been eliminated at that moment. If this is not the case, as many of the remaining symbols as possible must be found by means of the $+/-$ distribution of the signs, and Σ_1- and Σ_3-relations.

If in the end n symbols have not been determined 2^n Fourier-series have to be calculated.

The differences with the procedure, presented in this report are the following:

a) In the first stage of our procedure we choose as few symbols as possible maintaining a high minimum probability p for the acceptance of a sign. Moreover the possibility of pseudo translations is taken into account.
   Beurskens introduces a fixed number of symbols indiscriminately.
b) In the last stage of Beurskens's procedure symbols may be eliminated by means of rather weak relations. In doing so one runs the risk of introducing a wrong relation and missing the correct solution.
   We, however, use a multi-solution procedure, in which all unknown symbols are considered simultaneously. With the computed consistency criteria a classification of the various solutions is made, so that the correct solution can be found, even if it would turn out not to be the "best" one.

5.7.2 The method of Germain and Woolfson (1968)

The input limitations of the program of Germain and Woolfson are the following:

1) Maximum number of E's: 250.
2) Maximum number of triplet interactions: 1000.
3) Maximum number of symbols: 6.

ad 1: In a Fourier-synthesis one requires about ten reflections per atom in the asymmetric unit to get good images of all atoms, so that sometimes 250 signed E's are insufficient.

ad 2: In the stage at which the symbols are analysed it is important to have as many triplet interactions as possible. Limiting the number of triplet interactions to 1000 may in certain cases be a serious handicap.

ad 3: We have found it necessary to choose more than six symbols in most of our structure determinations to avoid the necessity of using probabilities lower than .99.

Therefore our programs do not have limitations of this kind.

In the program of Germain and Woolfson the procedure of solving the unknown symbols is essentially based on searching for relations between the symbols. The probability of such a relation is measured by the frequency of occurrence. They also include the possibility that one of the accepted relations may fail and this results in six more possible solutions.

In our pre-selection of the possible solutions we use probabilities instead of frequencies of occurrence, and only very strong relations are accepted. All remaining unknown symbols are considered simultaneously.

The last stage of the program of Germain and Woolfson generates the signs of all reflections for the seven solutions and three figures-of-merit are calculated, namely the number of sign relationships which hold, the sum of the probabilities of the sign relationships which hold and $\Sigma\, S_1\, S_2\, S_3\, E_1\, E_2\, E_3$.

The last figure-of-merit is equivalent to the sum of the five consistency criteria 2, 4, 5, 7 and 8 used in our program ASA 3.

Germain and Woolfson do not use criteria for the absence of density on the centres of symmetry. This may lead to false solutions.

Our program has the possibility of refining a selected number of solutions.

5.8 *Automatic program for the non-centrosymmetric case*

5.8.1 Probability considerations

In the non-centrosymmetric case KARLE and KARLE (1966) give the expression for a phase angle of reflection H:

$$\varphi_H \approx \frac{\sum\limits_{H'} |E_{H'}\, E_{H-H'}| (\varphi_{H'} + \varphi_{H-H'})}{\sum\limits_{H'} |E_{H'}\, E_{H-H'}|} \tag{5.8.1.1}$$

with the variance given by

$$V = \frac{\pi^2}{3} + [I_0(\alpha)]^{-1} \sum_{h=1}^{\infty} \frac{I_{2n}(\alpha)}{n^2} + 4[I_0(\alpha)]^{-1} \sum_{n=1}^{\infty} \frac{I_{2n+1}(\alpha)}{(2n+1)^2} \tag{5.8.1.2}$$

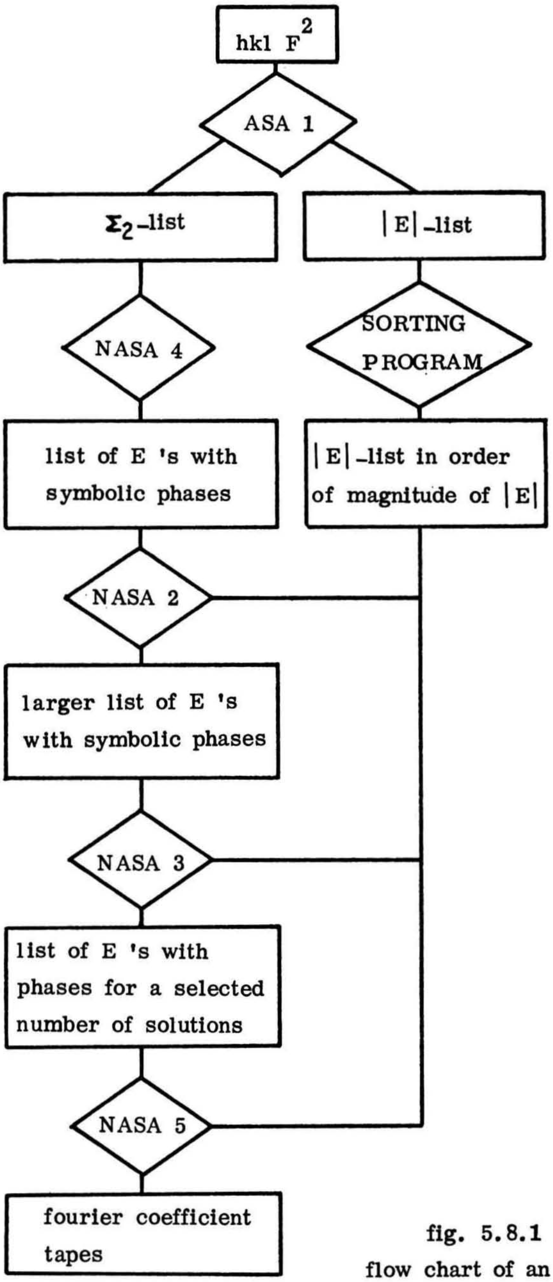


fig. 5.8.1
flow chart of an automatic computer program for the symbolic addition in the non-centrosymmetric case.
◇ is a program, □ is an input-output tape.

where $I_1(\alpha)$ are Bessel functions and

$$\alpha = 2\sigma_3\sigma_2^{-3/2}|E_H|\{[\sum_{H'}|E_{H'}E_{H-H'}|\sin(\varphi_{H'}+\varphi_{H-H'})]^2 + \\ + [\sum_{H'}|E_{H'}E_{H-H'}|\cos(\varphi_{H'}+\varphi_{H-H'})]^2\}^{1/2} \qquad (5.8.1.3)$$

Karle and Karle also give a curve showing the variance V as a function of α and a working rule of $V \leqslant .5$ for acceptance of a phase indication. Thus in a computer program we can use a maximum value of V or minimum value of α for accepting a phase angle.

5.8.2 Scheme for a computer program

In fig. 5.8.1 the flow diagram for an automatic computer program is shown.

The programs NASA 4 and NASA 2 can be essentially equal to the programs ASA 4 and ASA 2. Only small changes have to be introduced. Consistency criteria for program NASA 3, a modification of program ASA 3, can be developed in the following way:

If there is maximum agreement between the different contributors to α, then expression 5.8.1.3. reduces to:

$$\alpha_{\max} = 2\sigma_3\sigma_2^{-3/2}|E_H|\sum_{H'}|E_{H'}E_{H-H'}| \qquad (5.8.2.1)$$

Normally such an agreement does not exist and α is given by 5.8.1.2. With $P=(\alpha_{\max}+\alpha)/2$ and $Q=(\alpha_{\max}-\alpha)/2$ consistency criteria can be developed in analogy to the criteria of 4.3.2. P and Q play the role of sumplus and summinus respectively if sumplus > summinus and the roles of P and Q are interchanged if sumplus < summinus.

In the centrosymmetric case 1/2P and 1/2Q are in fact identical to sumplus and summinus respectively if sumplus > summinus; the reverse applies to the other case.

Program NASA 5 is a phase-refining program, based on the tangent-formula:

$$\operatorname{tg}\varphi_H = \frac{\sum_{H'} E_{H'}E_{H-H'}\sin(\varphi_{H'}+\varphi_{H-H'})}{\sum_{H'} E_{H'}E_{H-H'}\cos(\varphi_{H'}+\varphi_{H-H'})} \qquad (5.8.2.2)$$

(Karle and Karle, 1966).

Maybe it is possible to build NASA 5 into NASA 3.

CHAPTER 6

THE CRYSTAL STRUCTURE OF 1,1,4-TRI-METHYL-5-CARBOXYL-CYCLOPENTENE-4

6.1 *Introduction*

The formula of 1,1,4-tri-methyl-5-carboxyl-cyclopentene-4, also referred to as C9-acid, is $C_9H_{14}O_2$. The structural formula and the numbering of the atoms are given in fig. 6.6.1.

C9-acid is the first intermediate stage in a many-step synthesis of the five-membered-ring analogue of vitamin-A. (HUISMAN and BAAS, 1968).

6.2 *Crystal data*

The crystals of the C9-acid are monoclinic. Systematically missing are the h0l-reflections with $h+l=2n+1$ and the 0k0-reflections with $k=2n+1$. The space group is $P2_1/n$.

The unit-cell dimensions have been measured from Weissenberg photographs about [010] and $[10\bar{1}]$.

The photographs were made with CuKα-radiation and calibrated with Al-powder lines.

$a = 11.8151$ (9) Å
$b = 6.0363$ (8) Å
$c = 12.5605$ (16) Å
$\beta = 100.877$ (12) Å

The density, measured by flotation, is 1.17 g.cm^{-3}. The calculated density with $Z=4$ is 1.16 g.cm^{-3}.

6.3 *Intensity data*

6.3.1 Measurements

The intensities have been collected with the Nonius Automatic Three-Circle Diffractometer.

Data:

radiation: CuKα
scan: θ–2θ (moving crystal, moving counter)
scan angle: 2.0°
maximum θ-value: 69°
reference reflections: 400, $\bar{3}03$, $\bar{1}05$, 404, 1000, 808, $\bar{2}012$, 505
maximum peak height: 5000 counts per second (stronger reflections were reduced by attenuation filters)
mounting axis of the crystal: [010]
cross-section of the crystal: 0.04 × 0.25 mm
length of the crystal: 0.40 mm

6.3.2 Reduction

A reflection has been considered as significantly above zero, if net counts >2.5 times the square root of (total plus background).

It appeared that after completion of the measurements the intensities of the reference reflections had dropped by some 20 %. With the help of these reflections the other intensities were corrected for this effect.

Theoretically 1641 reflections were measurable within the limitations of the diffractometer and of these 1128 were found as significantly above zero.

The $(LP)^{-1}$ correction was applied.

The absorption correction was ignored.

6.4 *Structure determination*

The structure was determined by the symbolic addition method for centrosymmetric crystals (KARLE and KARLE, 1966).

The computer programs, described in chapter 4, were used.

6.4.1 Statistical data

In order to get normalized structure factor amplitudes $|E_h|$ the expression

$$|E_h|^2 = \frac{K^2 \exp\left(2B \frac{\sin^2\theta}{\lambda^2}\right) |F_h|^2}{\varepsilon \Sigma f_j{}^2 \frac{(\sin^2\theta)}{\lambda}}$$

was used in the program, described in 3.2. Scale factor K and temperature factor B ($=4.4$ Å^2) were estimated from a Wilson-plot. As the space group is $P2_1/n$, $\varepsilon=2$ for the h0l-reflections and the 0k0 reflections and $\varepsilon=1$ for all other reflections.

The computed values for the distribution and statistics of $|E|$ were as follows:

$\langle|E|\rangle = .746$ (.798)
$\langle|E|^2\rangle = .995$ (1.000)
$\langle|E^2-1|\rangle = 1.065$ (.968)
reflections with $|E|>3.0$: .4 % (.3 %)
reflections with $|E|>2.5$: 1.8 % (1.2 %)
reflections with $|E|>2.0$: 5.2 % (5.0 %)
reflections with $|E|>1.4$: 17.0 % (16.2 %)
reflections with $|E|>1.0$: 32.0 % (31.7 %)

In brackets the theoretical values are given as calculated by KARLE, HAUPTMAN, KARLE and WING, (1958).

6.4.2 Starting set of symbolically signed reflections

The 86 reflections with $|E|>2.0$ were separated from the rest of the data and with this group a Σ_2-list was computed.

The reflections $-4+1+13$ ($|E|=3.32$), $-9+5+7$ (3.42) and $-6+3+6$ (3.14) were used for defining the origin.

It was necessary to define the signs of six strong reflections in terms of a symbol.

In this way it was possible to get a starting set of 32 reflections with symbolic sign. Any of these reflections is signed with a probability $p > .993$.

6.4.3 Increasing the set of symbolically signed reflections

With the help of computer program SA 2 (see 4.2) the starting set was increased. After two cycles two very strong relations between the symbols were found. Each of them appeared more than 10 times in the Σ_2-list with probabilities $> .99$. In the next cycle these relations were accepted.

In the last cycle all reflections with $|E| > 1.6$ were considered. At the end of this cycle 92 reflections had obtained a symbolic sign. As no further progress could be made with the SA 2 program, it was decided to start eliminating symbols.

6.4.4 Finding the best solutions

The last step was the elimination of the 4 remaining symbols. The corresponding $2^4 = 16$ solutions, given in table 6.4.1, were analysed with program SA 3. (see 4.3).

TABLE 6.4.1
Signs of the unknown symbols for the 16 solutions. Included are the relations $R = -$ and $A = B$.

Number of the solution	Sign of symbol					
	A	B	S	R	X	Y
1	+	+	+	−	+	+
2	+	+	−	−	+	+
3	+	+	+	−	−	+
4	+	+	−	−	−	+
5	+	+	+	−	+	−
6	+	+	−	−	+	−
7	+	+	+	−	−	−
8	+	+	−	−	−	−
9	−	−	+	−	+	+
10	−	−	−	−	+	+
11	−	−	+	−	−	+
12	−	−	−	−	−	+
13	−	−	+	−	+	−
14	−	−	−	−	+	−
15	−	−	+	−	−	−
16	−	−	−	−	−	−

The results for all reflections above $|E| = 1.6$ with the 92 signed reflections as basis are given in table 6.4.2 for each of the 16 sign combinations. On the basis of the sums of E_{hkl} in the parity subgroups eee, eoo, oeo and ooe it was not possible to eliminate any of the solutions.

TABLE 6.4.2
First analysis of the consistency of the 16 solutions. The meaning of the consistency criteria is given in 4.3.2.

Number of the solution	The number of very good reflections	The sum total of their arguments	The number of good reflections	A	B	The number of bad reflections	A′	B′
1	82	1922.2	56	1189.0	202.9	66	664.7	324.7
2	78	1832.8	54	1087.8	197.8	72	783.3	401.9
3	88	1911.9	45	967.7	181.2	71	831.9	410.9
4	111	2585.9	45	984.8	138.8	48	401.5	192.5
5	102	2269.9	34	658.3	113.9	68	840.1	421.4
6	108	2483.6	50	1104.0	136.9	46	406.2	172.7
7	79	1820.6	57	1215.1	173.2	68	727.4	367.2
8	90	2064.0	45	865.2	131.0	69	827.6	415.7
9	90	1991.8	43	882.6	133.9	71	872.4	422.9
10	136	3182.0	35	737.2	97.2	33	223.2	63.9
11	98	2248.9	48	1055.7	159.1	58	567.4	272.4
12	73	1717.2	54	1068.1	156.4	77	881.2	480.6
13	77	1785.4	53	1144.8	190.5	74	793.2	389.7
14	82	1947.9	48	941.8	168.3	74	810.3	435.3
15	95	2103.8	35	722.6	128.3	74	865.8	483.0
16	106	2393.3	47	1042.9	148.0	51	497.2	222.1

By means of the consistency criteria discussed in 4.3.2 five solutions were selected from table 6.4.2 namely: 4, 6, 10, 11 and 16. The number of signed reflections for these solutions was extended to 117.

A new cycle of computations with SA 3 gave the final results for all reflections with $|E|>1.5$ as shown in table 6.4.3. The total argument

TABLE 6.4.3.
Second consistency analysis of the five best solutions from table 6.4.2.

Number of the solution	The number of very good reflections	The sum total of their arguments	The number of good reflections	A	B	The number of bad reflections	A′	B′
4	102	3020.2	89	2545.0	317.3	56	847.5	443.4
6	100	2925.1	99	2874.9	350.5	48	653.8	369.1
10	139	4072.1	79	2454.9	253.4	29	282.3	110.7
11	91	2721.8	101	2945.2	386.6	55	739.5	380.4
16	96	2834.6	99	2809.1	359.8	52	763.6	405.6

against solution number 10 is 364 and this is less than half the arguments against any of the other four solutions.

This solution gave the signs of 183 E-values and these were used as coefficients in a Fourier-summation. All these signs were later found to be correct.

The fourier-summation (E-map) showed 11 well defined peaks, from which the structure was found without any ambiguity (see Fig. 6.4.1).

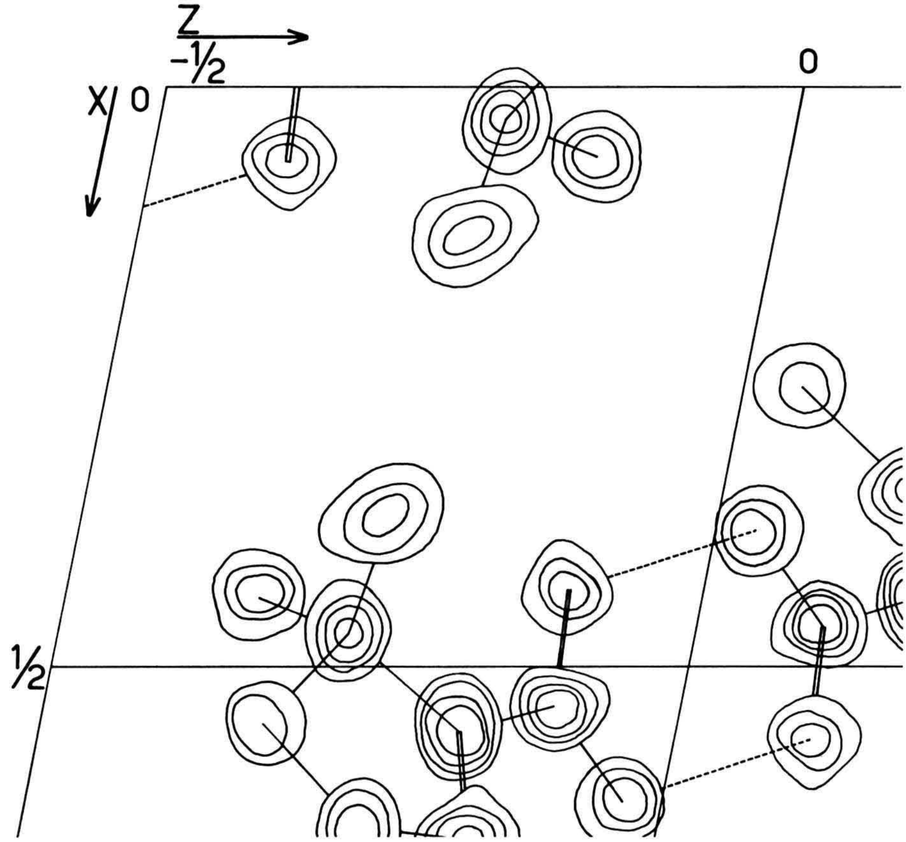


Fig. 6.4.1. Composite E-map of C9-acid viewed along the y-axis.

The structure factor calculation for this model, with the scaling factor K and the isotropic over all temperature factor B as estimated from the Wilson plot, gave an R-value of 23 %.

6.5 *Refinement*

The structure was refined by successive cycles of block diagonal least squares, using only the observed reflections. The weighting scheme of Cruickshank (1961) was used.

The fudge factors were .8, as ordinarily used. The atomic scattering factors were those from The International Tables for X-ray Crystallography.

After four cycles of individual isotropic refinement the R-index was reduced to 15.7 %.

In the next four cycles individual anisotropic temperature parameters were used. The R-factor dropped to 10.5 %. Because there were no shifts of importance anymore, a difference Fourier synthesis was computed, with the reflections with $\sin \theta/\lambda < .45$ as coefficients in order to locate the H-atoms. All hydrogen-atoms, the H-atom of the COOH-group included, were found. Their peak heights varied between .3 and .7 $e \cdot Å^{-3}$.

Two additional least squares cycles with fixed H-atoms and three cycles with isotropic refinement of the H-atoms followed, in which the R-factor was reduced to 6.0 %. Two final cycles were carried out omitting the reflections 200, $\bar{1}01$, $\bar{2}02$, $\bar{3}11$, $\bar{2}11$, $\bar{2}12$, and $\bar{1}12$, which presumably had appreciable extinction errors or measuring errors as a result of the dead time effect of our counting unit. The final R-value is 5.11 %.

In the last cycle the shifts in the parameters were all less than their estimated standard deviations.

A final difference Fourier synthesis, which was calculated with the observed reflections only, showed a number of peaks which would be significant in terms of the standard deviation for the electron density (.03 $e. Å^{-3}$), as calculated with the formula of CRUICKSHANK (1949).

Some of these peaks are lying on the bonds. Three other peaks can be interpreted as alternative positions for the hydrogen atoms of C(7), indicating disorder in the conformation of methyl group C(7). The alternative positions and the main positions are related to each other by a rotation of approximately 60° about the C(4)–C(7) bond.

The occupancy rate of the alternative positions is found to be 15 %. This result has not been used for further refinements.

6.6 *The crystal structure in numbers and figures*

In table 6.6.1 the final positional and thermal parameters of the heavy atoms are listed together with their e.s.d.'s. The e.s.d.'s are those given by the method of block diagonal least squares.

The positional and thermal parameters for the hydrogen atoms are given in table 6.2.2.

In fig. 6.6.1 the bond lengths are given with their e.s.d.'s and fig. 6.6.2 shows the bond angles, calculated from the coordinates of tables 6.6.1 and 6.6.2. Some other important intramolecular distances are given in fig. 6.6.3.

Newman projections along all C–C-bonds are shown in fig. 6.6.4. Various planes through the atoms of the conjugated double bond system were calculated by the method of SCHOMAKER, WASER, MARSH and BERGMAN (1959). The equations of these planes and the distances of the atoms from these planes are given in table 6.6.3.

TABLE 6.6.1.

Final positions of the heavy atoms in fractional coordinates (x, y, z) and the anisotropic thermal parameters (β_{ij}) from the expression

$$\exp\{-(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}l^2+2\beta_{12}hk+2\beta_{23}kl+2\beta_{31}lh)\}$$

The e.s.d. are given in parenthesis.

Atom	x	y	z	$\beta_{11}.10^5$	$\beta_{22}.10^5$	$\beta_{33}.10^5$	$2\beta_{12}.10^5$	$2\beta_{23}.10^5$	$2\beta_{31}.10^5$
C (1)	.52509 (18)	.53249 (39)	.27368 (17)	685 (16)	2275 (65)	433 (13)	−42 (57)	292 (49)	111 (23)
C (2)	.44856 (23)	.38110 (47)	.32925 (20)	970 (22)	3046 (83)	534 (16)	−196 (71)	724 (61)	287 (30)
C (3)	.35898 (20)	.28151 (42)	.23865 (20)	860 (20)	2450 (71)	665 (17)	−516 (65)	524 (62)	340 (30)
C (4)	.35303 (18)	.44424 (39)	.14689 (17)	635 (16)	2307 (64)	495 (14)	111 (55)	133 (50)	288 (24)
C (5)	.44210 (17)	.58592 (36)	.16745 (16)	575 (14)	2048 (59)	423 (12)	80 (52)	218 (47)	223 (22)
C (6)	.46641 (17)	.76168 (37)	.09434 (17)	591 (14)	2134 (62)	461 (12)	103 (53)	369 (50)	218 (22)
C (7)	.25789 (20)	.42285 (50)	.05057 (20)	729 (18)	3475 (89)	632 (18)	−318 (71)	16 (67)	−7 (29)
C (8)	.63014 (22)	.40794 (52)	.25045 (22)	787 (20)	3957 (98)	714 (19)	1091 (78)	782 (77)	203 (32)
C (9)	.56333 (25)	.73726 (51)	.34334 (21)	1204 (26)	3135 (87)	555 (17)	−772 (82)	−184 (67)	36 (33)
O(10)	.56490 (14)	.84022 (33)	.10204 (15)	692 (12)	3786 (66)	800 (14)	−582 (50)	1559 (51)	170 (22)
O(11)	.38015 (14)	.83661 (35)	.02353 (15)	765 (13)	3809 (69)	746 (13)	−214 (51)	1708 (51)	−75 (21)

TABLE 6.6.2

Final positional (x, y, z in fractions) and thermal (B in Å²) parameters of the hydrogen atoms with their e.s.d.'s in brackets.

	x	y	z	B
H (1)	.4096 (28)	.9459 (69)	.9716 (30)	6.0 (1.0)
H (2)	.2769 (23)	.4961 (52)	.9839 (24)	3.2 (.6)
H (3)	.2459 (28)	.2576 (68)	.0287 (30)	5.4 (.9)
H (4)	.1873 (28)	.4730 (64)	.0655 (28)	4.7 (.8)
H (5)	.3863 (22)	.1255 (48)	.2084 (22)	2.4 (.6)
H (6)	.2869 (18)	.2641 (43)	.2583 (18)	1.3 (.5)
H (7)	.4098 (24)	.4701 (55)	.3814 (25)	3.8 (.7)
H (8)	.4993 (23)	.2550 (54)	.3780 (24)	3.2 (.6)
H (9)	.5001 (25)	.8217 (55)	.3568 (26)	3.9 (.7)
H(10)	.6064 (31)	.6924 (76)	.4173 (32)	6.9 (1.1)
H(11)	.6078 (25)	.8334 (55)	.3042 (25)	3.6 (.7)
H(12)	.6828 (26)	.3667 (60)	.3124 (28)	4.4 (.8)
H(13)	.6728 (22)	.4927 (50)	.2045 (23)	2.8 (.6)
H(14)	.6015 (27)	.2477 (64)	.2070 (27)	5.1 (.8)

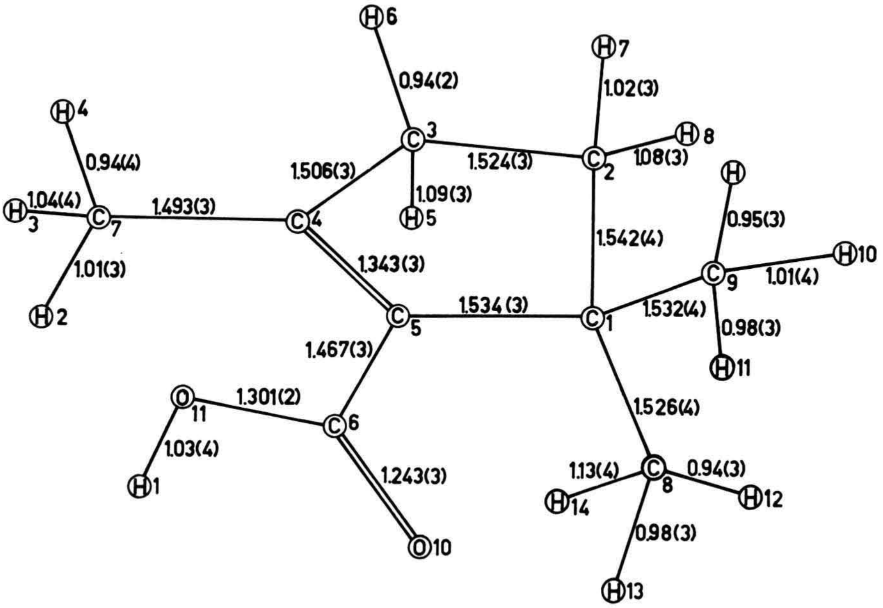


Fig. 6.6.1. Bond lengths in C9-acid.

In figures 6.6.5a and 6.6.5b the intermolecular distances smaller than 4.00 Å are shown.

6.7 *Discussion of the structure*

6.7.1 Conformation of the conjugated system

The conformation of the conjugated system including the double bond in the ring and the C=O bond is all-*trans*.

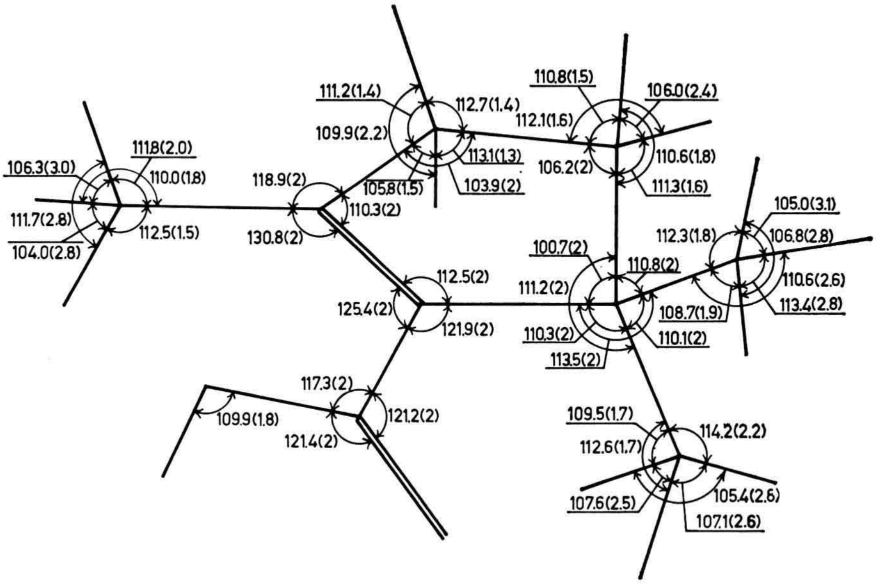


Fig. 6.6.2. Bond angles in C9-acid.

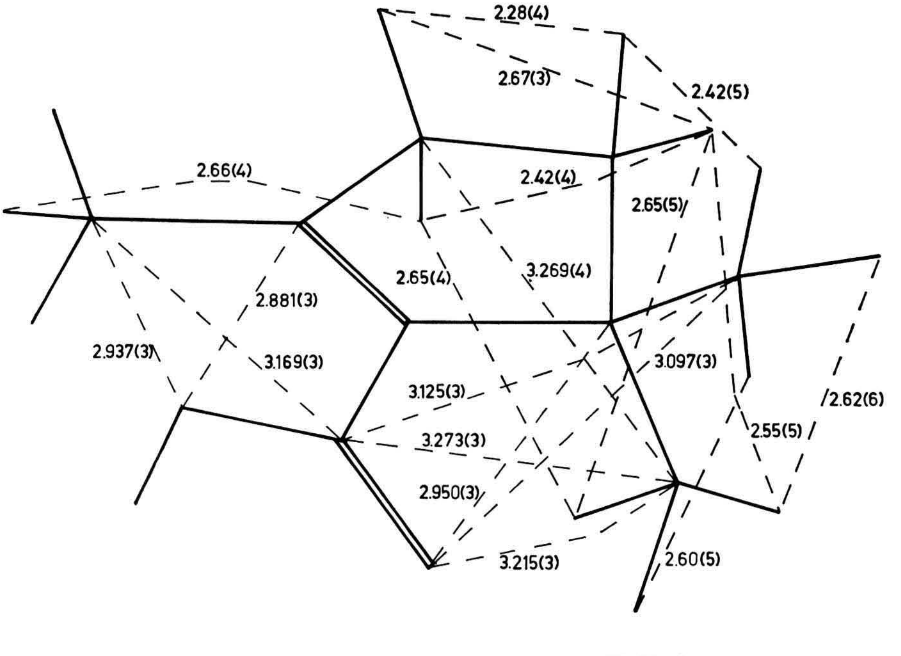


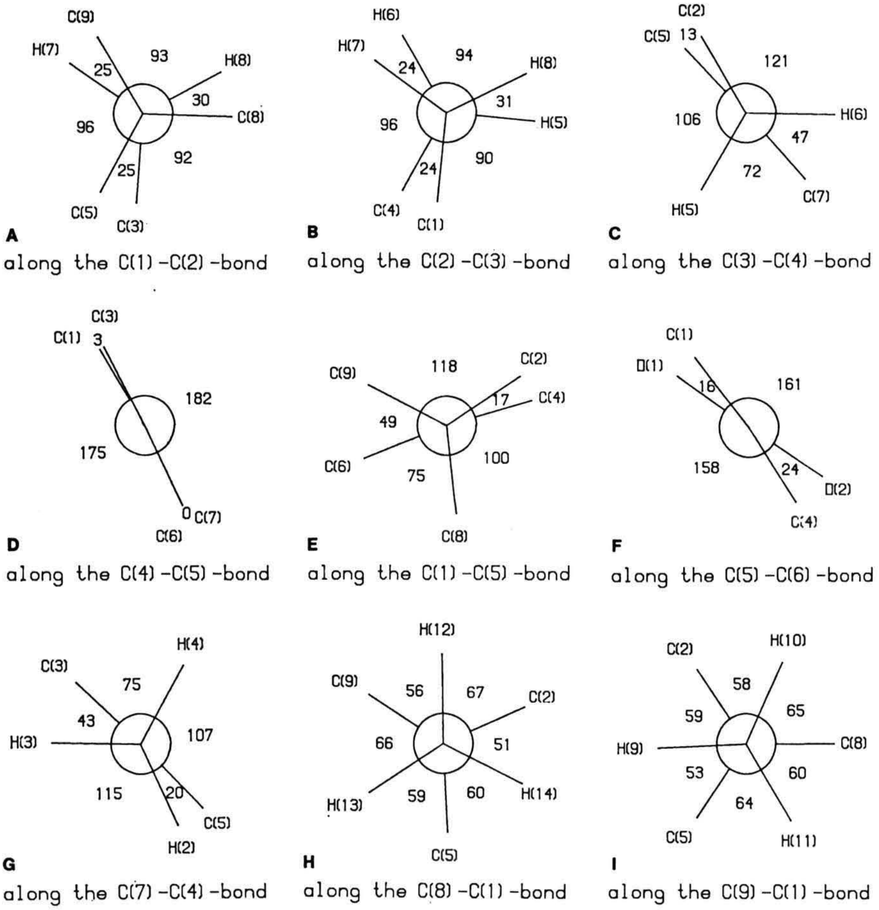
Fig. 6.6.3. Short intramolecular C–C and H–H distances.

6.7.2. Bondlengths of the heavy atoms (fig. 6.6.1).

The bond distances in the ring are all normal in comparison with the mean values in other structures in the vitamin-A series (Bart and MacGillavry, 1968b). There is one exception namely the C(2)–C(3)-bond. In nearly all structures this bond is fairly short (1.483 Å). As is pointed out by Sly (1964) a correction for thermal motion might lengthen the C(2)–C(3)-bond. In the C9-acid a normal distance of 1.524 Å is found. The thermal parameters of C(2) and C(3) are not excessively different from those of the other C and O atoms. Also the B-values of the hydrogen atoms attached to C(2) and C(3) are among the smallest found in the molecule.

6.7.3 Bond angles of the heavy atoms (fig. 6.6.2)

The endo-angles in the cyclopentene ring are all reduced as compared with cyclohexene; the angles C(1)–C(5)–C(4) and C(5)–C(4)–C(3) by more than 10°, the others by 5 to 10°. The reason for this is the strain in the



A along the C(1) -C(2) -bond

B along the C(2) -C(3) -bond

C along the C(3) -C(4) -bond

D along the C(4) -C(5) -bond

E along the C(1) -C(5) -bond

F along the C(5) -C(6) -bond

G along the C(7) -C(4) -bond

H along the C(8) -C(1) -bond

I along the C(9) -C(1) -bond

Fig. 6.6.4. Newman projections along all C–C-bonds.

TABLE 6.6.3

Planes through various parts of the molecule

Plane number	Plane through atoms:	Formula
1	C(1), C(3), C(4), C(5), C(6), C(7), O(10), O(11)	$-.578x+.632y+.517z-.447=0$
2	C(1), C(3), C(4), C(5), C(6), C(7)	$-.634x+.627y+.453z-.056=0$
3	C(5), C(6), O(10), O(11)	$-.324x+.684y+.654z-2.199=0$

x, y and z are the orthogonal coordinates in Å, obtained by the axes transformation of Cruickshank, 1961.

Distances of the atoms to the planes in Å.

Plane number	C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)	C(9)	O(10)	O(11)
1	+.12		+.02	−.03	+.06	+.00	−.20			+.33	−.31
2	−.04	+.36	+.01	+.02	+.04	+.00	−.03	−1.46	+.94		
3					+.00	−.01				+.00	+.00

Angles between the planes in degrees.

$<(1,2)=4.9°$; $<(1,3)=16.8°$; $<(2,3)=21.5°$.

five-membered ring. This strain is also demonstrated by the Newman-projections along C(1)–C(2) and along C(2)–C(3) (see fig. 6.4.4a resp. b) in which the ring dihedral angles are 24.5 and 23.7°. In cyclohexane, which is free from strain, the corresponding angles are 55° (Wohl, 1964).

The angle C(5)–C(4)–C(7) is increased somewhat because of the steric interaction between O(11) and methylgroup C(7).

6.7.4 C–H-bond lengths and the corresponding valence angles (fig. 6.6.1 and 6.6.2)

The mean value of the C–H-bond lengths is 1.01 Å with an experimental s.d. of 0.06 Å, the mean value of the angles C–C–H and H–C–H is 109.8° with an experimental s.d. of 2.9 (estimated s.d.'s .03–.04 Å and 1.5–3.0° respectively). Apparently the model for the electron density of hydrogen, as used in the refinements, is better in the direction perpendicular to the C–H-bond than in the bond direction itself.

The mean value of the C–H distances, as normally found from X-ray Diffraction, is approximately .1 Å smaller than the value accepted from Neutron Diffraction experiments.

6.7.5 Planarity of the conjugated system

From table 6.6.3 is seen that the conjugated system as a whole (plane 1) is not planar, the deviations from planarity being significant.

The two individual double bond systems are reasonably flat, though the displacements of the atoms C(1), C(3), C(4), C(5), C(6) and C(7) from the plane of the double bond C(4)=C(5) are significant. The planes of the two systems make an angle of 21.5°. This is clearly shown in the Newman projection along C(6)–C(5). (Fig. 6.4.4f).

The rotation of the COOH-group of 21° about the C(5)–C(6)-bond is caused by steric interaction of the mythyl groups C(8) and C(9) with O(10). The position of these two methyl groups with respect to the ring is determined by the puckering of the cyclopentene ring through C(2). A flat ring would make it possible for the whole conjugated system to be planar but would lead to completely eclipsed conformations about C(1)–C(2) and C(2)–C(3) and a very short distance between O(11) and methylgroup C(7).

6.7.6 The conformation of the methyl groups.

The Newman projections 6.6.4 g, h and i give the attachments of the methyl groups to the ring.

The system C(5), C(4), C(7) can be compared with propene, in which the double bond eclipses a hydrogen atom of the methyl group (Herschbach and Krisher, 1958). As shown in fig. 6.6.4 g H(2) of methyl group C(7) is rotated 20° about C(4)–C(7) out of the eclipsed position. This is

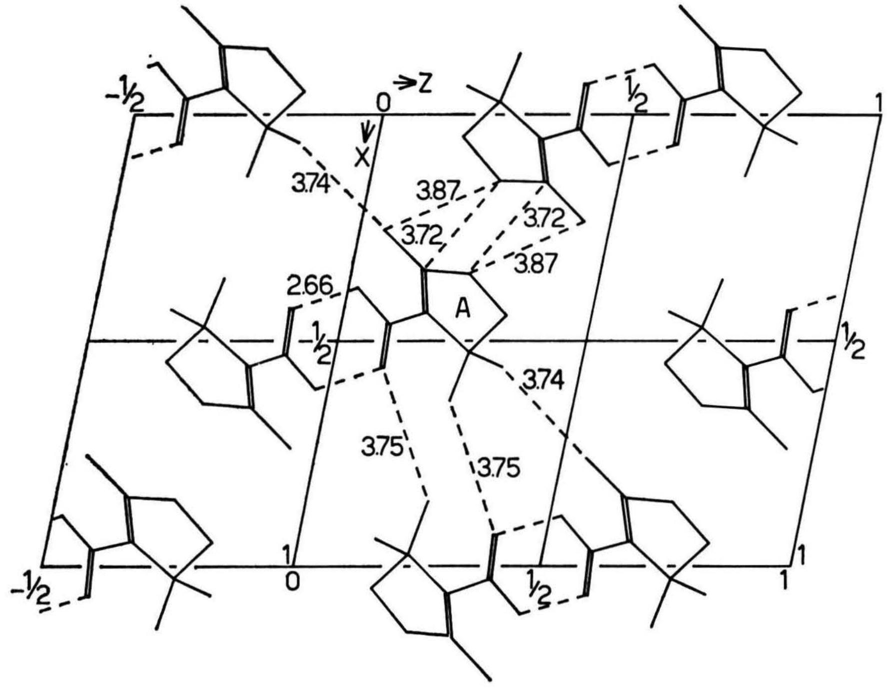


Fig. 6.6.5a

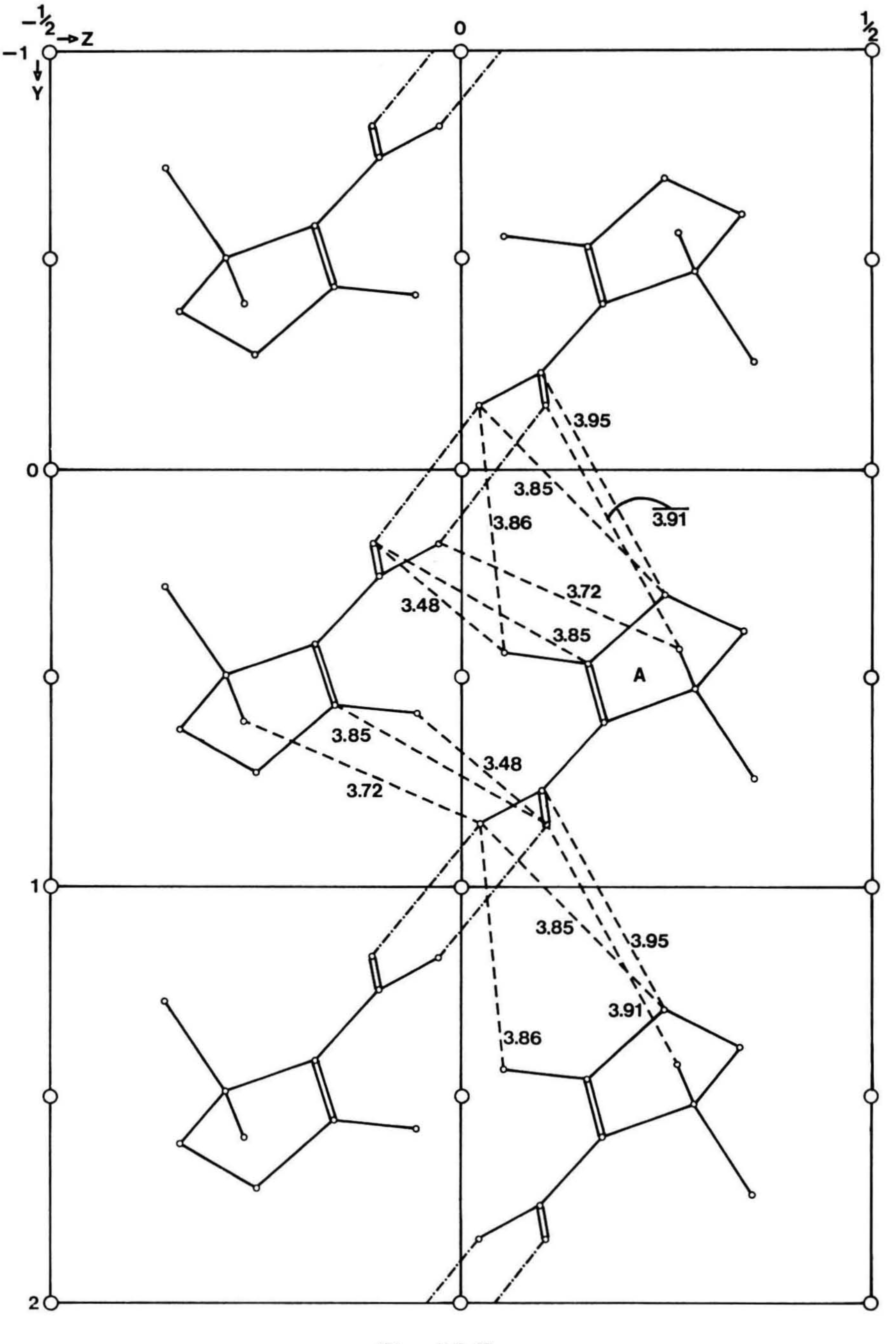


Fig. 6.6.5b

Fig. 6.6.5. The crystal structure of C9-acid viewed along the y-axis (fig. 6.6.5 a) and along the x-axis (fig. 6.6.5 b). The intermolecular distances between molecule A and the adjacent molecules are given partly in fig. a and partly in fig. b.

caused by steric interaction of H(2) with O(11). The distance between these atoms is 2,35 Å.

The methyl groups C(8) and C(9) are attached to the ring fairly well in the staggered conformation (see fig. 6.6.4 h and i).

6.7.7 Packing (fig. 6.6.5 a, b)

The molecules are bound in pairs by hydrogen-bonds of 2.665 Å between the COOH-groups.

The "double molecules" are arranged in fairly close-packed rows along the [010]-axis. In these rows (see fig. 6.6.5 b) most of the shorter contacts ($<$3.9 Å) occur.

Between different rows only a few short contacts are found, as is shown in figure 6.6.5 a.

CHAPTER 7

The crystal structure of *retro* vitamin-A acid

7.1 *Introduction*

Retro vitamin-A acid is an intermediate in the vitamin-A synthesis of Huisman, Smit, Van Leeuwen and Van Rij (1956). In fig. 7.6.4 the structural formula and the numbering of the atoms are given.

7.2 *Crystal data*

Retro vitamin-A acid crystallizes in the monoclinic system. Systematic absences are the h0l-reflections with h = 2n and the 0k0-reflections with k = 2n + 1. Thus the space group is $P2_1/a$.

The unit cell dimensions were obtained by means of zero layer Weissenberg photographs about two principal axes, calibrated with Al-powder lines.

$a = 16.644$ (3) Å
$b = 15.470$ (2) Å
$c = 7.193$ (2) Å
$\beta = 98.92$ (2) °

The density measured by flotation is 1.10 $g.cm^{-3}$. For $Z = 4$ the calculated density is 1.09 $g.cm^{-3}$.

7.3 *Intensity data*

7.3.1 Measurements

The intensities were collected with the Nonius Automatic Single Crystal Diffractometer.

Some data:

Radiation:	$CuK\alpha$
scan:	θ–2θ (moving crystal, moving counter)
scan angle:	2.0° ($\theta < 55°$), 2.5° ($\theta \geqslant 55°$)
maximum θ-value:	69.0°
reference reflections:	10 hol-reflections and 10 reflections at $\chi = 30°$
maximum peak height:	5000 counts per second (stronger reflections were reduced by using attenuation filters).
mounting axis of the crystal:	[010]
length of the crystal:	.72 mm
cross-section of the crystal:	.55 × .6 mm

7.3.2 Reduction

In the accessible part of the reciprocal lattice 2371 reflections were found to be significant (i.e. netcount > 2.5 times the square root of total + background). The total number of reciprocal points in this part of the lattice is 3429.

To the net counts the $(LP)^{-1}$-correction was applied. The absorption correction was ignored ($\mu = 6$ cm^{-1}).

7.4 *Structure determination*

The structure has been determined by means of the Symbolic Addition method for centrosymmetric crystals (KARLE and KARLE, 1966), partly by hand (see chapter 3), partly with the help of the semi-automatical programs (see chapter 4).

7.4.1 First attempts

Before applying the symbolic addition method to *retro* vitamin-A acid we had tried to determine the structure by means of the Patterson method. The orientation of the double bond system was easily found. However, nothing that could be related to the carboxylic group was found.

Therefore, in setting up the trial models we were guided by packing considerations and, what appeared to us a reasonable configuration of the carboxylic end of the molecule (see fig. 1.4.1). Our attempts were unsuccessful because the configuration around the double bond C(12)–C(13) proved later to be *cis* instead of *trans*.

7.4.2 Statistical data

The normalized structure factor amplitudes $|E_H|$ were obtained from the expression:

$$|E_H|^2 = \frac{K^2 \exp\left(2B\,\dfrac{\sin^2\theta}{\lambda^2}\right) |F_H|^2}{\epsilon \Sigma f_j^2 \left(\dfrac{\sin^2\theta}{\lambda^2}\right)}$$

$K = .21$ and $B = 5.3$ Å^2 were found from a Wilson-plot. $\epsilon = 2$ for h01- and 0k0-reflections and $\epsilon = 1$ for all others.

The computed statistical quantities are:

$\langle|E|\rangle = .728$ (.798)
$\langle|E|^2\rangle = .987$ (1.000)
$\langle|E^2-1|\rangle = 1.036$ (.968)
reflections with $|E| > 3.0$: .6 % (.3 %)
reflections with $|E| > 2.5$: 1.5 % (1.2 %)
reflections with $|E| > 2.0$: 5.2 % (4.6 %)
reflections with $|E| > 1.4$: 16.0 % (16.2 %)
reflections with $|E| > 1.0$: 31.4 % (31.7 %)

The theoretical values, given in parentheses, are those of KARLE, HAUPTMAN, KARLE and WING (1958).

7.4.3 Set of symbolically signed reflections

With program SA 1 (see chapter 3) the 261 strongest |E|-values were separated from the rest of the |E|-data and distributed over the eight parity groups. The parity groups with $k=2n$ contained two times as many reflections as the groups with $k=2n+1$, indicating a pseudo translation of $\frac{1}{2}$ b.

The 261 reflections were used to compute a Σ_2-list. Reflections $\bar{4}$ 0 1 ($|E|=3.48$), 10 3 4 (3.62) and $\bar{1}$ 4 6 (3.33) were chosen as origin defining reflections. We used seven symbols, five in the parity groups with $k=2n$ and two in the other parity groups. In several cycles of sign searching by hand the symbolic signs of 98 reflections were found with probabilities $>.99$.

In two cycles with program SA 2 (see chapter 4) 43 additional reflections were signed. During the last cycle two very strong relations between the symbols were incorporated.

7.4.4 Analysis of the symbols

With the 141 symbolically signed reflections an analysis of the five remaining symbols was carried out with program SA 3. (see chapter 4). On the basis of the consistency criteria two of the 32 different solutions were far better than all others. For these solutions a new cycle with program SA 3 was computed with 173 signed reflections as input data. The difference between the consistency criteria of both solutions were very small, as shown in table 7.4.1.

7.4.5 E-map of the better solution

A Fourier synthesis with 299 E's of the better solution was calculated. The positions of 21 peaks in this E-map made good chemical sense.

TABLE 7.4.1

The consistency criteria, calculated from 173 signed reflections. The meaning of criteria A, B, A′ and B′ is given in 4.3.2.

Number of the solution	The number of very good reflections	The sum total of their arguments	The number of good reflections	A	B	The number of bad reflections	A′	B′
1	205	3840	172	2792	286	65	240	94
2	177	3559	184	2924	308	81	314	147

A structure factor calculation for the atoms corresponding to the 17 highest peaks resulted in an R-index of 40 %. Subsequently a Fourier synthesis with 650 F's indicated the positions of the five remaining atoms.

A structure factor calculation for all atoms led to an R-index of 27.5 % and further refinement proceeded smoothly. All 299 reflections, used in the E-map, were found to have been correctly signed.

7.4.6 E-map of the second solution

The E-map of the second solution showed two molecules related by a shift of a single bond in the plane of the conjugated system. This is schematically shown in fig. 7.4.2. Both molecules have impossible hydrogen bond distances. The atoms of the correct structure are situated midway between the corresponding atoms of these two molecules.

7.5 *Refinement*

7.5.1 Block diagonal least squares

The refinement of the structure was carried out by a block diagonal least squares method, using CRUICKSHANK's (1961) weighting scheme and the atomic scattering factors from the International Tables for X-ray Crystallography, vol. III.

Four cycles of least squares with individual isotropic thermal parameters for all atoms reduced the R-value to 16 %.

Anisotropic thermal parameters were then introduced and after three more cycles the R-value had dropped to 10 %. A difference synthesis showed all hydrogen atoms and also some disorder of the out-of-plane atoms C(2), C(17) and C(18) of the ring (see 7.5.2).

In the two following least squares cycles the hydrogen atoms were included except one, which gave overlap with the disordered atom C*(18). The hydrogen atoms were refined with individual isotropic thermal parameters. The R-value dropped to 6.8 %.

7.5.2 Disorder in the ring

The disorder is caused by the occurrence of molecules with a different conformation of the ring, C(2), C(17) and C(18) being reflected in the plane of C(1), C(3), C(4), C(5) and C(6). We shall indicate the reflected atoms by C*(2) C*(17) and C*(18).

The positions of C*(2), C*(17) and C*(18) and also of H(4) were derived from a second difference synthesis. From the peak heights of the disordered atoms it was estimated that about 10 % of the molecules had the alternative conformation. The residual electron density in the rest of the unit cell varied between +.15 and −.15 e.Å^{-3}.

Because C*(18) and H(4) overlap, in the next least squares cycles the positions of these atoms were kept fixed. The final R-value was 5.3 %.

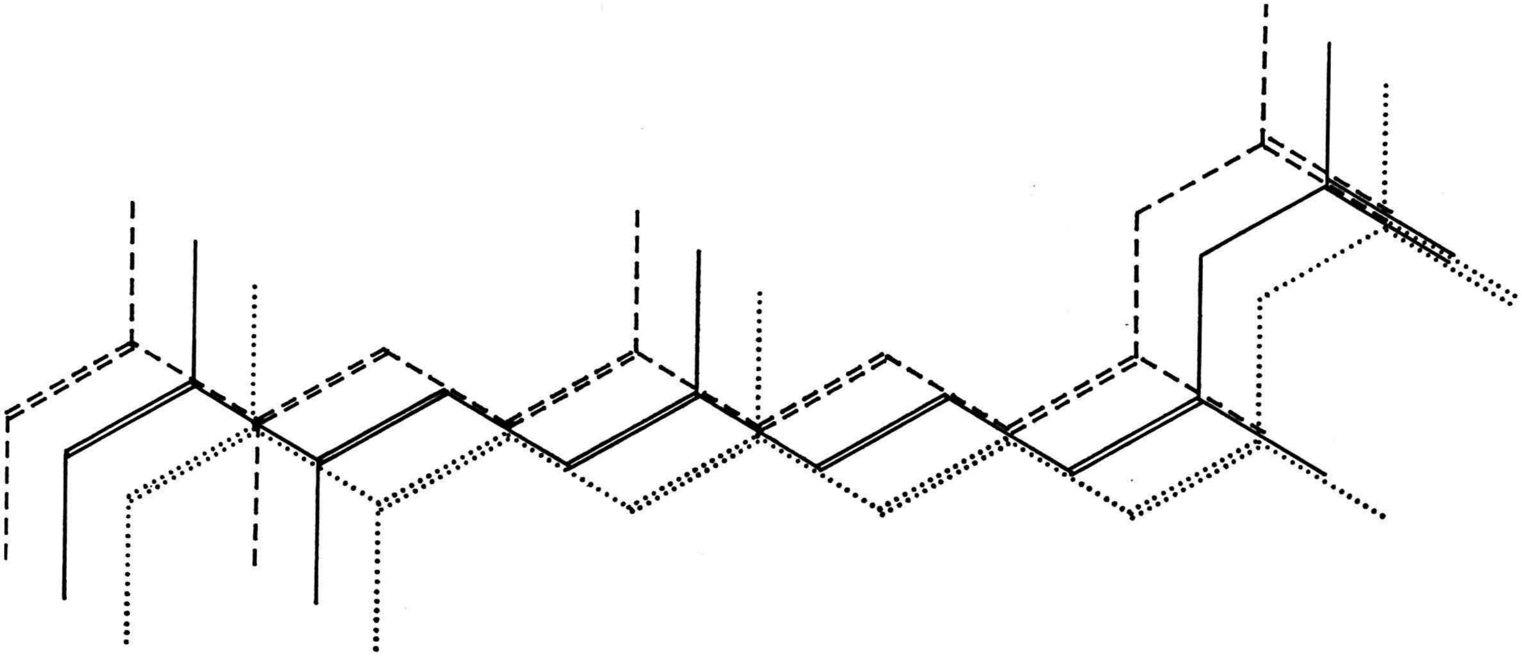


Fig. 7.4.2. Schematical drawing of the relation between the two molecules of the false second E-map and the correct molecule.

legend: ——— correct molecule
...... false molecule 1
- - - - - false molecule 2

7.6 *The crystal structure in numbers and figures*

In table 7.6.1 the final positional and thermal parameters of the carbon and oxygen atoms are listed together with their e.s.d.'s. The e.s.d.'s are those given by the method of block diagonal least squares.

The positional and thermal parameters of the hydrogen atoms are given in table 7.6.2. For their numbering, see fig. 7.6.4. Fig. 7.6.1 shows the bond lengths with their e.s.d.'s between brackets and fig. 7.6.2 gives the bond angles between the heavy atoms. The bond angles C–C–H and H–C–H are given in table 7.6.3.

Important intramolecular distances between the heavy atoms are given in fig. 7.6.3 and the shorter intramolecular hydrogen hydrogen and hydrogen oxygen distances in fig. 7.6.4. Newman projections along all carbon-carbon bonds are shown in fig. 7.6.5.

Planes through various parts of the conjugated double bond system have been calculated by the method of SCHOMAKER, WASER, MARSH and BERGMAN, (1959). The results of these calculations are summarized in table 7.6.4.

In figure 7.6.6 the packing of the molecules is shown. Intermolecular distances smaller than 4.0 Å are indicated in this figure and in table 7.6.5.

7.7 *Discussion of the structure*

7.7.1 Influence of the conformational disorder

The disorder manifested itself in the occurrence of three satellite peaks, which were interpreted as mirror images of C(2), C(17) and C(18) in the plane of C(1), C(3), C(4), C(5) and C(6). Now it is not to be expected that only C(2), C(17) and C(18) will be affected by the disorder.

Therefore in principle, the positions of the other atoms as found from the least squares refinement are averages of the positions in both conformations. These effects will have been incorporated in the temperature factors. As a result the bond distances as found may well deviate considerably from their real values.

7.7.2 Configuration of the molecule

Contrary to expectation the configuration about bond C(12)=C(13) is *cis* instead of *trans*. In *retro-β*-ionylidene acetic acid the same situation exists at the carboxylic end of the molecule (chapter 8).

The connection between the six membered ring and the chain is *trans*.

7.7.3 Bondlengths of the heavy atoms. (fig. 7.6.1)

It is to be noted that no attempt has been made to correct the bond lengths for thermal motion.

The short distance C(2)–C(3) of 1.483 Å and the significant difference between C(1)–C(17) and C(1)–C(18) may be considered as due to the disorder (see 7.7.1).

TABLE 7.6.1.

Final parameters and their e.s.d.'s of the carbon and oxygen atoms. The positional parameters x, y and z are given in fractions. The thermal parameters, based on the expression $\beta_{11} h^2 + \beta_{22} k^2 + \beta_{33} l^2 + 2\beta_{12} hk + 2\beta_{23} kl + 2\beta_{31} lh$ have been multiplied with 10^5.

Atom	x	y	z	β_{11}	β_{22}	β_{33}	$2\beta_{12}$	$2\beta_{23}$	$2\beta_{31}$
C (1)	.33669 (15)	.00146 (14)	.18168 (33)	536 (11)	383 (10)	2239 (52)	21 (17)	−44 (37)	106 (39)
C (2)	.33775 (24)	−.04704 (20)	−.00692 (50)	876 (20)	488 (15)	3249 (87)	9 (28)	−567 (58)	670 (67)
C (3)	.26228 (26)	−.03505 (22)	−.14398 (43)	1120 (25)	727 (18)	2536 (68)	−204 (35)	−836 (59)	−361 (66)
C (4)	.24232 (20)	.05817 (21)	−.16727 (37)	725 (15)	818 (18)	1939 (55)	−18 (28)	−65 (51)	−228 (46)
C (5)	.26649 (13)	.11868 (16)	−.03981 (30)	385 (9)	563 (12)	1960 (47)	−24 (17)	238 (38)	262 (33)
C (6)	.31333 (12)	.09699 (13)	.14470 (28)	353 (8)	393 (9)	1968 (45)	−31 (14)	108 (32)	237 (30)
C (7)	.33097 (13)	.16079 (13)	.27374 (31)	421 (9)	354 (9)	2213 (49)	−7 (15)	75 (34)	249 (33)
C (8)	.36941 (14)	.15738 (13)	.46690 (31)	473 (10)	365 (9)	2192 (49)	−37 (16)	−58 (35)	194 (35)
C (9)	.39556 (13)	.22608 (13)	.57759 (31)	394 (9)	378 (9)	2229 (48)	−20 (15)	−76 (34)	358 (33)
C(10)	.43522 (14)	.21029 (14)	.76694 (31)	449 (10)	409 (9)	2180 (49)	−32 (16)	−243 (35)	403 (35)
C(11)	.47479 (14)	.26659 (14)	.89061 (32)	448 (10)	399 (9)	2291 (51)	−20 (16)	−289 (36)	439 (35)
C(12)	.51492 (14)	.24086 (14)	1.07398 (32)	480 (10)	402 (10)	2138 (48)	4 (16)	−223 (35)	458 (35)
C(13)	.56327 (14)	.28660 (15)	1.20427 (31)	433 (9)	476 (11)	2152 (49)	59 (17)	−312 (38)	478 (34)
C(14)	.58369 (15)	.38094 (15)	1.17645 (33)	479 (11)	512 (11)	2292 (51)	−179 (18)	−477 (41)	581 (38)
C(15)	.54807 (14)	.43695 (14)	1.31320 (32)	447 (10)	417 (10)	2200 (49)	−128 (17)	−285 (37)	241 (35)
C(16)	.24113 (17)	.21129 (19)	−.08256 (37)	547 (12)	678 (15)	2505 (57)	265 (22)	815 (49)	195 (42)
C(17)	.42491 (25)	−.00942 (24)	.27815 (70)	704 (18)	601 (18)	6687 (168)	565 (31)	−784 (88)	−1053 (91)
C(18)	.27997 (26)	−.03950 (20)	.29887 (57)	1002 (23)	429 (14)	4493 (109)	119 (30)	908 (64)	1802 (83)
C(19)	.38740 (19)	.31791 (15)	.50743 (38)	743 (15)	384 (10)	2829 (64)	22 (21)	−55 (42)	−72 (49)
C(20)	.60323 (19)	.24602 (20)	1.38474 (39)	679 (15)	658 (15)	2495 (62)	166 (25)	−80 (50)	−133 (47)
O (1)	.47999 (11)	.47445 (12)	1.24837 (24)	585 (9)	641 (10)	2327 (39)	252 (15)	−149 (32)	147 (29)
O (2)	.58083 (11)	.44499 (13)	1.47671 (25)	560 (9)	809 (12)	2452 (41)	247 (17)	−1122 (36)	−198 (30)

TABLE 7.6.2

Parameters of the disordered carbon atoms and the hydrogen atoms with their e.s.d.'s between brackets. x, y and z are given in fractional coordinates, B in Å^2. The e.s.d.'s of the coordinates of C*(18) and H(4) are assumed.

Atom	x	y	z	B
C* (2)	.2695 (14)	−.0518 (16)	0.416 (34)	4.1 (.5)
C*(17)	.4075 (22)	−.0139 (24)	.1377 (51)	7.2 (.8)
C*(18)	.3150 (40)	−.0230 (50)	.4400 (100)	7.1 (.8)
H (1)	.4370 (22)	−.0688 (24)	.2734 (49)	8.4 (1.0)
H (2)	.4620 (23)	.0230 (25)	.2395 (53)	9.1 (1.1)
H (3)	.4215 (35)	.0067 (35)	.4305 (74)	14.4 (1.7)
H (4)	.2620 (40)	−.0030 (50)	.4150 (100)	6.2 (1.3)
H (5)	.2235 (19)	−.0356 (21)	.2121 (44)	6.5 (.8)
H (6)	.2871 (19)	−.0944 (21)	.3201 (44)	7.6 (.8)
H (7)	.3516 (19)	−.1116 (21)	.0322 (43)	5.4 (.8)
H (8)	.3858 (25)	−.0132 (27)	−.0820 (56)	7.8 (1.1)
H (9)	.2581 (21)	−.0628 (23)	−.2635 (49)	6.5 (.9)
H(10)	.2266 (32)	−.0650 (36)	−.0468 (73)	12.9 (1.7)
H(11)	.2142 (18)	.0747 (19)	−.2808 (41)	5.7 (.7)
H(12)	.2165 (18)	.2174 (20)	−.2128 (43)	5.3 (.8)
H(13)	.2838 (18)	.2479 (22)	−.0754 (43)	5.7 (.8)
H(14)	.2107 (18)	.2348 (20)	.0127 (42)	4.9 (.7)
H(15)	.3158 (13)	.2176 (14)	.2288 (30)	2.3 (.5)
H(16)	.3785 (15)	.1007 (16)	.5297 (35)	3.8 (.6)
H(17)	.3725 (19)	.3217 (20)	.3810 (43)	5.3 (.8)
H(18)	.3619 (27)	.3522 (30)	.5894 (63)	10.0 (1.3)
H(19)	.4381 (27)	.3521 (29)	.5463 (63)	9.5 (1.3)
H(20)	.4306 (14)	.1510 (16)	.8058 (34)	3.1 (.5)
H(21)	.4763 (14)	.3260 (15)	.8564 (33)	2.9 (.5)
H(22)	.5064 (16)	.1791 (18)	1.1034 (37)	4.0 (.6)
H(23)	.5852 (21)	.1821 (24)	1.3948 (48)	7.6 (.9)
H(24)	.5904 (19)	.2758 (21)	1.5020 (45)	6.0 (.9)
H(25)	.6619 (22)	.2575 (24)	1.3996 (51)	7.5 (.9)
H(26)	.6436 (16)	.3898 (18)	1.2000 (37)	4.3 (.7)
H(27)	.5610 (14)	.4009 (15)	1.0479 (33)	2.8 (.5)
H(28)	.4594 (27)	.5114 (30)	1.3716 (61)	10.7 (1.5)

The methyl groups C(16), C(19), and C(20) have significantly shortened bondlengths (mean value 1.507 Å) as compared with the accepted H_3C–C-length in a saturated system. BART (1967) found the same lengths in his carotenoids and ascribed this effect to hyperconjugation.

7.7.4 Bond angles of the heavy atoms (fig. 7.6.2)

The angle deformations in *retro*-vitamin-A acid can be interpreted on the basis of the accepted Van der Waals radius of hydrogen (1.2 Å).

The endo angles of the cyclo-hexene ring are all normal in comparison with related structures (BART and MACGILLAVRY, 1968b), and within the e.s.d.'s equal to the angles found in *retro*-β-ionylidene acetic acid. (chapter 8).

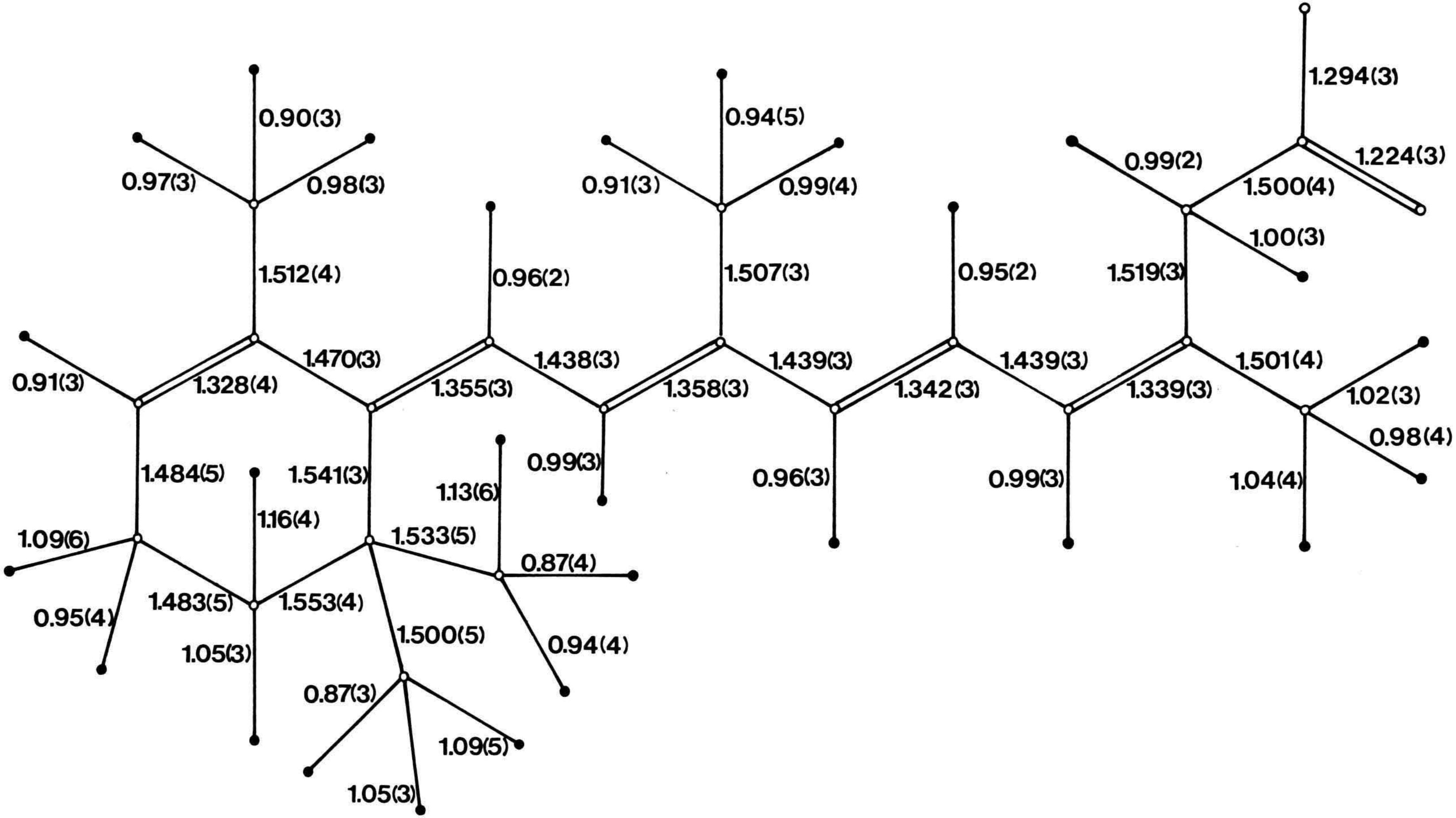


Fig. 7.6.1. Bond lengths.

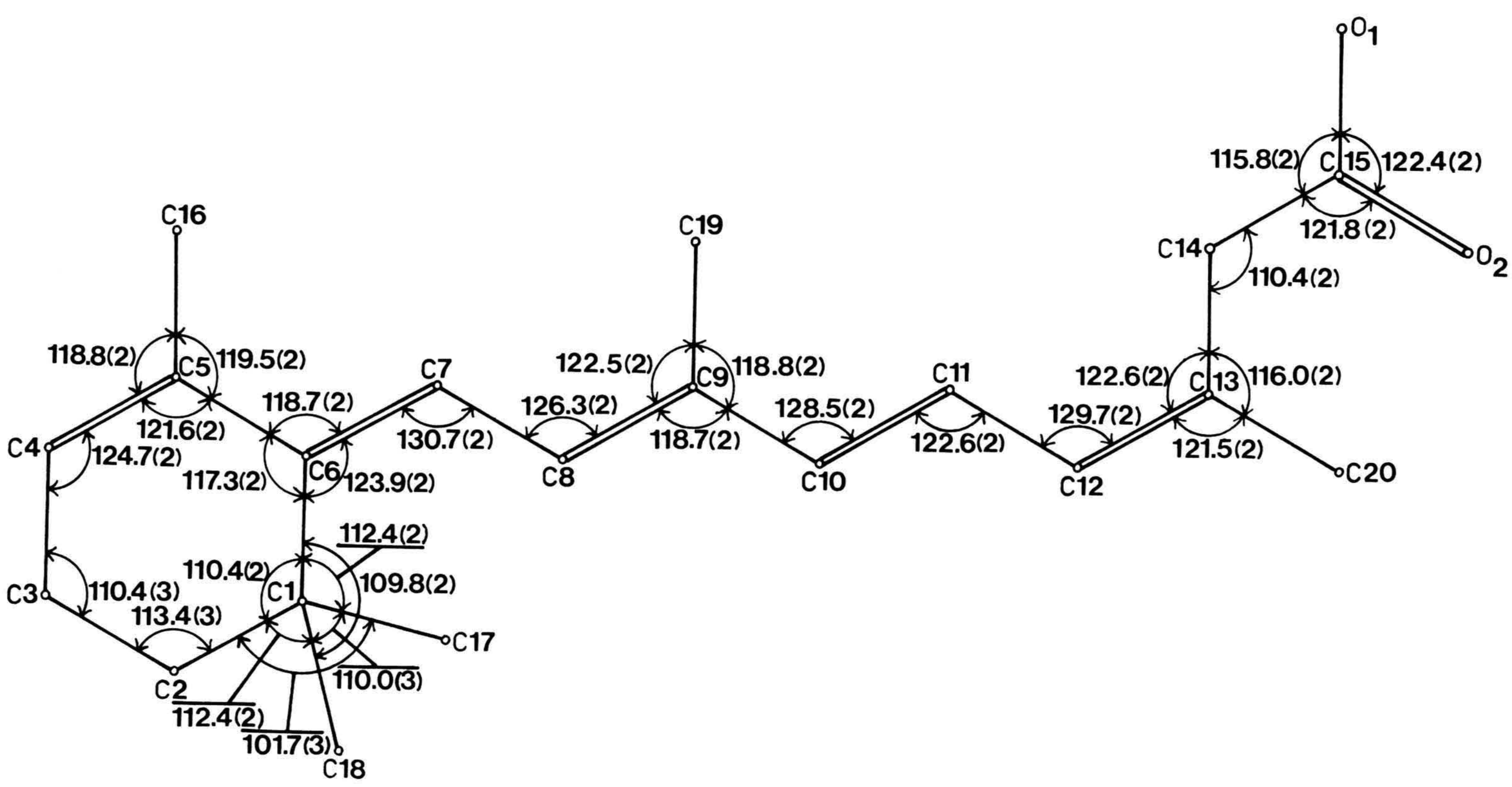


Fig. 7.6.2. Bond angles of the heavy atoms.

TABLE 7.6.3

C—C—H- and H—C—H-bond angles in °, with their e.s.d.'s between brackets.

C(1) -C(2) -H(7)	105 (2)	C(12)-C(11)-H(21)	118 (1)	H(2) -C(17)-H(3)	109 (4)
C(1) -C(2) -H(8)	107 (2)	C(11)-C(12)-H(22)	114 (1)	C(1) -C(18)-H(4)	119 (3)
C(3) -C(2) -H(7)	115 (2)	C(13)-C(12)-H(22)	117 (1)	C(1) -C(18)-H(5)	103 (2)
C(3) -C(2) -H(8)	102 (2)	C(13)-C(14)-H(26)	110 (2)	C(1) -C(18)-H(6)	116 (2)
H(7) -C(2) -H(8)	115 (3)	C(13)-C(14)-H(27)	111 (1)	H(4) -C(18)-H(5)	96 (3)
C(2) -C(3) -H(9)	119 (2)	C(15)-C(14)-H(26)	107 (2)	H(4) -C(18)-H(6)	115 (3)
C(2) -C(3) -H(10)	91 (3)	C(15)-C(14)-H(27)	107 (2)	H(5) -C(18)-H(6)	104 (3)
C(4) -C(3) -H(9)	111 (2)	H(26)-C(14)-H(27)	110 (2)	C(9) -C(19)-H(17)	113 (2)
C(4) -C(3) -H(10)	111(3)	C(5) -C(16)-H(12)	111 (2)	C(9) -C(19)-H(18)	110 (3)
H(9) -C(3) -H(10)	115 (4)	C(5) -C(16)-H(13)	113 (2)	C(9) -C(19)-H(19)	112 (3)
C(3) -C(4) -H(11)	117 (2)	C(5) -C(16)-H(14)	112 (2)	H(17)-C(19)-H(18)	121 (3)
C(5) -C(4) -H(11)	119 (2)	H(12)-C(16)-H(13)	102 (3)	H(17)-C(19)-H(19)	110 (3)
C(6) -C(7) -H(15)	115 (1)	H(12)-C(16)-H(14)	117 (3)	H(18)-C(19)-H(19)	88 (4)
C(8) -C(7) -H(15)	115 (1)	H(13)-C(16)-H(14)	102 (3)	C(13)-C(20)-H(23)	111 (2)
C(7) -C(8) -H(16)	119 (1)	C(1) -C(17)-H(1)	107 (2)	C(13)-C(20)-H(24)	114 (2)
C(9) -C(8) -H(16)	115 (1)	C(1) -C(17)-H(2)	118 (2)	C(13)-C(20)-H(25)	109 (2)
C(9) -C(10)-H(20)	113 (1)	C(1) -C(17)-H(3)	103 (3)	H(23)-C(20)-H(24)	106 (3)
C(11)-C(10)-H(20)	119 (1)	H(1) -C(17)-H(2)	112 (4)	H(23)-C(20)-H(25)	117 (3)
C(10)-C(11)-H(21)	119 (1)	H(1) -C(17)-H(3)	107 (4)	H(24)-C(20)-H(25)	99 (3)

The angles C(17)–C(1)–C(6) and C(6)–C(7)–C(8) are enlarged by 4° and 6° respectively. This is due to steric interaction between H(3) and H(16). The distance between these atoms is still very short (1.82 Å). (see fig. 7.6.4). The same large angles have been found in the comparable structures of β-ionylidene-crotonic acid (Eichhorn and MacGillavry) and *retro*-β-ionylidene acetic acid (chapter 8).

The chain of the molecule is curved in its plane. The angles at the methyl-side of the chain are all larger than the expected values of 125 and 110° and the angles at the other side are smaller than the expected values of 125°. This is caused by the steric interaction of H(15) with H(17), of H(19) with H(21) and of H(21) with H(27). (see fig. 7.6.4). This bending of the chain has been found in related structures, as reviewed by BART and MACGILLAVRY (1968b).

7.7.5 The hydrogen atoms

The mean value of the C–H bond lengths in the methyl groups and methylene groups is 1.00 Å with an experimental standard deviation of .08 Å. The corresponding e.s.d.'s vary between .03 and .06 Å.

The positions of the hydrogen atoms attached to the double bond system have been determined much more accurately than the other hydrogen atoms, judging from the experimental standard deviation (.03 Å) of the mean value of the corresponding C–H-distances (.96 Å). The e.s.d.'s in this case vary between .02 and .03 Å.

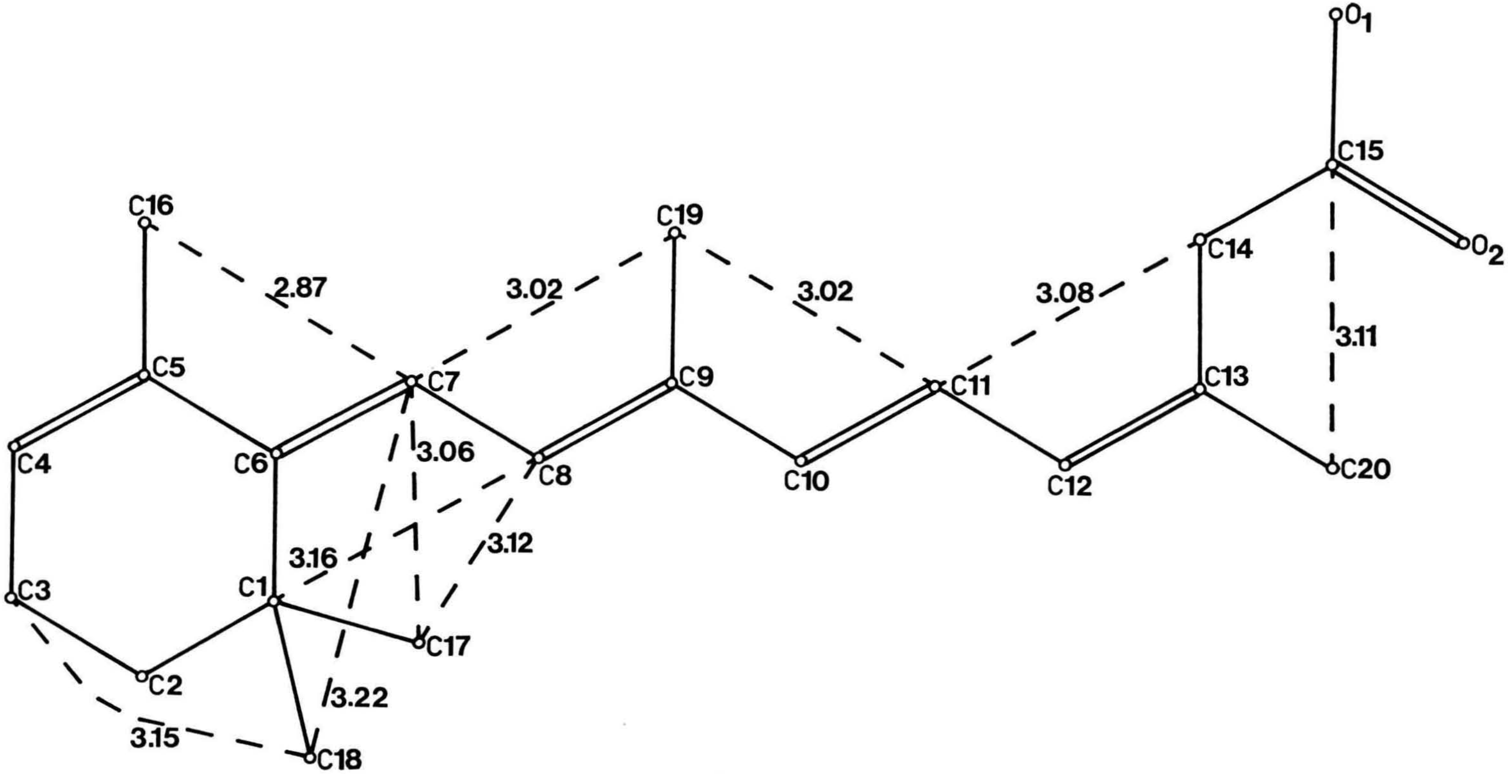


Fig. 7.6.3. Intramolecular distances between the heavy atoms.

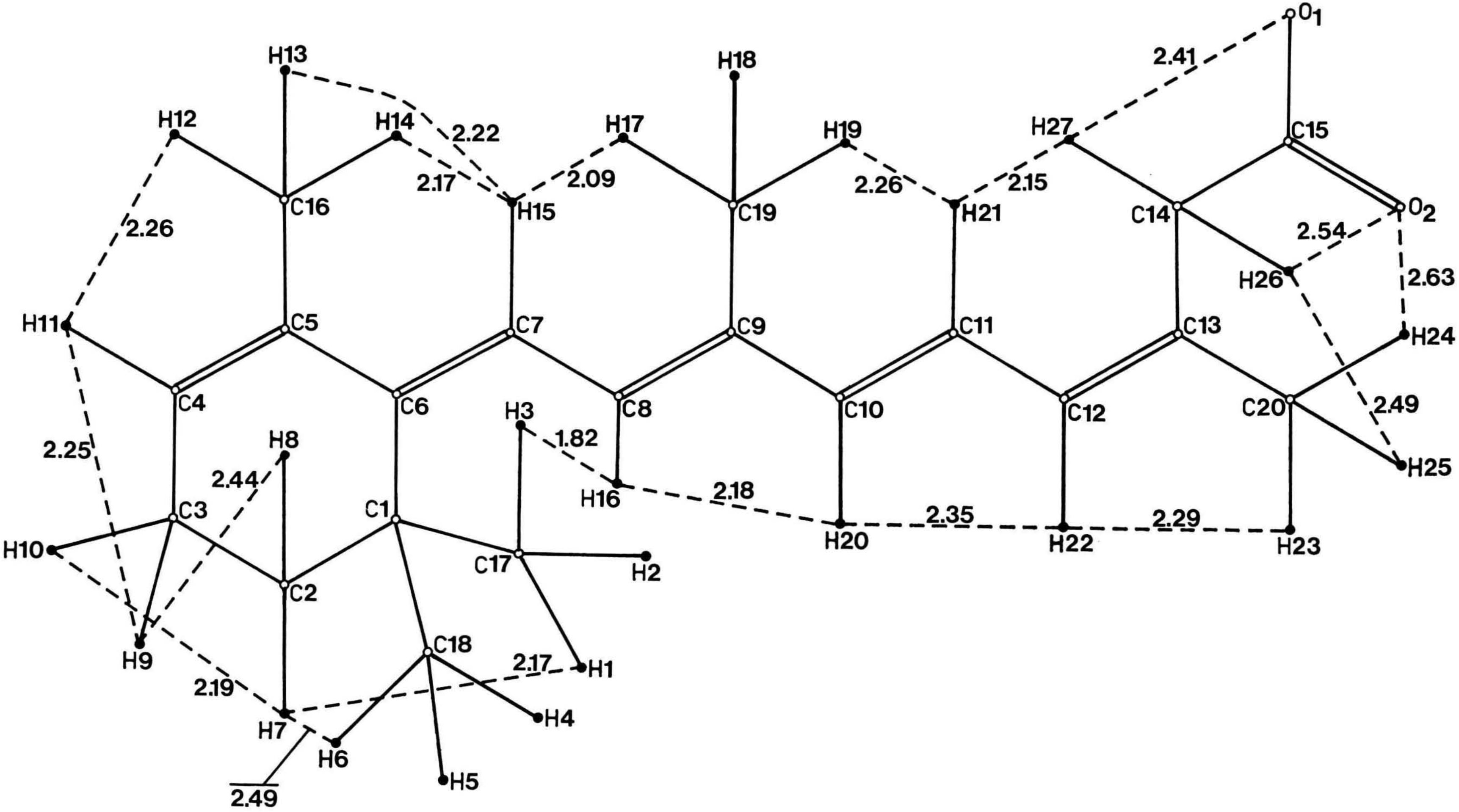
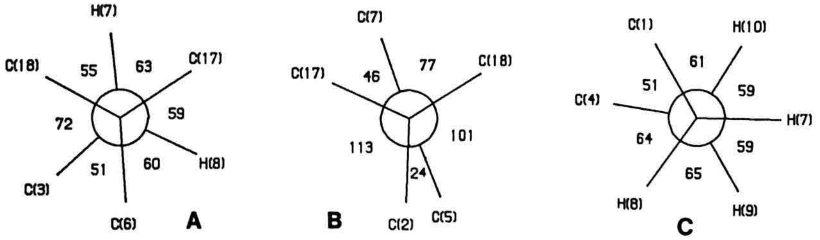
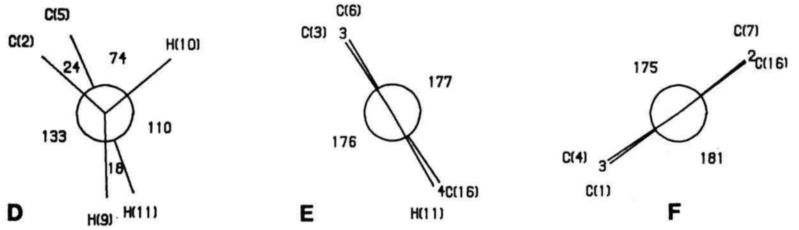


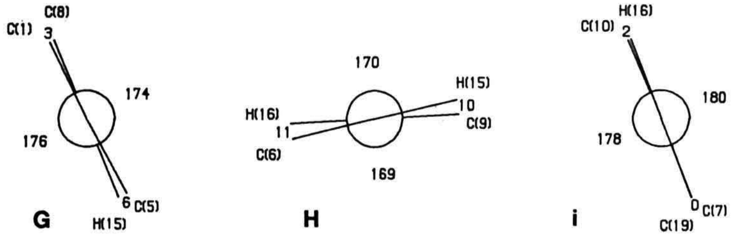
Fig. 7.6.4. Short intramolecular hydrogen hydrogen distances.



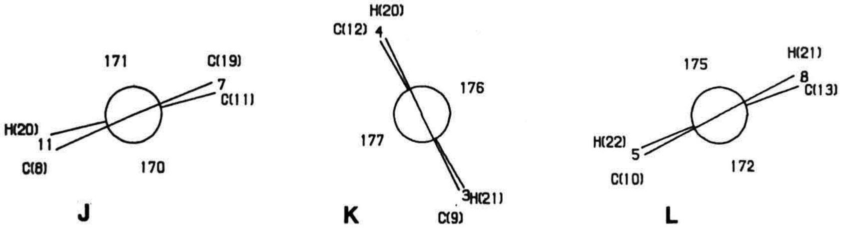
along the C(1)-C(2)-bond along the C(1)-C(6)-bond along the C(2)-C(3)-bond



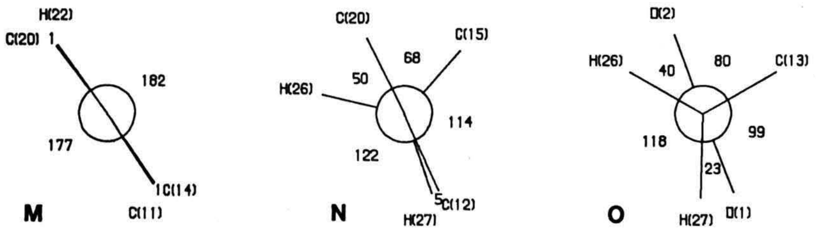
along the C(3)-C(4)-bond along the C(4)-C(5)-bond along the C(5)-C(6)-bond



along the C(6)-C(7)-bond along the C(7)-C(8)-bond along the C(8)-C(9)-bond



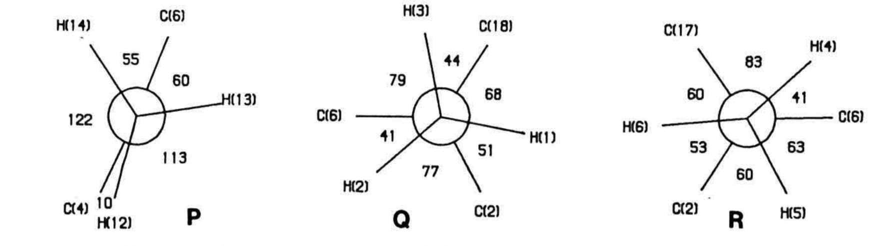
along the C(9)-C(10)-bond along the C(10)-C(11)-bond along the C(11)-C(12)-bond



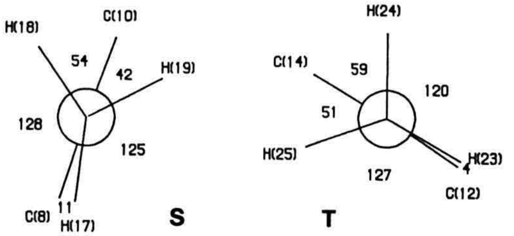
along the C(12)-C(13)-bond along the C(13)-C(14)-bond along the C(14)-C(15)-bond

Fig. 7.6.5 (To be continued)

Fig. 7.6.5 (continued)



along the C(16)-C(5)-bond along the C(17)-C(1)-bond along the C(18)-C(1)-bond



along the C(19)-C(9)-bond along the C(20)-C(13)-bond

Fig. 7.6.5. Newman projections along all carbon carbon bonds.

7.7.6 The planarity of the conjugated system (table 7.6.4)

From the distances to plane 1 it can be seen that the conjugated system is definitely bent out of this plane, the central part being on one side and the end parts on the other side of the plane.

From Newman-projections (fig. 7.6.5) the torsion angles about the individual bonds can be found to be

C(4)=C(5) 0°, C(5)–C(6) 2° cc (counter clock wise),
C(6)=C(7) 3° c (clock wise), C(7)–C(8) 10° c,
C(8)=C(9) 0°, C(9)–C(10) 7° c, C(10)=C(11) 3° cc,
C(11)–C(12) 7° c, and C(12)=C(13) 1° cc.

Thus going from head to tail the overall torsion is clockwise.

7.7.7 The orientation of the methyl groups

The hydrogen atoms in a methyl group attached to a carbon atom with single bonds only tend to be staggered with respect to the substituents on this atom. In a methyl group attached to a carbon atom carrying a double bond one of the hydrogen atoms of the methyl group usually eclipses the double bond. This is rather nicely born out by our results.

The orientations of the methyl groups are shown in the Newman projections fig. 7.6.5 p, q, r, s, and t.

The methyl groups C(16), C(19) and C(20) attached to double bonded carbon atoms do not deviate significantly from the expected orientation.

TABLE 7.6.4

Planes through the double bond systems of *retro* vitamin-A acid. In the expressions x, y and z are the orthogonal coordinates in Å relative to the standard axes.

Plane number	Plane through the atoms	Formula of the plane
1	C(1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 19, 20)	$+.920x+.067y-.385z-4.309=0$
2	C(3, 4, 5, 6, 16)	$+.913x+.143y-.381z-4.445=0$
3	C(1, 5, 6, 7, 8)	$+.929x+.176y-.326z-4.601=0$
4	C(7, 8, 9, 10, 19)	$+.950x-.057y-.308z-4.190=0$
5	C(9, 10, 11, 12)	$+.916x-.178y-.361z-3.321=0$
6	C(11, 12, 13, 14, 20)	$+.870x-.264y-.417z-2.263=0$
7	C(14, 15), O(1, 2)	$-.556x-.788y+.266z+7.084=0$

Distances of the atoms to the planes in Å.

Atom	Plane number						
	1	2	3	4	5	6	7
C (1)	.17		+.00				
C (2)							
C (3)	.22	+.00					
C (4)	.09	−.01					
C (5)	.05	+.02	−.02				
C (6)	.05	−.01	+.02				
C (7)	−.10		+.03	+.01			
C (8)	−.24		−.03	−.01			
C (9)	−.19			−.02	+.02		
C(10)	−.31			+.01	−.02		
C(11)	−.11				−.01	+.02	
C(12)	−.21				+.01	−.02	
C(13)	+.08					−.02	
C(14)	+.60					+.00	+.00
C(15)							−.01
C(16)	−.09	+.00					
C(17)							
C(18)							
C(19)	+.04			+.00			
C(20)	−.03					+.02	
O (1)							+.00
O (2)							+.00

Important angles between plane A and plane B in degrees.

A ↓ B →	1	2	3	4	5	6	7
1	×	4.4	7.1	8.6	14.2	19.4	48.2
2		×	3.8				
3			×	13.5			
4				×	7.8		
5					×	6.5	
6						×	67.3
7							×

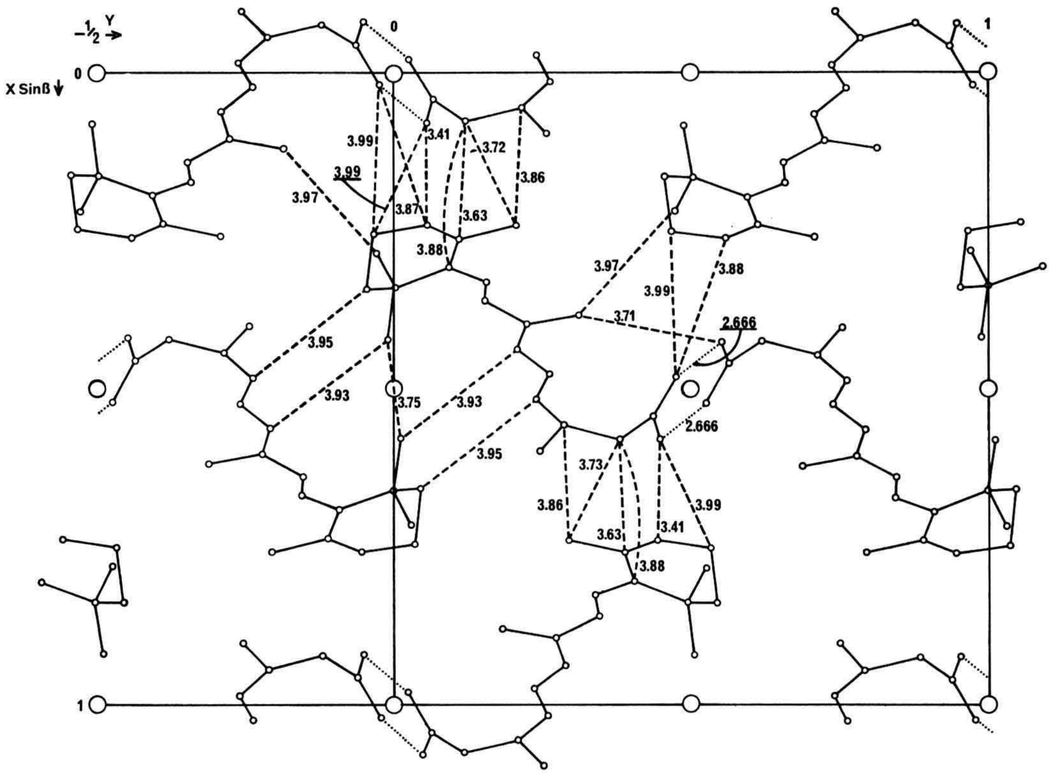


Fig. 7.6.6. The crystal structure of *retro* vitamin-A acid viewed along the z-axis. Intermolecular distances smaller than 4.00 Å have been indicated.

Methyl group C(18) forms part of a nicely staggered conformation. It has to be remarked that the angles corresponding with H(4) are very doubtful because the position of H(4) has not been refined.

Methyl group C(17) is rotated by about 10° out of the fully staggered conformation by the steric interaction of H(3) with H(6) (see fig. 7.6.4). In *retro*-β-ionylidene acetic acid the same has been found (see 8.7.6).

7.7.8 The attachment of the COOH-group to the conjugated chain

The angle between the plane of the COOH-group and the plane of the last double bond of the chain is 67° (see table 7.6.4). The angle of the COOH-group with the overall plane of the conjugated system is 48°.

CHAPTER 8

THE CRYSTAL STRUCTURE OF RETRO-β-IONYLIDENE ACETIC ACID

8.1 *Introduction*

Retro-β-ionylidene acetic acid, also referred to as *retro*-C15-acid, is an intermediate compound of the vitamin-A synthesis of Huisman, Smit, Van Leeuwen en Van Rij (1956).

The structural formula is given in fig. 8.6.1 together with the numbering of the atoms.

8.2 *Crystal data*

The crystals are monoclinic. With a choice of axes such that β is very nearly 90°, the space group is $P2_1/n$.

The cell constants were obtained by means of calibrated zero layer Weissenberg diagrams about two principal axes.

$a = 9.986\ (2)$ Å
$b = 7.816\ (4)$ Å
$c = 17.791\ (4)$ Å
$\beta = 89.55\ (2)\ ^\circ$

The density, measured by flotation, is 1.11 g.cm^{-3}. For $Z=4$ the calculated density is 1.12 g.cm^{-3}.

8.3 *Intensity Data*

8.3.1 Measurements

The intensity data have been collected with the Nonius Automatic Single Crystal Diffractometer.

Data:

Radiation:	CuKα
Scan:	θ–2θ
Scan-angle:	2° ($\theta \leqslant 55°$)–2.5° ($\theta > 55°$)
maximum θ:	68°5
reference reflections:	0 0 12, 0 5 10, 0 4 0
maximum peak height:	5000 counts. sec^{-1} (Stronger reflections were reduced by using attenuation filters).
mounting axis of the crystal:	[1 0 0]
cross-section of the crystal:	.5 × .5 mm
length of the crystal:	.8 mm

8.3.2 Reduction

To the net counts the $(LP)^{-1}$-correction was applied. The absorption correction was ignored because of the low absorption coefficient $\mu = 6\ \text{cm}^{-1}$.

In the course of the measurements the reference intensities decreased with 10 %. A correction for this effect was applied.

A reflection was considered significant if the net count was greater than 2.5 times the standard deviation. In the observed part of the reciprocal space all 2595 reflections were measured and of these, 2139 were significantly above zero.

8.4 *Structure determination*

The structure was determined by the symbolic addition method for centro symmetric structures (KARLE and KARLE, 1966), using the automatic computer programs described in chapter 5.

8.4.1 First Attempts

The first attempts to solve the structure have been made by Mrs. S. Paul–Roy. She used a sharpened Patterson series (sharpening factor: $\sin^2\theta.\ \exp(-2.3.\ \sin^2\theta)$), from which the orientation of the planes of the conjugated part of the molecule and of the carboxylic group were found. However, it was not possible to derive a trial structure which could be refined.

With the same Patterson-series we tried to find a starting model by using the Symmetry Minimum Function (SIMPSON, DOBROTT and LIPSCOMB, 1965), for which the necessary computer programs were written. These attempts were not successful either.

8.4.2 Statistical data

The normalized structure factors $|E_{hkl}|$ were calculated by the expression

$$|E_{hkl}|^2 = \frac{K^2 \exp\left(2B\,\frac{\sin^2\theta}{\lambda^2}\right)|F_{hkl}|^2}{\epsilon\,\Sigma\, f_j{}^2\,\frac{(\sin^2\theta)}{\lambda^2}}$$

K and B ($=4{,}1$ Å^2) were estimated from a Wilson plot. Because the spacegroup is $P2_1/n$, $\epsilon = 2$ for the hol- and the oko-reflections and $\epsilon = 1$ for all others. The computed values for the distribution of $|E|$ were as follows:

$$\begin{aligned}
\langle|E|\rangle &= .802\ \ (.798)\\
\langle|E^2|\rangle &= 1.009\ \ (1.000)\\
\langle|E^2-1|\rangle &= 1.027\ \ (.968)
\end{aligned}$$

reflections with $|E| > 3.0$: .3 % (.3 %)
reflections with $|E| > 2.5$: 1.5 % (1.2 %)
reflections with $|E| > 2.0$: 5.2 % (4.6 %)
reflections with $|E| > 1.4$: 17.4 % (16.2 %)
reflections with $|E| > 1.0$: 32.8 % (31.7 %)

The theoretical values, given in parentheses, have been calculated by KARLE, HAUPTMAN, KARLE and WING (1958). The program ASA 1 delivered a list of all reflections with their $|E|$-values and a Σ_2 list of the reflections with $|E| > 2.1$ both on punch tape and line-printer paper.

8.4.3 The Symbolic Addition Procedure

With Program ASA 4 the starting set of symbolic signs was produced. Some of the results of this program were as follows:

a) The origin was fixed by giving the reflections $\overline{7}$ 0 9 ($|E| = 3.30$), 7 5 4 (3.08) and 3 1 13 (3.08) the sign +.

b) With 4 Symbols and the origin-defining reflections as generators it was possible to give 53 reflections a symbolic sign with a probability of $p > .99$.

c) At this stage 52 reflections belonged to the group $h+k+l = 2n$ and only one reflection (3 1 13) to the group $h+k+l = 2n+1$. By choosing three more symbols 10 more reflections were found in the last group.

d) Thus the resulting starting set of symbolically signed reflections was built up by means of 7 symbols and consisted 63 reflections, all signed with a probability greater than .985.

From the fact that the reflections with $h+k+l = 2n+1$ are rather weak we conclude that quite a lot of atoms in the unit cell have a y-coordinate of nearly zero. As a result the n-glide reflection plane causes a pseudo I-centered cell.

With program ASA 2 the group of signed reflections was increased to 128. The lower bound of $|E|$ in the last stage was 1.8.

The first cycle of program ASA 3 reduced the number of reliable solutions to 56. Some of the rejected solutions were eliminated on the basis of far too large sums of E_{hkl} in the parity groups eee, eoo, oeo and ooe. The others were eliminated by the consistency criteria.

In the second cycle the reflections for computing Fourier series were selected. The consistency criteria of the 10 best solutions are given in table 8.4.1. The lower bound of $|E|$ was 1.5.

Solution number 3 had the lowest sum of arguments against. The reflection tape to compute the Fourier synthesis of this solution consisted of 270 reflections. Eventually the signs of these reflections all proved to be correct.

The E-map (fig. 8.4.1) showed 18 peaks one of which was an artefact. This peak was a continuation of the zig-zag of the conjugated chain, as shown in the figure.

A structure factor calculation with the positions as found from the E-map with the exception of the artefact and with the scale factor and overall temperature factor as estimated from the Wilson-plot gave an R-value of 30 %.

TABLE 8.4.1

The consistency criteria for the ten best solutions. The meaning of the criteria is given in 4.3.2.

Number of the solution	The number of very good reflections	The sum total of their arguments	The number of good reflections	A	B	The number of bad reflections	A′	B′
2	168	3749.8	90	2361.1	320.6	124	794.7	272.5
3	196	4351.0	74	2052.0	283.6	112	639.5	172.5
9	171	3869.1	76	2134.5	307.0	135	866.2	321.9
13	175	3995.1	78	2112.2	305.6	129	795.0	290.7
15	183	4425.1	56	1462.6	214.4	143	988.2	408.4
19	195	4543.3	55	1535.9	209.6	132	870.0	339.8
25	179	4264.0	60	1603.4	209.9	143	1014.0	407.4
26	201	4786.6	45	1268.9	181.3	136	921.8	340.0
30	178	4022.8	75	2074.4	288.6	129	832.5	280.5
55	192	4587.1	47	1319.1	177.5	143	1012.8	402.2

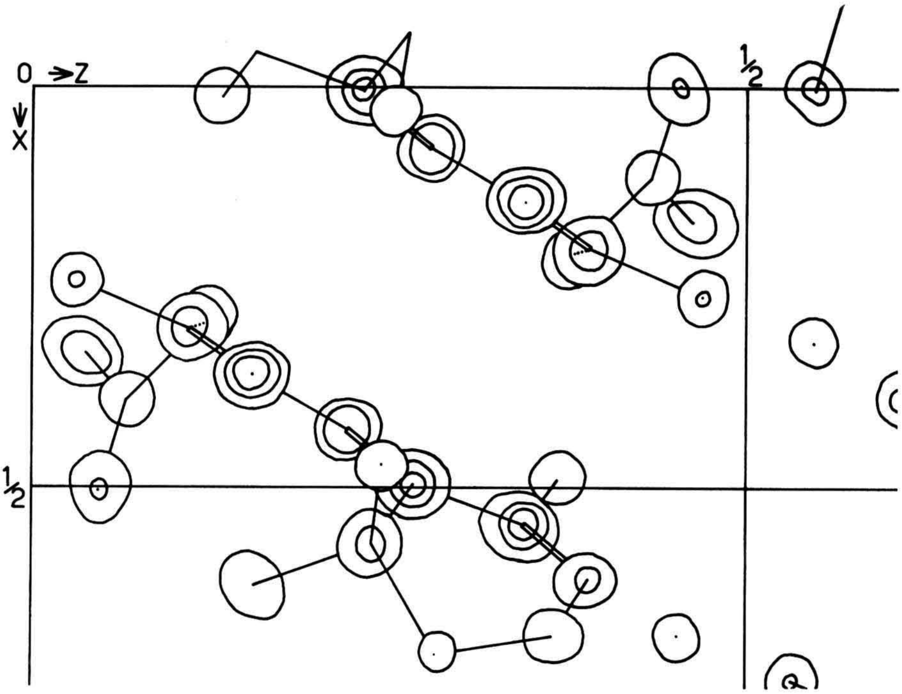


Fig. 8.4.1. Composite E-map of *retro*-β-ionylidene acetic acid.

8.5 *Refinement*

8.5.1 Least squares

The refinement of the positional and thermal parameters was carried out by block diagonal least squares, using the weighting scheme of CRUICKSHANK (1961). The atomic scattering factors were those from the International Tables for X-ray Crystallography, vol. III.

Three cycles of isotropic refinement, using random selections of the observed structure factors, reduced the R value to 21 %. With other random selections four cycles of anisotropic least squares were carried out. At this stage the R-value was 11 % and a difference synthesis was calculated. All hydrogen atoms were located except the one of the carboxylic group. The peak heights of the atoms ranged from .35 e.Å^{-3} to .7 e.Å^{-3}.

The refinement was continued with anisotropic thermal parameters for the carbon and oxygen atoms and isotropic thermal parameters for the hydrogen-atoms. Ten very strong reflections were omitted because they had appreciable measuring errors as a result of the dead time effect of our counting circuit. In three cycles the R-value decreased to 5.3 %.

8.5.2 Disorder in the carboxylic group

The two C–O distances in the carboxylic group were both 1.25 Å at this stage. The oxygen atoms moreover had very great anisotropic thermal parameters as compared with C(11). Two carboxylic groups, related by a centre of symmetry, are connected by hydrogen bonds of 2.6 Å. In a new difference synthesis two overlapping maxima were found corresponding with hydrogen positions belonging to each of the oxygen atoms. From these facts we concluded that the carboxylic group showed positional disorder. In about half the number of molecules throughout the crystal O(1) is the double bonded oxygen atom and in the other half O(1) is single bonded.

A cycle of least squares with two half hydrogen atoms with isotropic thermal parameters of 5.0 Å^2 at the positions as found from the difference Fourier followed. This cycle reduced the R-factor to 5.14 % and produced no shifts greater than the corresponding e.s.d.'s. The refinement was terminated.

8.5.3 The final difference synthesis

The difference synthesis mentioned in 8.4.2 showed a residual electron density within $\pm .2$ e.Å^{-3} with a standard deviation of approximately .02 e.Å^{-3}, calculated with the formula of CRUICKSHANK (1949).

Midway between all pairs of bonded atoms residual electron density was found, the mean height of which is .12 e.Å^{-3}. In the neighbourhood of both oxygen atoms two well shaped peaks of about .2 e.Å^{-3} were found, which could not be interpreted.

8.6 *The crystal structure in numbers and figures*

The final parameters of the carbon and oxygen atoms are listed in table 8.6.1. The parameters of the hydrogen atoms are given in table 8.6.2. The e.s.d.'s are those given by the least squares procedure.

In figure 8.6.1 the bond lengths are given with their e.s.d.'s and figure 8.6.2 shows the bond angles. No attempt was made to correct these for thermal movement. Other important intramolecular distances are given in figures 8.6.3 and 8.6.4. Figure 8.6.5 shows Newman projections along all the C–C-bonds.

By the method of Schomaker, Waser, Marsh and Bergman, (1959) several planes through the atoms of the conjugated bond system were calculated. The results are given in table 8.6.3. The intermolecular distances are shown in the figures 8.6.6 a and b.

8.7 *Discussion of the structure*

8.7.1 Conformation of the conjugated system

The conformation of the molecule was already known from the structure determination of the p.Br-anilide of retro-C15-acid. (Paul–Roy, Schenk and MacGillavry, 1969).

The most unexpected feature of the structure is the cis configuration about the double bond C(8)=C(9).

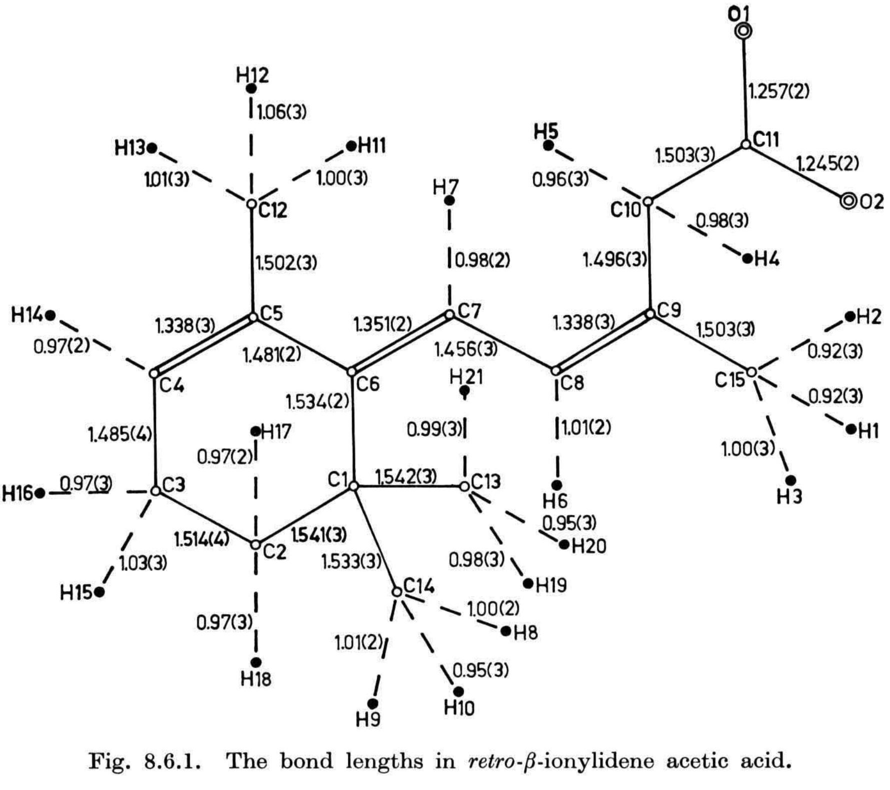


Fig. 8.6.1. The bond lengths in *retro-β*-ionylidene acetic acid.

TABLE 8.6.1.

Final parameters and their e.s.d.'s of the carbon and oxygen atoms. The positional parameters x, y and z are given in fractions. The thermal parameters based on the expression $\beta_{11} h^2 + \beta_{22} k^2 + \beta_{33} l^2 + 2\beta_{12} hk + 2\beta_{23} kl + 2\beta_{31} lh$ have been multiplied with 10^5.

Atom	x	y	z	β_{11}	β_{22}	β_{33}	$2\beta_{12}$	$2\beta_{23}$	$2\beta_{31}$
C (1)	.57306 (18)	.23292 (22)	.23709 (10)	759 (17)	1254 (28)	320 (6)	−44 (35)	−51 (21)	96 (16)
C (2)	.70044 (20)	.27188 (27)	.28233 (13)	873 (20)	1579 (35)	473 (8)	−411 (42)	−159 (27)	−67 (20)
C (3)	.67827 (25)	.26953 (33)	.36663 (14)	1244 (27)	2285 (47)	445 (9)	−162 (56)	−499 (33)	−457 (25)
C (4)	.61296 (22)	.10647 (32)	.38901 (12)	1137 (23)	2263 (43)	306 (7)	498 (51)	−135 (28)	−173 (20)
C (5)	.54053 (18)	.00962 (24)	.34248 (10)	811 (18)	1568 (31)	279 (6)	410 (38)	−4 (22)	67 (16)
C (6)	.50804 (16)	.06688 (21)	.26540 (10)	673 (15)	1206 (26)	265 (5)	219 (32)	−106 (19)	97 (14)
C (7)	.41859 (18)	.97561 (22)	.22533 (10)	815 (18)	1238 (28)	286 (6)	−87 (36)	−14 (21)	44 (16)
C (8)	.35550 (18)	.01808 (23)	.15425 (10)	869 (19)	1318 (29)	296 (6)	−16 (37)	−42 (21)	34 (17)
C (9)	.29546 (18)	.90914 (25)	.10728 (10)	831 (18)	1606 (32)	273 (6)	129 (38)	−160 (22)	70 (16)
C(10)	.29049 (20)	.71965 (25)	.11879 (11)	970 (20)	1503 (32)	329 (6)	−301 (40)	−303 (24)	191 (18)
C(11)	.38226 (17)	.62376 (24)	.06630 (10)	815 (18)	1422 (30)	301 (6)	−163 (37)	−205 (22)	−48 (16)
C(12)	.48526 (24)	.84303 (31)	.37109 (12)	1356 (26)	2018 (42)	348 (7)	74 (53)	416 (29)	−8 (21)
C(13)	.62060 (22)	.22032 (29)	.15461 (12)	1121 (23)	1906 (39)	361 (7)	−341 (47)	167 (27)	356 (21)
C(14)	.47434 (21)	.38190 (25)	.24579 (13)	1012 (21)	1251 (30)	475 (8)	135 (41)	−57 (26)	128 (21)
C(15)	.22812 (27)	.97195 (34)	.03726 (13)	1601 (32)	2388 (50)	319 (7)	611 (62)	−249 (31)	−380 (24)
O (1)	.34547 (16)	.47807 (21)	.04470 (11)	1078 (17)	1972 (30)	628 (8)	−672 (36)	−1036 (25)	385 (18)
O (2)	.49175 (16)	.68785 (22)	.04765 (12)	1103 (18)	1997 (31)	736 (9)	−790 (38)	−967 (28)	731 (20)

TABLE 8.6.2

Final parameters and their e.s.d.'s of the hydrogen atoms.
x, y, z are given in fractional coordinates, B in Å².

Atom	x	y	z	B
H (1)	.2623 (30)	.9146 (40)	—.0037 (17)	5.1 (7)
H (2)	.1389 (33)	.9412 (45)	.0349 (19)	5.9 (8)
H (3)	.2346 (30)	.0985 (40)	.0281 (18)	5.4 (7)
H (4)	.2002 (26)	.6734 (32)	.1107 (15)	3.3 (5)
H (5)	.3149 (26)	.6934 (32)	.1698 (14)	3.4 (5)
H (6)	.3563 (21)	.1436 (26)	.1398 (12)	1.7 (4)
H (7)	.3881 (19)	.8662 (25)	.2463 (11)	1.1 (4)
H (8)	.3898 (24)	.3599 (30)	.2179 (14)	2.9 (5)
H (9)	.4494 (24)	.3909 (31)	.3007 (14)	2.9 (5)
H(10)	.5133 (29)	.4866 (36)	.2290 (16)	4.3 (6)
H(11)	.3856 (28)	.8402 (36)	.3702 (16)	4.3 (6)
H(12)	.5206 (27)	.7423 (32)	.3366 (14)	3.1 (5)
H(13)	.5165 (26)	.8128 (33)	.4235 (14)	3.6 (6)
H(14)	.6208 (25)	.0685 (32)	.4409 (14)	3.3 (5)
H(15)	.6168 (32)	.3684 (40)	.3830 (17)	5.2 (7)
H(16)	.7631 (30)	.2820 (38)	.3926 (17)	4.7 (7)
H(17)	.7679 (25)	.1885 (32)	.2676 (14)	3.0 (5)
H(18)	.7337 (27)	.3831 (33)	.2668 (15)	3.7 (6)
H(19)	.6792 (30)	.1234 (38)	.1437 (17)	4.9 (7)
H(20)	.6741 (28)	.3152 (36)	.1398 (16)	4.4 (6)
H(21)	.5502 (27)	.2062 (34)	.1162 (15)	3.8 (6)

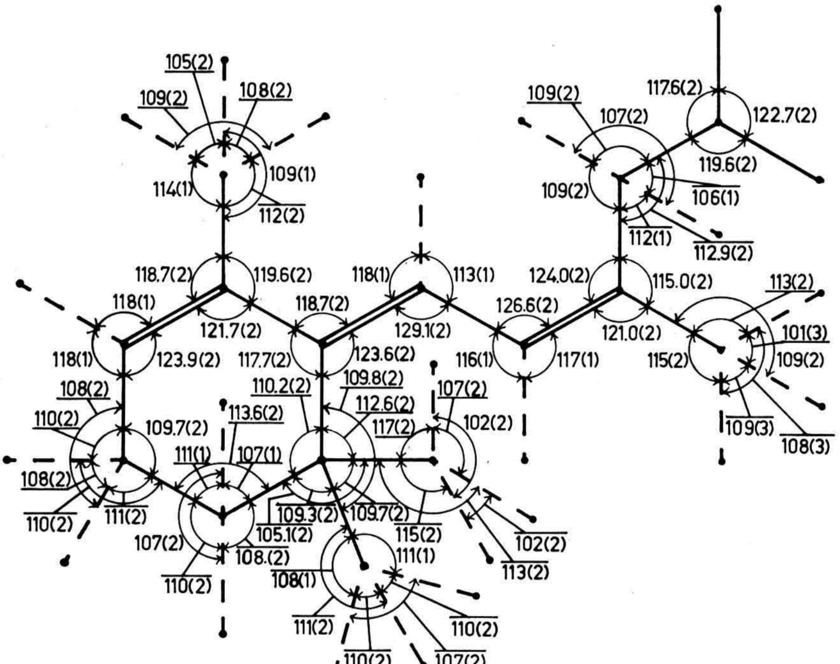


Fig. 8.6.2. Bond angles in *retro*-β-ionylidene acetic acid.

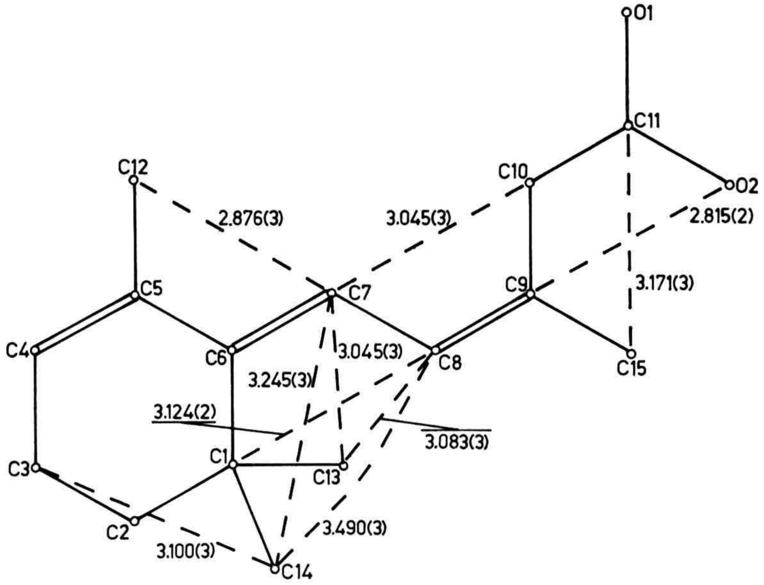


Fig. 8.6.3. Important intramolecular carbon carbon distances in *retro-β*-ionylidene acetic acid.

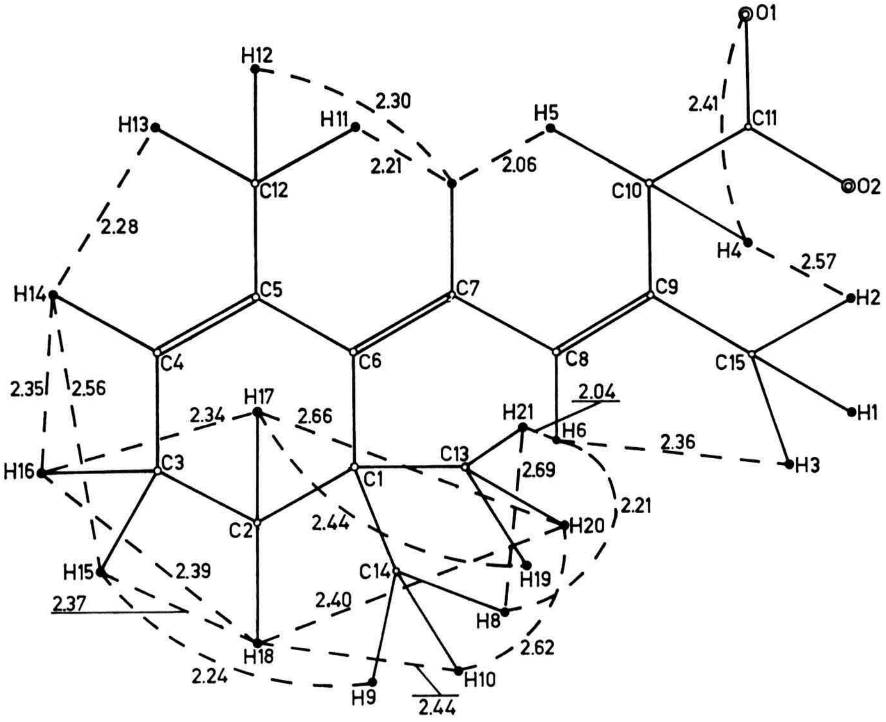


Fig. 8.6.4. Short intramolecular hydrogen hydrogen distances.

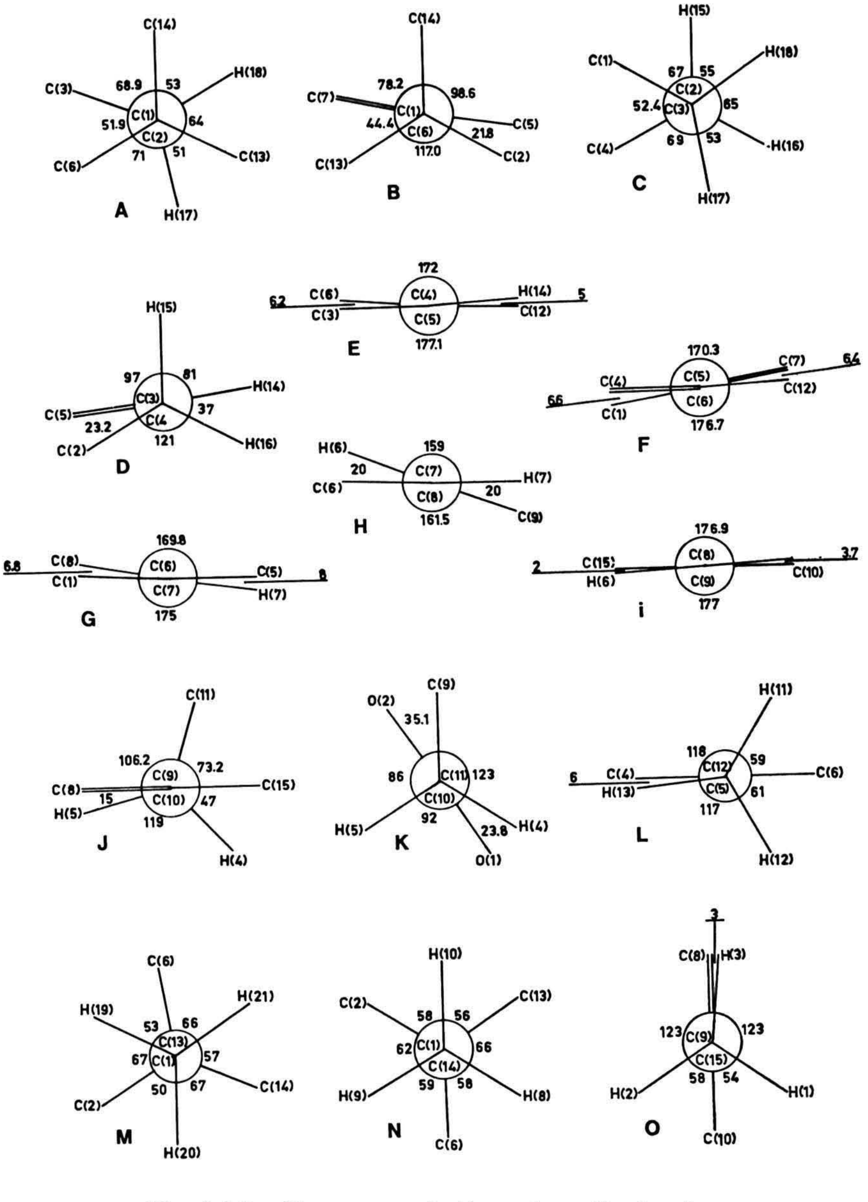


Fig. 8.6.5. Newman projections along the bonds:

a) C(1)–C(2), b) C(1)–C(6), c) C(2)–C(3),
d) C(3)–C(4), e) C(4)–C(5), f) C(5)–C(6),
g) C(6)–C(7), h) C(7)–C(8), i) C(8)–C(9),
j) C(9)–C(10), k) C(10)–C(11), l) C(12)–C(5),
m) C(13)–C(1), n) C(14)–C(1), o) C(15)–C(9).

TABLE 8.6.3

Planes through various parts of the conjugated chain of the molecule. In the given formula's x, y and z are the orthogonal coordinates in Å, obtained by the axes transformation of CRUICKSHANK, (1961).

Plane number	Plane through the atoms:	Formula of the plane
1	C(3, 4, 5, 12, 6, 1, 7, 8, 9, 10, 15)	$.814x-.411y-.411z+1.272=0$
2	C(3, 4, 5, 12, 6)	$.813x-.475y-.338z+1.393=0$
3	C(1, 5, 6, 7, 8)	$.726x-.548y-.415z+2.850=0$
4	C(7, 8, 9, 10, 15)	$.862x-.098y-.498z-\ .910=0$
5	C(10, 11), O(1, 2)	$-.447x+.423y-.788z+\ .586=0$

Distances of the atoms to the planes in Å.

Plane number	C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)	C(9)	C(10)	C(11)
1	+.27		+.07	−.10	−.04	+.07	−.08	−.22	−.02	+.47	
2	+.07	+.69	+.02	−.04	+.01	−.01					
3	+.00				−.05	+.03	+.07	−.06			
4							−.02	+.03	+.00	+.00	
5										−.00	+.01

Plane number	(C12)	C(13)	C(14)	C(15)	O(1)	O(2)
1	−.16			−.26		
2	+.01	+.99	−1.34			
3						
4				−.01		
5					−.00	−.00

Important angles between plane A and plane B in degrees.

A ↓ B→	1	2	3	4	5
1	×	5.5	9.3	18.9	77.7
2		×	7.8		
3			×	27.6	
4				×	88.0
5					×

In most of the carotenoïds and Vitamin-A-related structures determined so far except *cis-* and *trans-β*-ionylidene-crotonic acid, the conformation about the single bond C(6)–C(7) is such, that the plane of the double bond in the ring is rotated approximately 40° out of the single-*cis* position with respect to the double bond system of the chain. In *trans-β*-ionylidene-crotonic-acid the conformation is single-*trans* (EICHHORN and MACGILLAVRY, 1959).

In retro-C15-acid the conformation about the double bond C(6)–C(7) is also *trans*.

8.7.2 Bond lengths of the heavy atoms. (fig. 8.6.1)

The bond lengths in the cyclohexene ring are all normal in comparison with the mean values of related structures from the vitamin-A series (BART and MACGILLAVRY, 1968b) with one exception. For the C(2)–C(3)-bond we find a fairly normal distance of 1.514 Å while the mean value is 1.483 Å.

The bond lengths in the chain are also normal.

There are two types of C(n)–CH_3 bonds. In the one type Carbon atom C(n) is an atom in a system of double bonds, in the other C(n) has single bonds only. The mean C–CH_3 distances for the two types are 1.502 and 1.538 respectively. The difference of .036 Å is significant. The difference in length between C(6)–C(1) and the mean value of C(4)–C(3) and C(9)–C(10) is significant too. Similar differences were also found in the above mentioned structures.

Both differences were explained by BART (1967) in terms of hyper conjugation.

As discussed in 8.5.2 the lengths of C(11)–O(1) and C(11)–O(2) are nearly equal indicating conformational disorder in the COOH-group.

8.7.3 Bond angles of the heavy atoms (fig. 8.6.2)

Because of the similarity between the rings in *retro* vitamin-A acid and in canthaxanthin each of the endo angles in the first ring corresponds to an angle in the second. There are no significant differences between corresponding bond angles.

The endo angles do not differ much from the corresponding angles in non-cyclic compounds. The Newman projections along C(1)–C(2) and C(2)–C(3) (fig. 8.6.5 a, c) show normal dihedral ring angles of 52° as compared with the corresponding dihedral angle in cyclohexane (WOHL, 1964). From these results we can conclude that the ring is relatively free from strain.

The large angle C(6)–C(7)–C(8) of 129.1° is caused by steric interaction between H(6) on the one hand and H(8) and H(21) on the other. The distances H(6)–H(8) and H(6)–H(21) are 2.21 and 2.04 Å respectively, which is considerably shorter than twice the accepted Van der Waals radius for hydrogen (1.2 Å). (See fig. 8.6.4). In β-ionylidene-crotonic acid, which also has the *trans*-conformation about C(6)–C(7) (see 8.7.1), the corresponding angle is 130.0°. (EICHHORN and MACGILLAVRY, 1959).

8.7.4 C–H-bond lengths and the corresponding valence angles

The mean value of the C–H bond lengths is .98 Å with an experimental standard deviation of .03 Å (the corresponding estimated standard deviations vary between .02 and .03 Å). The mean value of the angles

C–C–H and H–C–H is 109° with an experimental standard deviation of 3° (the estimated standard deviations vary between 2° and 3°).

The accepted value of a C–H-bond length, as measured by neutron diffraction, is approximately .1 Å greater than the mean value of .98 as found for the C–H-lengths in *retro*-C15-acid. Normally one finds this apparent bond contraction from X-ray experiments.

8.7.5 The planarity of the conjugated system

Table 8.6.3 shows that the conjugated system as a whole (in the table referred to as plane 1) is not planar at all. The planes through the individual double bonds have significant deviations from planarity too. From the Newman projections (fig. 8.6.5 e, f, g, h and i) the torsions about the individual bonds are calculated as C(4)=C(5) : 6° clock wise, C(5)–C(6) : $6\frac{1}{2}^\circ$ counter clock wise, C(6)=C(7) : 7° c., C(7)–C(8) : 20° c., C(8)=C(9) : $3\frac{1}{2}^\circ$ c.

So the overall torsion of the conjugated system is clockwise, going from head to tail.

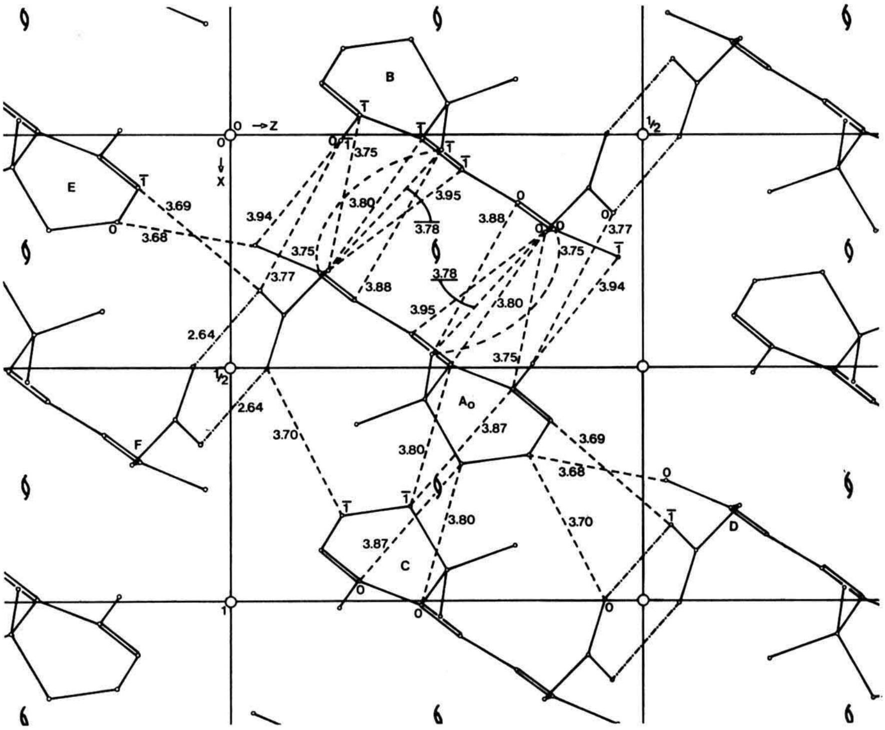


Fig. 8.6.6 a

Fig. 8.6.6. The crystal structure of *retro*-β-ionylidene acetic acid viewed along the y-axis (fig. 8.6.6.a) and along the x-axis (fig. 8.6.6.b). Some of the intermolecular distances <4.0 Å between molecule A_0 and adjacent molecules are given in fig. a and all others in fig. b.

The torsion angles of the single bonds, mentioned above, need not have the same values as the angles between the different planes of the molecule, because these planes include all atoms of the individual double bonds. The torsion angles however give a better picture of the deviation from planarity in the conjugated system.

8.7.6 The conformation of the methyl groups.

The attachments of the methyl groups are given in the Newman projections 8.6.5 l, m, n and o.

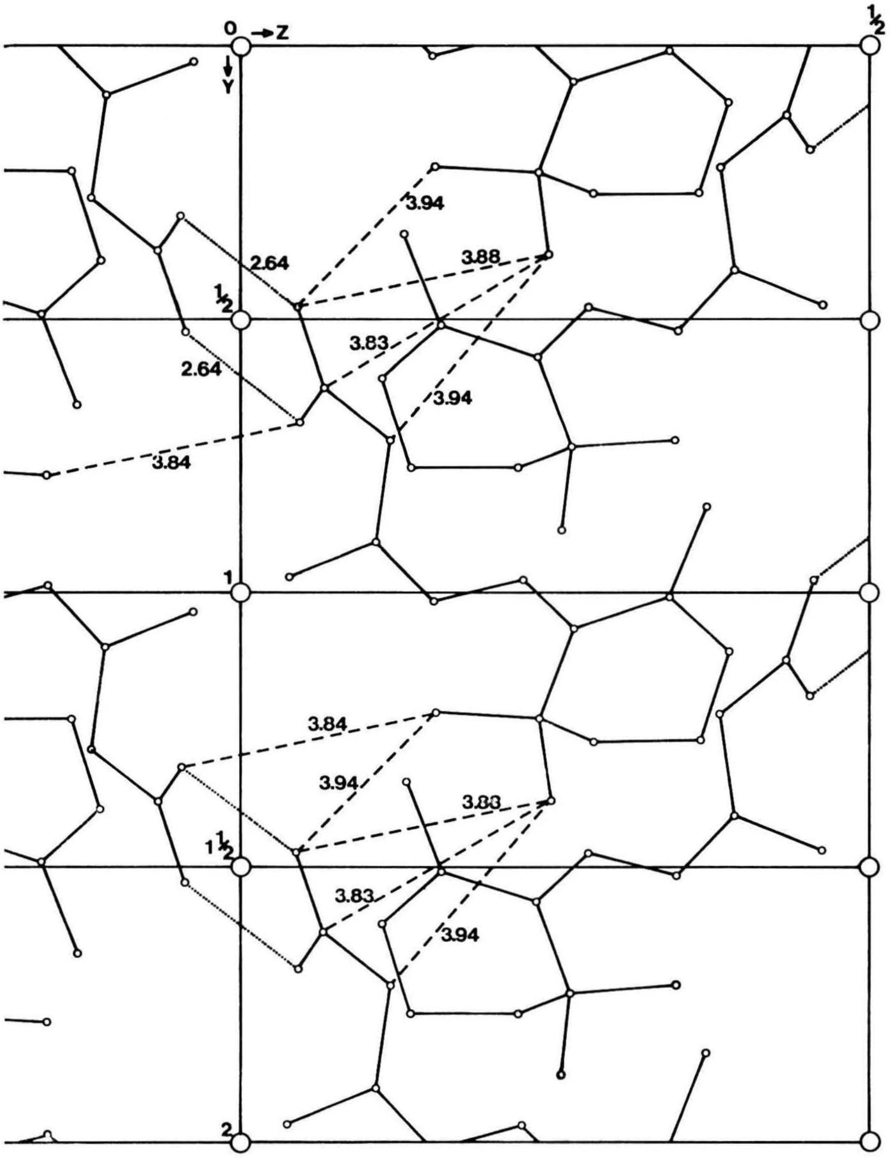


Fig. 8.6.6 b

In the methyl groups C(12) (fig. 8.6.5 l) and C(15) (fig. 8.6.5 o) a hydrogen atom eclipses the double bond fairly well, as also found in propene (HERSCHBACH and KRISHER, 1958).

Methyl group C(14) is attached in the staggered conformation as shown in figure 8.6.5. Methyl group C(13) is rotated by about 7° out of the fully staggered conformation. This is caused by the steric hindrance between H(21) and H(6). The distance H(21)–H(6) of 2.04 Å is the shortest non-nearest-neighbours hydrogen-hydrogen distance (see fig. 8.6.4).

8.7.7 The attachment of the COOH-group to the conjugated chain

C(10) forms the connection between the conjugated chain and the carboxylic group. H(5) of this methylene group is rotated by 15° about C(9)–C(10) out of the eclipsed position with respect to the double bond C(8)–C(9) (fig. 8.6.5 j). This is caused by steric interaction of H(5) with H(7), the distance between the atoms being 2.06 Å (fig. 8.6.4).

The position of plane of the COOH-group is fixed by the steric interaction of C(9) with O(2) (fig. 8.6.3) and the interaction of H(4) with O(1) (fig. 8.6.4) in such a way that the dihedral angle between O(2) and C(9) along C(10)–C(11) is 35° (fig. 8.6.5 k).

The angle between the plane of the COOH-group and the plane of the double bond C(8)–C(9) is 88°. The angle between the carboxylic group and the overall plane of the conjugated system is 78° (table 8.6.3).

8.7.8 Packing

The packing of the structure is illustrated in the figures 8.6.6 a and b.

Molecules related by centres of symmetry $\frac{1}{2}\,\frac{1}{2}\,0$ and $0\,0\,\frac{1}{2}$ respectively are connected by hydrogen bonds of 2.64 Å between their carboxylic groups to "double molecules".

Molecule A_0 is packed at the positive x-side between the molecules C_{-1} and C_0. These molecules are generated from A_0 by the screw axis $x = 3/4$, $z = 1/4$. Only ring to ring contacts occur. (see fig. 8.6.6 a).

At the negative x-side A_0 is packed between the molecules B_{-1} and B_0, generated from A_0 by the screw axis $x = 1/4$, $z = 1/4$. The ring of A_0 has many short contacts (3.7–3.8 Å) with the chain of B_0 and so does the chain of A_0 with the ring of B_{-1} (fig. 8.6.6 a).

The other half of the "double molecule" A_0–F_{-1}, the molecule F_{-1}, has the same contacts as A_0 with the molecules, generated from C_{-1}, C_0, B_{-1} and B_0 by the centre of symmetry $\frac{1}{2}\,\frac{1}{2}\,0$.

Molecule A_0 further has some short contacts (3.7 Å) with the molecules D_0, D_{-1}, E_{-1} and E_0 and some contacts of 3.8–3.9 Å with the molecules F_0 and A_1.

REFERENCES

BART, J. C. J., Thesis, Amsterdam (1967).

————, and C. H. MACGILLAVRY, Acta Cryst. **B24**, 1569 (1968a).

————, and ————, Acta Cryst. **B24**, 1587 (1968b).

BEURSKENS, P. T., Thesis, Utrecht (1965).

CRUICKSHANK, D. W. J., Acta Cryst. **2**, 65 (1949).

————, Computing Methods and the Phase Problem in X-ray Crystal Analysis, p. 45, 70 Oxford, Pergamon Press. (1961).

EICHHORN, E. L. and C. H. MACGILLAVRY, Acta Cryst. **12**, 872 (1959).

GERMAIN, G. and M. M. WOOLFSON, Acta Cryst. **B24**, 91 (1968).

HAUPTMAN, H. and J. KARLE, Solution of the Phase Problem 1, The centrosymmetric Crystal. A.C.A. Monograph No 3. (1953).

————, and ————, Acta Cryst. **12**, 93 (1959).

HERSCHBACH, D. R. and L. C. KRISHER, J. Chem. Phys. **28**, 728 (1958).

HUISMAN, H. O., A. SMIT, P. H. VAN LEEUWEN and J. H. VAN RIJ, Rec. Trav. Chim. **75**, 977 (1956).

————, and J. L. BAAS, Private communication. (1968).

KARLE, I. L., H. HAUPTMAN, J. KARLE and A. B. WING, Acta Cryst. **11**, 257 (1958).

KARLE, J. and I. L. KARLE, Acta Cryst. **21**, 849 (1966).

KOCH, B. and C. H. MACGILLAVRY, Program and Abstracts. I.U.C. Summer Meeting, Roma (1963).

PAUL–ROY, S., H. SCHENK and C. H. MACGILLAVRY, To be published. (1969).

SCHOMAKER, V., J. WASER, R. E. MARSH and G. BERGMAN, Acta Cryst. **12**, 600 (1959).

SIMPSON, P. G., R. D. DOBROTT, and W. N. LIPSCOMB, Acta Cryst. **18**, 169 (1965).

SLY, W. G., Acta Cryst. **17**, 511 (1964).

STAM, C. H. and C. H. MACGILLAVRY, Acta Cryst. **16**, 62 (1963).

STERLING, C., Acta Cryst. **17**, 1224 (1964).

WOHL, R. A., Chimia **18**, 219 (1964).

SUMMARY

As pointed out in chapter 1 the object of the investigations was twofold:

1. The determination of the crystal structures of some compounds related to vitamin-A.
2. The automation of the Symbolic Addition Method (**Karle** and **Karle**, 1966).

The chapters 2 and 3 describe the Symbolic Addition Method and include a study about the proper conditions for successful application. Computer programs, which enable semi-automatic and automatic Symbolic Addition are discussed in the chapters 4 and 5 respectively.

The crystal structures of 1,1,4-tri-methyl-5-carboxyl-cyclopentene-4, *retro* vitamin-A acid and *retro*-β-ionylidene acetic acid are described in the chapters 6, 7 and 8. The main features of the structures are given in chapter 1 (1.8, 1.9 and 1.11). The structures have been determined with the computer programs discussed in the chapters 4 and 5. Previous attempts to solve the structure of *retro* vitamin-A acid and *retro*-β-ionylidene acetic acid by means of the Patterson Method were unsuccessful.

ACKNOWLEDGEMENTS

First of all I wish to thank Prof. Dr. C. H. MacGillavry, whose interest and knowledge were very important in the investigations presented in this report. I would also like to thank Dr. C. H. Stam for many helpful discussions about the investigations and the resulting manuscript.

I am indebted to Prof. Dr. J. M. Bijvoet, Prof. Dr. E. Havinga, Prof. Dr. H. O. Huisman and Dr. C. Romers for their critical reading of the manuscript.

To Mr. A Kreuger I express my gratitude for carrying out most of the experimental work.

Further I wish to record the help of Miss J. Zeeman and my wife with the typing and correcting of the manuscript and the help of Mr. B. Moleman with the computations and drawings.

The calculations were carried out on the Electrologica X1 and X8 computers of the Mathematical Centre, Amsterdam. The cooperation with the computing group of the Centre is gratefully acknowledged.