

Structure of the Tannin Geraniin Based on Conventional X-ray Data at 295 K and on Synchrotron Data at 293 and 120 K

PETER LUGER,^{a,*} MANUELA WEBER,^a SETSUO KASHINO,^b YOSHIKI AMAKURA,^c TAKASHI YOSHIDA,^c TAKUO OKUDA,^c GEZINA BEURSKENS^d AND ZBIGNIEW DAUTER^e

^aInstitut für Kristallographie, Fachbereich Chemie, Freie Universität Berlin, Takustrasse 6, 14195 Berlin, Germany,

^bDepartment of Chemistry, Faculty of Science, Okayama University, Okayama 700-8530, Japan, ^cFaculty of Pharmaceutical Sciences, Okayama University, Okayama 700-8530, Japan, ^dLaboratory of Crystallography, University of Nijmegen, The Netherlands, and ^eEMBL Outstation c/o DESY Hamburg, Notkestrasse 85, 22603 Hamburg, Germany. E-mail: luger@chemie.fu-berlin.de

(Received 30 April 1997; accepted 6 January 1998)

Abstract

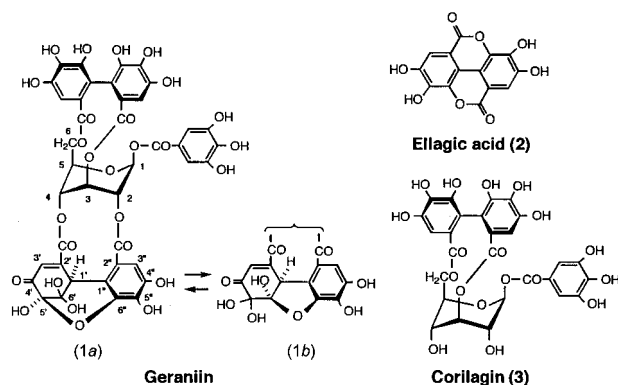
Geraniin, C₄₁H₂₈O₂₇·7H₂O, is the main tannin from *Geranium thunbergii* Sieb. et Zucc., which is one of the most popular folk medicines and also an official antidiarrheic drug in Japan. Conventional X-ray diffractometer data at 295 K and synchrotron data at 293 and 120 K (at EMBL/DESY Hamburg with a MARRESEARCH imaging plate) were measured. The structure could not be determined with any of the currently distributed direct-methods programs, but was easily solved with *DIRDIF* using several conformers of the hexahydroxydiphenyl group as input fragments into a vector-search procedure. A molecular structure was established where a cyclohexenetrione moiety attached to O4 of a central glucose unit was in a hydrated six-membered hemiacetal ring structure, as expected from the NMR study of crystalline geraniin. Owing to the high oligocyclic substitution the glucopyranosyl ring is in the unusual ¹C₄ conformation. Refinement of the low-temperature synchrotron data allowed identification of almost all the H atoms, even within the seven water molecules, so that a rather complex system of approximately 30 different hydrogen bonds can be studied in some detail.

1. Introduction

Geraniin (1) was isolated in 1977 as the main tannin from *Geranium thunbergii* Sieb. et Zucc. in an attempt to obtain the active principle of this medicinal plant, which is one of the most popular folk medicines and also is an official antidiarrhetic drug in Japan (Okuda *et al.*, 1977).

Geraniin gave gallic acid, ellagic acid (2) and corilagin (3) on hydrolysis in boiling water and was shown to be a corilagin derivative esterified at O2 and O4 of D-glucopyranose. Chemical and NMR evidence showed that the chemical structure in the crystal was (1a) with a dehydrohexahydroxydiphenyl (DHHDP) group forming a

hydrated six-membered hemiacetal ring and that the molecule exists as an equilibrium mixture of (1a) and (1b) in aqueous solution (Okuda *et al.*, 1982). Geraniin has been regarded as a key compound among a large number of ellagitannins with a ¹C₄ glucopyranose core isolated from various medicinal plants since 1975 (Okuda *et al.*, 1995). Geraniin analogs which have the DHHDP group in the molecule include furosinin, didehydrogeraniin, mallotusinic acid and macarinins and many dimers, such as euphobins A–H. Many other ellagitannins (chebulagic acid, repandusinic acid, phyllanthusiins A–C and elaeocarpusin *etc.*) are also considered to be the metabolites capable of being produced by structural modifications at the DHHDP unit of geraniin (Okuda *et al.*, 1995). It is also of interest that ellagic acid liberated upon hydrolysis of geraniin and other ellagitannins has antitumor activity (Mukhtar *et al.*, 1984; Sayer *et al.*, 1982; Wood *et al.*, 1982).



X-ray analysis was performed to confirm the stereo-structure, to establish the spatial molecular geometry and to study intra- and intermolecular hydrogen-bonding interactions of the crystalline heptahydrate of (1). The present study is the first X-ray crystallographic elucidation of an ellagitannin, which should provide some 'hard' structural data on this class of compounds,

Table 1. *Experimental details*

	GERAJ	GERAH	GERAT
Crystal data			
Chemical formula	C ₄₁ H ₂₈ O ₂₇ ·7H ₂ O	C ₄₁ H ₂₈ O ₂₇ ·7H ₂ O	C ₄₁ H ₂₈ O ₂₇ ·7H ₂ O
Chemical formula weight	1078.76	1078.76	1078.76
Cell setting	Orthorhombic	Orthorhombic	Orthorhombic
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
<i>a</i> (Å)	13.055 (2)	13.000 (4)	12.960 (2)
<i>b</i> (Å)	29.139 (4)	29.100 (4)	29.110 (2)
<i>c</i> (Å)	11.5760 (10)	11.600 (4)	11.510 (2)
<i>V</i> (Å ³)	4403.7 (9)	4388 (2)	4342.3 (11)
<i>Z</i>	4	4	4
<i>D</i> _x (Mg m ⁻³)	1.627	1.633	1.650
<i>D</i> _m (Mg m ⁻³)	1.62	—	—
Density measured by	Flotation in aqueous NaI	—	—
Radiation type	Cu <i>K</i> α	Synchrotron	Synchrotron
Wavelength (Å)	1.5418	0.86	0.928
No. of reflections for cell parameters	20	See text, §2	See text, §2
θ range (°)	5–14	See text, §2	See text, §2
μ (mm ⁻¹)	1.22	0.146	0.148
Temperature (K)	295	293 (2)	120 (2)
Crystal form	Prismatic	Prismatic	Prismatic
Crystal size (mm)	0.44 × 0.15 × 0.06	0.4 × 0.1 × 0.04	0.4 × 0.1 × 0.04
Crystal color	Yellow	Yellow	Yellow
Data collection			
Diffractometer	Rigaku AFC-5	Imaging Plate	Imaging Plate
Data collection method	$\omega/2\theta$ scans	φ scans	φ scans
Absorption correction	None	None	None
No. of measured reflections	4326	23617	29 510
No. of independent reflections	3894	4027	2699
No. of observed reflections	3408	3598	2638
Criterion for observed reflections	$F > 3\sigma(F)$	$F > 2\sigma(F)$	$F > 2\sigma(F)$
<i>R</i> _{int}	0.024	0.059	0.030
θ_{\max} (°)	60	30.33	27.04
Range of <i>h, k, l</i>	0 → <i>h</i> → 14 0 → <i>k</i> → 32 -1 → <i>l</i> → 13	0 → <i>h</i> → 15 0 → <i>k</i> → 34 0 → <i>l</i> → 13	0 → <i>h</i> → 12 0 → <i>k</i> → 30 0 → <i>l</i> → 11
No. of standard reflections	3	—	—
Frequency of standard reflections (min)	Every 97 reflections	—	—
Intensity decay (%)	1	—	—
Refinement			
Refinement on	<i>F</i>	<i>F</i> ²	<i>F</i> ²
$R[F^2 > 2\sigma(F^2)]$	0.046	0.0557	0.0379
$wR(F^2)$	0.064	0.1535	0.0978
<i>S</i>	1.25	1.074	1.032
No. of reflections used in refinement	3408	4027	2682
No. of parameters used	793	691	691
H-atom treatment	All H-atom parameters refined	H-atom parameters not refined	H-atom parameters not refined
Weighting scheme	$w = 1/[\sigma^2(F_o) - 0.0224F_o + 0.0027F_o^2]$	$w = 1/[\sigma^2(F_o^2) + (0.1263P)^2 + 0.6630P]$, where $P = (F_o^2 + 2F_c^2)/3$	$w = 1/[\sigma^2(F_o^2) + (0.0569P)^2 + 6.5145P]$, where $P = (F_o^2 + 2F_c^2)/3$
$(\Delta/\sigma)_{\max}$	0.56	0.367	-1.053
$\Delta\rho_{\max}$ (e Å ⁻³)	0.24	0.551	0.350
$\Delta\rho_{\min}$ (e Å ⁻³)	-0.44	-0.400	-0.298
Extinction method	None	<i>SHELXL93</i> (Sheldrick, 1993)	<i>SHELXL93</i> (Sheldrick, 1993)
Extinction coefficient	—	0.0017 (8)	0.0022 (4)
Source of atomic scattering factors	<i>International Tables for X-ray Crystallography</i> (1974, Vol. IV)	<i>International Tables for Crystallography</i> (1992, Vol. C, Tables 4.2.6.8 and 6.1.1.4)	<i>International Tables for Crystallography</i> (1992, Vol. C, Tables 4.2.6.8 and 6.1.1.4)
Computer programs			
Data collection	<i>AFC</i> (Rigaku Co., 1990)	Imaging Plate	Imaging Plate
Cell refinement	<i>AFC</i> (Rigaku Co., 1990)	Imaging Plate	Imaging Plate

Table 1 (*cont.*)

	GERAJ	GERAH	GERAT
Data reduction	TEXSAN (MSC, 1985)	Imaging Plate	Imaging Plate
Structure solution	Not given	SHELXS86 (Sheldrick, 1990)	SHELXS86 (Sheldrick, 1990)
Structure refinement	HBLS (Ashida, 1973)	SHELXL93 (Sheldrick, 1993)	SHELXL93 (Sheldrick, 1993)
Preparation of material for publication	DAPH (Ashida, 1973)	SHELXL93 (Sheldrick, 1993)	SHELXL93 (Sheldrick, 1993)

the absence of which has recently been noted in the review article of Quideau & Feldman (1996).

2. X-ray experiments

Crystals of (1) were grown by slow evaporation from an acetonitrile/water solution. Three single-crystal X-ray data sets were measured, see also Table 1. The first data set was collected with conventional Cu $K\alpha$ radiation at room temperature at the Institute for Protein Research, Osaka University, Japan, on a Rigaku AFC-5 four-circle diffractometer (data set GERAJ). Two further data collections were executed with synchrotron radiation at EMBL/DESY (Hamburg) using a MARRESEARCH imaging plate. The synchrotron experiments have been performed at two different beam lines, the room-temperature data GERAH at the wiggler line BW7A equipped with a 180 mm MAR imaging plate scanner and the 120 K data GERAT at the line X11 with a 300 mm MAR scanner. The different geometrical setups and wavelengths used resulted in different resolutions being achieved, 0.85 Å for GERAH and 1.02 Å for GERAT, and therefore a different number of independent reflections. The cell dimensions in the synchrotron experiments were estimated from the postrefinement procedure during data merging. They are strongly correlated to the wavelength and crystal-to-detector distance, which may not be known with very high absolute accuracy. Therefore, the accuracy of the cell dimensions estimated from these experiments may have larger systematic errors than those from the diffractometric experiment. The most relevant crystallographic data are given in Table 1.

3. Structure solution with DIRDIF

The structure could not be determined with any of the currently distributed direct-methods programs, but was easily solved using vector-search techniques. Although there is no obvious large and rigid fragment in this rather flexible molecule, a number of suitable 20-atom fragments were generated by stepwise rotation around the central bond in an idealized hexahydroxydiphenoyl group. Owing to possible strain in the macrocycle containing the search group, and because of the symmetry of the group, the rotation in steps of 15° was cut down to obtain a total of five different conformers of the search group. A rotation of 15° causes a displace-

ment of at most 0.6 Å for some of the atoms, so that one of the generated models should fit the fragment in the structure with a maximum error of 0.3 Å (and an expected average deviation of less than 0.1 Å).

Each of these conformers was input (one at the time) to the *DIRDIF* program system (Beurskens, Beurskens *et al.*, 1992), which implies a rotation scan, using vector-search methods (*ORIENT*; Beurskens *et al.*, 1987), automatically followed by positioning using reciprocal correlation functions (*TRACOR*; Beurskens, Gould *et al.*, 1987), phase expansion and refinement (*PHASEX*; Beurskens & Smykalla, 1991), and further expansion of the partial structure. One of the conformers indeed led to the complete structure, thus a 20-atom fragment was sufficient to solve this 75-atom (non-H) problem.

4. Refinements

Both synchrotron data sets (GERAH and GERAT) were refined (based on F^2) with *SHELXL93* (Sheldrick, 1993). In addition to the title molecule the presence of

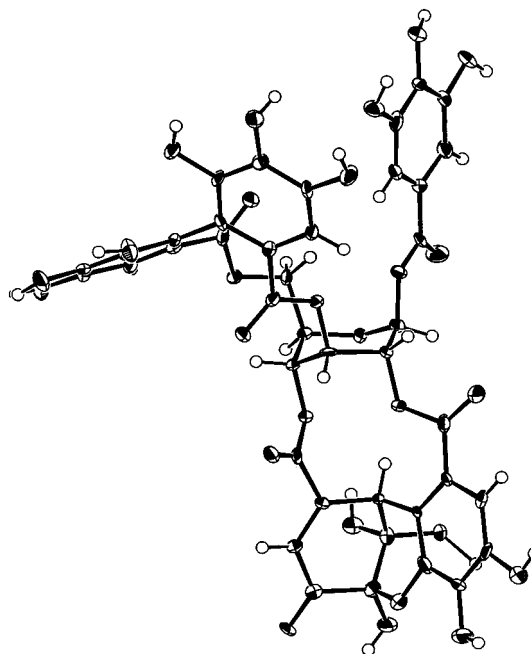


Fig. 1. ORTEPII (Johnson, 1976) plot of geraniin (GERAT, 120 K); the thermal ellipsoids are drawn at a probability of 50%.

seven water molecules could clearly be confirmed. At room temperature some of the water O atoms tended to have large anisotropic displacement parameters (ADP's), so that a disorder for two of them (O6W and O7W) was considered; however, in the 120 K data set GERAT, disorder was no longer observed and all atoms including their ADP's refined well to convergence.

The GERAT data set also allowed H atoms to be determined from difference syntheses. All H atoms of the geraniin main molecule and those at the water O atoms O1W and O2W could be located without problems. However, at the water O atoms O3W–O7W the interpretation of the difference map was not clear-cut. It was decided to accept these peaks as possible water H atoms that could fit into a hydrogen bond with reasonable O···O contacts. All C–H H atoms were refined with the 'riding model' option, but no attempt was made to refine hydroxyl hydrogen positions. For the GERAH refinement the H atoms from the GERAT data were used.

The GERAJ structure was refined starting with the low-temperature synchrotron model by block-diagonal least-squares using *HBL5-V* (Ashida, 1973). Non-H atoms were refined anisotropically. Out of the 42 H atoms 29 were refined isotropically. The other 13 H atoms were excluded from the calculation at the final stage of the refinement, because their displacement parameters became extraordinarily large and geometries involving them became unreasonable during the course of refinements. The calculations were performed on a NEAC ACOS 3700 computer at the Okayama University Computer Center.

The 120 K molecular structure obtained after convergence of the refinements (see also Table 1) is displayed in the *ORTEP*II (Johnson, 1976) representations in Fig. 1.† The GERAJ and GERAH data sets give very comparable results for positional parameters and ADP's. For comparison with the low-temperature GERAT data, see also Fig. 2, which shows a plot of the U_{eq} values versus the corresponding atomic distances from the molecular mass center. The second-order least-squares curves of the U_{eq} distributions are rather close together for the room-temperature data sets GERAJ and GERAH, but differ uniformly by 0.008 \AA^2 over the whole range. The curves also show that the U_{eq} values vary from around 0.03 \AA^2 for the atoms close to the molecular center to $\sim 0.08 \text{ \AA}^2$ for the peripheral ones. At 120 K, for the GERAT data, thermal motion is remarkably reduced with U_{eq} values in the range $0.02\text{--}0.03 \text{ \AA}^2$.

If the conformity of bonding geometry for the three data sets is considered it was found that for the two room-temperature data sets corresponding non-H bond

lengths differ by more than 3σ in 13 cases (out of a total of 78 bonds), but more than 4σ in only two cases with the worst disagreement for C32–C37, 1.388 (6) and 1.419 (6) Å. If, however, the averages of the GERAJ and GERAH data are compared with the low-temperature GERAT bond lengths then only in three cases is the discrepancy greater than 3σ , and none exceed 4σ . Moreover, the expected tendency is visible in that a number of bonds, especially at terminal C=O or C–OH groups, tend to be longer and hence less affected by thermal motion for the low-temperature data set compared with room temperature.

For the following discussion of results the low-temperature structure GERAT will only be considered. Its atomic parameters are given in Table 2, and selected bond lengths and angles are in Table 3; the corresponding room-temperature quantities have been deposited.†

5. Results and discussion

The molecular structure including the seven water molecules (with part of the atomic numbering) is shown in Fig. 3. The geraniin molecule consists of four major fragments:

- the central β -D-glucopyranosyl ring;
- the dehydrohexahydroxydiphenoyl group (DHHDP) linked to O2 and O4 of the glucose;
- the hexahydroxydiphenoyl group (HHDP) attached to O3 and O6;
- the gallic acid residue at O1.

The β -D-glucopyranosyl ring is in the unusual 1C_4 conformation with all substituents in axial positions, which is otherwise only found in glucose molecules where the pyranosyl is involved in oligocyclic small ring systems as, for example, in anhydro ring closures. This

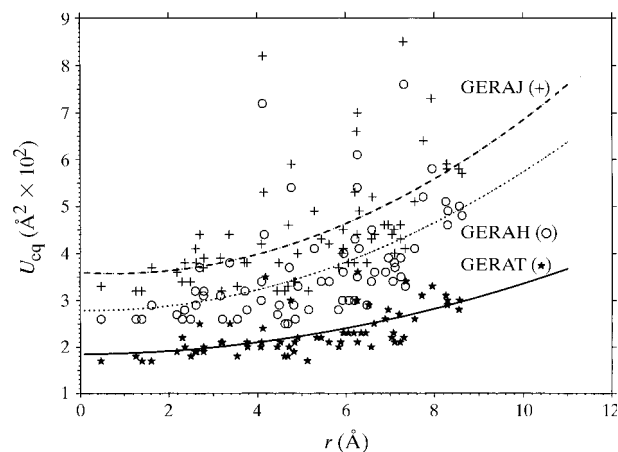


Fig. 2. Distribution of U_{eq} values versus atomic distances r from the molecular mass center for the three data sets. For definition of the data sets GERAJ, GERAH and GERAT, see §2 and Table 1.

† Supplementary data for this paper are available from the IUCr electronic archives (Reference: OH0063). Services for accessing these data are described at the back of the journal.

Table 2. Fractional atomic coordinates and equivalent isotropic displacement parameters (\AA^2)
$$U_{\text{eq}} = (1/3)\sum_i \sum_j U^{ij} a^i a^j \mathbf{a}_i \cdot \mathbf{a}_j$$

	x	y	z	U_{eq}
C1	0.6485 (5)	0.2733 (2)	0.0569 (5)	0.0195 (15)
O1	0.6415 (3)	0.24697 (11)	-0.0481 (3)	0.0193 (9)
C2	0.5427 (5)	0.2692 (2)	0.1104 (5)	0.0174 (14)
O2	0.5283 (3)	0.30431 (11)	0.1978 (3)	0.0194 (10)
C3	0.5236 (5)	0.2247 (2)	0.1777 (5)	0.0173 (14)
O3	0.5068 (3)	0.18881 (11)	0.0935 (3)	0.0178 (9)
C4	0.6164 (5)	0.2129 (2)	0.2535 (5)	0.0167 (14)
O4	0.6281 (3)	0.24951 (11)	0.3396 (3)	0.0181 (9)
C5	0.7171 (5)	0.2148 (2)	0.1896 (5)	0.0193 (14)
O5	0.7299 (3)	0.25807 (11)	0.1289 (3)	0.0185 (9)
C6	0.7374 (5)	0.1748 (2)	0.1052 (5)	0.0199 (15)
O6	0.7070 (3)	0.13320 (11)	0.1643 (3)	0.0206 (10)
C11	0.7169 (5)	0.2535 (2)	-0.1285 (5)	0.0202 (14)
O11	0.7881 (4)	0.27955 (13)	-0.1146 (4)	0.0299 (11)
C12	0.7008 (5)	0.2243 (2)	-0.2329 (5)	0.0221 (15)
C13	0.7666 (5)	0.2305 (2)	-0.3270 (5)	0.0214 (15)
C14	0.7551 (5)	0.2035 (2)	-0.4244 (5)	0.0208 (15)
O14	0.8164 (3)	0.20637 (13)	-0.5213 (3)	0.0280 (11)
C15	0.6774 (5)	0.1708 (2)	-0.4299 (5)	0.0210 (15)
O15	0.6650 (4)	0.14307 (13)	-0.5247 (3)	0.0286 (11)
C16	0.6085 (5)	0.1663 (2)	-0.3380 (5)	0.023 (2)
O16	0.5264 (4)	0.13634 (12)	-0.3437 (3)	0.0296 (11)
C17	0.6202 (5)	0.1923 (2)	-0.2387 (5)	0.0204 (14)
C21	0.4682 (5)	0.3413 (2)	0.1721 (5)	0.0206 (15)
O21	0.4370 (4)	0.34771 (13)	0.0741 (4)	0.0347 (12)
C22	0.4448 (5)	0.3694 (2)	0.2744 (5)	0.0191 (15)
C23	0.3569 (5)	0.3974 (2)	0.2629 (5)	0.0225 (15)
C24	0.3251 (5)	0.4243 (2)	0.3546 (5)	0.023 (2)
O24	0.2395 (4)	0.45211 (12)	0.3534 (3)	0.0305 (11)
C25	0.3792 (5)	0.4252 (2)	0.4564 (5)	0.0215 (15)
O25	0.3510 (4)	0.45150 (13)	0.5480 (3)	0.0302 (11)
C26	0.4661 (5)	0.3981 (2)	0.4667 (5)	0.0229 (15)
O26	0.5154 (3)	0.40166 (12)	0.5733 (3)	0.0274 (11)
C27	0.5015 (5)	0.3689 (2)	0.3785 (4)	0.0169 (14)
C31	0.4576 (5)	0.1508 (2)	0.1365 (5)	0.022 (2)
O31	0.4431 (3)	0.14514 (12)	0.2387 (3)	0.0249 (11)
C32	0.4246 (5)	0.1199 (2)	0.0414 (5)	0.0180 (14)
C33	0.3668 (5)	0.1377 (2)	-0.0479 (5)	0.0213 (15)
C34	0.3234 (5)	0.1091 (2)	-0.1279 (5)	0.0217 (15)
O34	0.2599 (4)	0.12712 (12)	-0.2113 (4)	0.0358 (12)
C35	0.3443 (5)	0.0623 (2)	-0.1233 (5)	0.0228 (15)
O35	0.2987 (4)	0.03243 (12)	-0.2028 (3)	0.0276 (11)
C36	0.4028 (5)	0.0444 (2)	-0.0352 (5)	0.0204 (15)
O36	0.4217 (3)	-0.00181 (11)	-0.0318 (3)	0.0278 (11)
C37	0.4455 (5)	0.0724 (2)	0.0507 (5)	0.0179 (14)
C41	0.5540 (5)	0.2562 (2)	0.4191 (5)	0.0206 (15)
O41	0.4834 (4)	0.23034 (12)	0.4396 (3)	0.0248 (11)
C42	0.5736 (5)	0.3006 (2)	0.4812 (5)	0.0199 (15)
C43	0.5695 (5)	0.3013 (2)	0.5965 (5)	0.0211 (15)
C44	0.6035 (6)	0.3414 (2)	0.6641 (5)	0.026 (2)
O44	0.6143 (4)	0.34185 (12)	0.7685 (3)	0.0326 (12)
C45	0.6193 (5)	0.3850 (2)	0.5908 (5)	0.022 (2)
O45	0.6740 (4)	0.41913 (12)	0.6472 (3)	0.0304 (11)
C46	0.6707 (5)	0.3739 (2)	0.4766 (5)	0.0225 (15)
O461	0.7637 (4)	0.35148 (12)	0.5063 (3)	0.0297 (11)
O462	0.6897 (3)	0.41307 (12)	0.4088 (3)	0.0250 (11)
C47	0.5990 (5)	0.3424 (2)	0.4083 (5)	0.0183 (15)
C61	0.6624 (5)	0.1002 (2)	0.0959 (5)	0.0201 (14)
O61	0.6633 (3)	0.10252 (12)	-0.0079 (3)	0.0238 (10)
C62	0.6121 (5)	0.0649 (2)	0.1692 (5)	0.021 (2)
C63	0.6658 (5)	0.0461 (2)	0.2596 (5)	0.022 (2)
C64	0.6194 (5)	0.0128 (2)	0.3299 (5)	0.023 (2)
O64	0.6814 (4)	-0.00610 (13)	0.4153 (4)	0.0336 (12)
C65	0.5182 (5)	0.0007 (2)	0.3110 (5)	0.0197 (15)

Table 2 (cont.)

	x	y	z	U_{eq}
O65	0.4614 (4)	-0.02944 (13)	0.3785 (3)	0.0307 (11)
O66	0.3618 (4)	0.01118 (13)	0.2052 (3)	0.0293 (11)
C66	0.4635 (5)	0.0209 (2)	0.2210 (5)	0.021 (2)
C67	0.5087 (5)	0.0524 (2)	0.1455 (5)	0.021 (2)
O1W	0.8542 (4)	0.29544 (14)	0.3409 (4)	0.0408 (13)
O2W	0.4468 (4)	0.10015 (13)	-0.5412 (4)	0.0331 (11)
O3W	0.0896 (4)	0.43940 (14)	0.2051 (4)	0.0365 (12)
O4W	0.6945 (4)	0.4334 (2)	0.8763 (4)	0.0479 (13)
O5W	0.4664 (5)	0.4683 (2)	0.7578 (4)	0.055 (2)
O6W	0.5818 (4)	-0.06796 (13)	0.5409 (4)	0.0460 (14)
O7W	0.3568 (5)	0.0272 (2)	-0.4272 (4)	0.059 (2)

Table 3. Selected geometric parameters (\AA , $^\circ$)

C1—O5	1.413 (7)	C23—C24	1.377 (8)
C1—O1	1.435 (6)	C24—C25	1.366 (8)
C1—C2	1.509 (9)	C24—O24	1.373 (7)
O1—C11	1.359 (7)	C25—O25	1.353 (7)
C2—O2	1.447 (6)	C25—C26	1.380 (9)
C2—C3	1.529 (7)	C26—O26	1.388 (7)
O2—C21	1.362 (7)	C26—C27	1.402 (8)
C3—O3	1.441 (6)	O26—C45	1.445 (8)
C3—C4	1.525 (8)	C27—C47	1.520 (9)
O3—C31	1.371 (7)	C41—O41	1.209 (7)
C4—O4	1.462 (6)	C41—C42	1.497 (8)
C4—C5	1.499 (8)	C42—C43	1.329 (8)
O4—C41	1.341 (7)	C42—C47	1.514 (7)
C5—O5	1.450 (6)	C43—C44	1.472 (8)
C5—C6	1.540 (7)	C44—O44	1.210 (7)
C6—O6	1.443 (6)	C44—C45	1.537 (8)
O6—C61	1.370 (7)	C45—O45	1.382 (7)
C21—O21	1.212 (7)	C45—C46	1.508 (8)
C21—C22	1.465 (8)	C46—O462	1.404 (7)
C22—C27	1.406 (8)	C46—O461	1.412 (7)
C22—C23	1.408 (8)	C46—C47	1.524 (9)
O5—C1—O1	111.9 (4)	C26—C27—C47	114.6 (5)
C1—C2—C3	115.0 (5)	C22—C27—C47	129.2 (5)
C4—C3—C2	110.6 (5)	C43—C42—C47	123.4 (5)
C5—C4—C3	113.5 (5)	C42—C43—C44	121.9 (5)
O5—C5—C4	111.6 (4)	C43—C44—C45	113.9 (5)
O5—C5—C6	109.4 (4)	O26—C45—C46	111.3 (5)
C4—C5—C6	115.5 (5)	O26—C45—C44	103.2 (5)
C1—O5—C5	118.1 (4)	C46—C45—C44	111.1 (4)
C27—C22—C23	120.6 (5)	O462—C46—O461	111.1 (5)
C24—C23—C22	120.0 (6)	O462—C46—C45	112.8 (4)
C25—C24—C23	121.0 (6)	O461—C46—C45	105.4 (5)
C24—C25—C26	118.8 (5)	O462—C46—C47	108.0 (5)
C25—C26—O26	114.1 (5)	O461—C46—C47	111.5 (4)
C25—C26—C27	123.5 (5)	C45—C46—C47	108.0 (5)
O26—C26—C27	122.3 (6)	C42—C47—C27	110.6 (5)
C26—O26—C45	121.8 (5)	C42—C47—C46	109.2 (5)
C26—C27—C22	116.1 (5)	C27—C47—C46	108.6 (4)

then results in rather strained ring systems which cause more or less remarkable distortions of the pyranosyl chair (Jeffrey & Yates, 1979; Strumpel *et al.*, 1984). In the geraniin pyranosyl ring the Cremer–Pople puckering parameters (Table 4; Cremer & Pople, 1975; Bülow & Luger, 1983), especially the θ value of $\sim 170^\circ$, indicate a slightly distorted 1C_4 chair, however, the puckering amplitude $Q = 0.474$ and the small endocyclic torsion angles along the ring bonds O5—C1, C1—C2 *etc.*,

Table 4. Cremer–Pople Puckering parameters of the nonplanar rings

Ring	Pyranose	Hemiacetal	Cyclohexene
Q (Å)	0.474	0.568	0.510
θ (°)	170.1	118.3	45.6
φ (°)	70	340	229
Conformation (out-of-plane)†	1C_4	H (C46,C47) $\longleftrightarrow E$ (C46)	H (C46,C45) $\longleftrightarrow E$ (C46)

The closest ideal conformations (θ , φ) are: hemiacetal: H (129.23, 330), E (125.26, 0); cyclohexene: H (50.77, 210), E (54.74, 230). † The conformational symbols indicate the six-membered ring forms C = chair, H = half chair, E = envelope.

45.2 (6), -41.6 (6), 43.6 (6), -49.2 (6), 51.8 (6), -50.3 (6)°, indicate a remarkable flattening of the ring, which tolerates more the 1,3-diaxial interactions of the axial substituents. Normally aldopyranosyl rings have torsion angles between 55 and 60°. In the pyranosyl ring the expected shortening of the O5—C1 bond [1.413 (7) Å] compared with O5—C5 [1.450 (6) Å] is visible. A further indication of the ring flattening is the large endocyclic bond angles, especially at O5, C1 and C2. Averages over the X-ray structures of the nine unsubstituted aldohexoses investigated so far give 113.6° for the angle at O5 and $\sim 110^\circ$ for the endocyclic

Table 5. Intra- and intermolecular hydrogen bonds (GERAT, 120 K)

$D-H \cdots A$	$D \cdots A$	$D-H$	$H \cdots A$	$D-H \cdots A$
O14—H14 \cdots O41 ⁱ	2.993 (6)	0.825 (4)	2.243 (4)	151.5 (3)
O15—H15 \cdots O14 ⁱⁱ	2.692 (6)	0.820 (4)	2.283 (4)	111.3 (3)
O16—H16 \cdots O2W ⁱⁱ	2.710 (6)	0.821 (4)	1.935 (4)	157.1 (3)
O24—H24 \cdots O3W ⁱⁱⁱ	2.612 (6)	0.815 (4)	1.886 (5)	147.8 (3)
O25—H25 \cdots O64 ⁱⁱⁱ	2.745 (6)	0.811 (4)	2.022 (4)	148.5 (3)
O34—H34 \cdots O462 ⁱⁱⁱ	2.713 (5)	0.816 (4)	2.047 (4)	138.5 (3)
O34—H34 \cdots O35 ⁱⁱ	2.803 (6)	0.816 (4)	2.385 (4)	112.7 (3)
O35—H35 \cdots O7W ⁱⁱ	2.695 (6)	0.814 (4)	1.921 (5)	158.4 (3)
O36—H36 \cdots O35 ⁱⁱ	2.723 (5)	0.829 (4)	2.252 (4)	116.2 (3)
O36—H36 \cdots O45 ^{iv}	2.932 (5)	0.829 (4)	2.224 (4)	143.6 (3)
O45—H45 \cdots O4W ⁱⁱ	2.682 (6)	0.816 (4)	1.930 (5)	153.0 (3)
O45—H45 \cdots O44 ⁱⁱ	2.758 (5)	0.816 (4)	2.357 (4)	111.1 (3)
O461—H461 \cdots O1W ⁱⁱ	2.769 (6)	0.824 (4)	1.956 (4)	168.2 (3)
O462—H462 \cdots O36 ^v	3.199 (5)	1.495 (4)	1.998 (4)	132.1 (2)
O64—H64 \cdots O6W ⁱⁱ	2.645 (6)	0.822 (4)	1.824 (4)	176.3 (3)
O65—H65 \cdots O6W ⁱⁱ	2.681 (6)	0.820 (4)	1.884 (5)	163.5 (3)
O66—H66 \cdots O35 ^{vi}	2.657 (6)	0.828 (4)	2.062 (4)	128.4 (3)
O66—H66 \cdots O65 ⁱⁱ	2.654 (6)	0.828 (4)	2.221 (4)	112.8 (3)
O1W—H11W \cdots O5 ⁱⁱ	3.119 (6)	0.798 (4)	2.467 (4)	139.8 (3)
O1W—H12W \cdots O16 ⁱ	2.987 (6)	0.999 (5)	2.007 (4)	166.8 (3)
O2W—H21W \cdots O15 ⁱⁱ	3.097 (6)	1.022 (5)	2.405 (5)	124.2 (3)
O2W—H22W \cdots O7W ⁱⁱ	2.756 (6)	0.804 (4)	1.992 (5)	158.6 (3)
O3W—H31W \cdots O61 ^{vii}	2.748 (6)	0.958 (4)	1.867 (4)	151.6 (3)
O3W—H32W \cdots O5W ^{viii}	2.848 (6)	1.245 (4)	2.511 (5)	92.4 (2)
O4W—H42W \cdots O66 ^{ix}	2.861 (7)	1.347 (5)	1.652 (4)	144.8 (3)
O5W—H52W \cdots O25 ⁱⁱ	2.883 (7)	1.079 (5)	1.817 (4)	168.8 (3)
O6W—H61W \cdots O5W ^x	2.622 (6)	0.987 (4)	1.679 (5)	158.5 (3)
O6W—H62W \cdots O21 ^{iv}	2.800 (6)	1.069 (4)	2.143 (4)	117.3 (3)
O7W—H71W \cdots O65 ^{xi}	3.091 (6)	1.159 (5)	1.932 (4)	179.8 (3)
O7W—H72W \cdots O4W ^{xv}	2.870 (6)	1.159 (5)	1.711 (5)	179.8 (2)

Symmetry codes: (i) $\frac{1}{2} + x, \frac{1}{2} - y, -z$; (ii) x, y, z ; (iii) $-\frac{1}{2} + x, \frac{1}{2} - y, 1 - z$; (iv) $1 - x, -\frac{1}{2} + y, \frac{1}{2} - z$; (v) $1 - x, \frac{1}{2} + y, \frac{1}{2} - z$; (vi) $\frac{1}{2} - x, -y, \frac{1}{2} + z$; (vii) $-\frac{1}{2} + x, \frac{1}{2} - y, -z$; (viii) $\frac{1}{2} - x, 1 - y, -\frac{1}{2} + z$; (ix) $\frac{1}{2} + x, \frac{1}{2} - y, 1 - z$; (x) $1 - x, -\frac{1}{2} + y, \frac{1}{2} - z$; (xi) $x, y, -1 + z$.

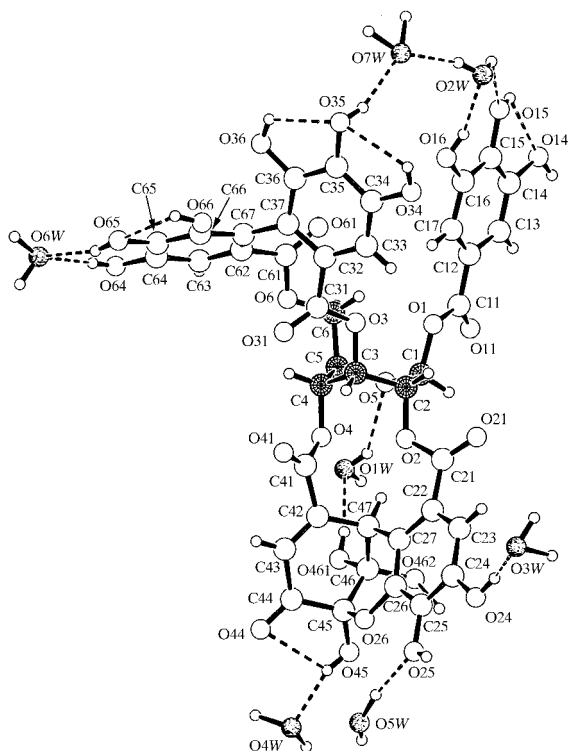


Fig. 3. Molecular structure with the seven water molecules and the atomic numbering scheme (SCHAKAL; Keller, 1989).

O—C—C and C—C—C angles, so that differences of more than 4° are seen here.

In the DHHDP fragment the existence of a six-membered hemiacetal ring [form (1a), see formula scheme] rather than a five-membered ring [form (1b)] is confirmed, as was already deduced from the NMR measurement immediately after dissolving the crystals in acetone- d_6 .

Of the two nonplanar rings of this tricyclic ring system, the cyclohexene ring and the hemiacetal ring are both in conformations between half-chair and envelope.

In the description of an envelope form, the out-of-plane atom is C46 in both cases, in the half-chair description, C46 and C47 are out-of-plane for the hemiacetal ring, and C45 and C46 for the cyclohexene ring; see also the Cremer–Pople parameters in Table 4. The double bond for C42=C43 is confirmed by the length 1.329 (8) Å.

A certain strain in this tricyclic ring system is indicated by some unusual bond angles with C27 as the apex. In the HHDP fragment the two ring planes are inclined to form an angle of 60.6 (3)° between the two least-squares planes. Bond lengths and angles in this

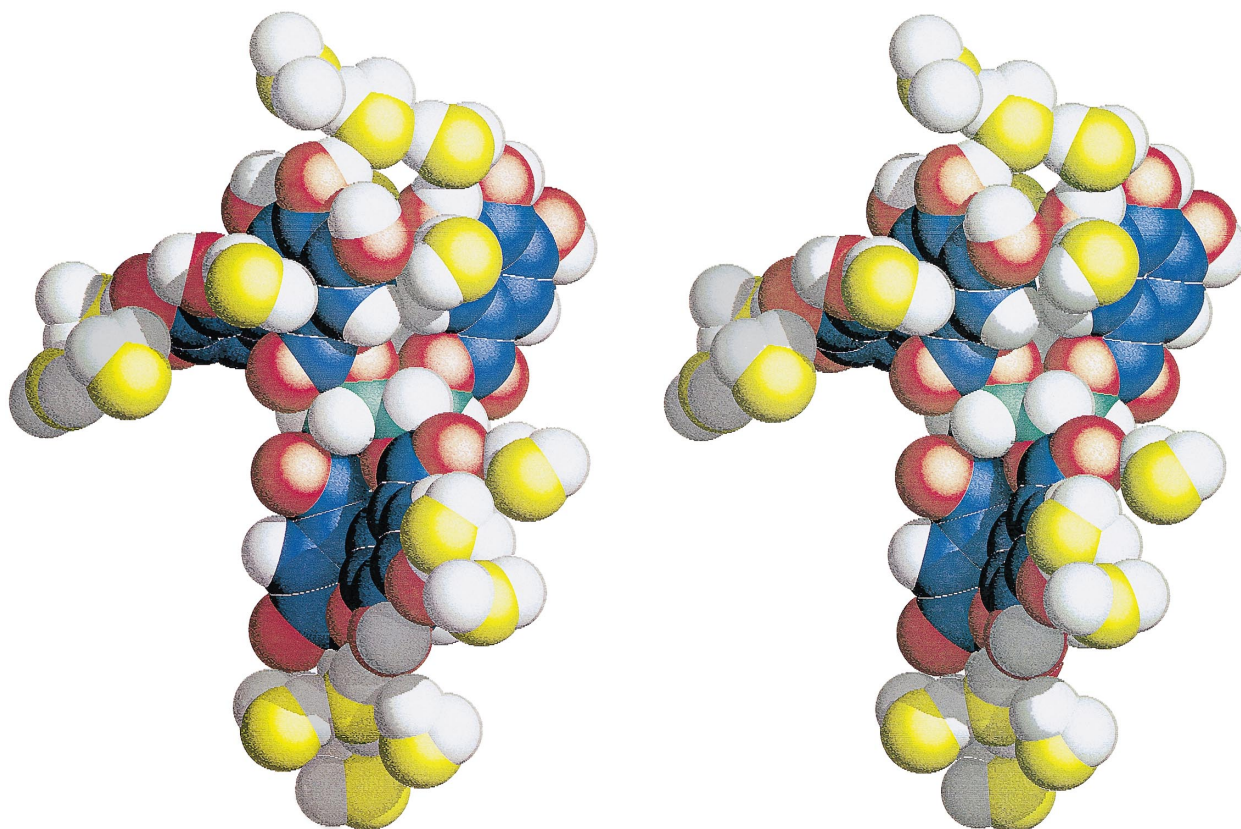


Fig. 4. Space-filling model with the nearest-neighbor water molecules. The atomic color codes are: pyranosyl ring, green; water molecules, yellow; otherwise, carbon, blue; oxygen, red; hydrogen, white (SCHAKAL; Keller, 1989).

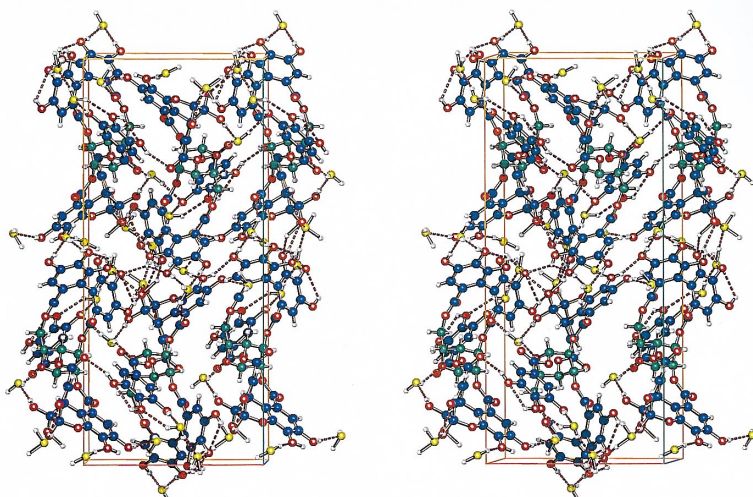


Fig. 5. Projection of the unit cell onto the *ab* plane, with hydrogen bonds indicated by dashed lines. The atomic color code is as described in the caption to Fig. 4. The axis directions are indicated by: *a* red, *b* green and *c* blue (SCHAKAL; Keller, 1989).

fragment and also in the gallic ester are in normal ranges and need no further discussion.

The gallic acid ester is completely planar.

There are links within the molecule through intra- and quasi-intramolecular (indirect) hydrogen bonds (see Fig. 3). Simple intramolecular hydrogen bonds are seen for a

number of vicinal hydroxyl groups, *e.g.* in the trihydroxyphenyl rings. One quasi-intramolecular indirect hydrogen bond connects the OH group of O461 in the DHHDP fragment *via* water molecule O1W with O5 of the pyranosyl ring. Another indirect link *via* the two water molecules O7W and O2W connects the HHDP

fragment with the gallic acid trihydroxyphenyl ring. This linkage closes a deep pocket which is otherwise open between these two fragments (see also Fig. 3).

A space-filling model of geraniin together with *all* water molecules related to the main molecule, either as donors or acceptors of a hydrogen bond, is displayed in Fig. 4.

The overall molecular shape is similar to that of the letter 'T', with most water molecules on the top and at the bottom of the 'T'. Almost no water shell is found on the outer side of the gallic acid ester fragment and the O4...O41...O44 region of the DHHDP group. These regions are linked by an intermolecular hydrogen bond, thus providing no space for water accumulation.

6. Crystal structure

In addition to the above-mentioned intra- and quasi-intramolecular hydrogen bonds a large number of intermolecular hydrogen bonds exist in the crystal lattice, see summary in Table 5. Each OH group of the main molecule acts as (at least!) one donor, the same holds for all but two water H atoms, so that a total of 30 hydrogen bonds were identified, establishing a complex three-dimensional network (see Fig. 5). In the crystal lattice a layer structure of sugars and nonsugars, with the *b* axis as the layer-to-layer axis, is present. The sugar layers are around $y = \frac{1}{4}, \frac{3}{4}, \dots$, while the nonsugar layers are at $y = 0, \frac{1}{2}$ etc.

References

- Ashida, T. (1973). *HBLS-V and DAPH. The Universal Crystallographic Computing System*, pp. 55–65. The Computation Center, Osaka University, Japan.
- Beurskens, P. T., Admiraal, G., Beurskens, G., Bosman, W. P., Garcia-Granda, S., Gould, R. O., Smits, J. M. M. & Smykalla, C. (1992). *DIRDIF92*. Technical Report of the Crystallography Laboratory, University of Nijmegen, The Netherlands.
- Beurskens, P. T., Beurskens, G., Strumpel, M. & Nordman, C. E. (1987). *ORIENT. Patterson and Pattersons*, edited by J. P. Glusker, B. K. Patterson & M. Rossi, pp. 356. Oxford: Clarendon Press.
- Beurskens, P. T., Gould, R. O., Bruins Slot, H. J. & Bosman, W. P. (1987). *Z. Kristallogr.* **179**, 127–159.
- Beurskens, P. T. & Smykalla, C. (1991). *PHASEX. Direct Methods of Solving Crystal Structures*, edited by H. Schenk, p. 281. New York: Plenum Press.
- Bülow, R. & Luger, P. (1983). *J. Appl. Cryst.* **16**, 431–432.
- Cremer, D. & Pople, J. A. (1975). *J. Am. Chem. Soc.* **97**, 1354–1358.
- Jeffrey, D. A. & Yates, J. H. (1979). *Carbohydr. Res.* **74**, 319–322.
- Johnson, C. K. (1976). *ORTEPII*. Report ORNL-5138. Oak Ridge National Laboratory, Tennessee, USA.
- Keller, E. (1989). *J. Appl. Cryst.* **22**, 19–22.
- Molecular Structure Corporation (1985). *TEXSAN. TEXRAY Structure Analysis Package*. MSC, 3200 Research Forest Drive, The Woodlands, TX 77381, USA.
- Mukhtar, H., Del Tito, B. J. Jr, Marcelo, C. L., Das, M. & Bickers, D. R. (1984). *Carcinogenesis*, **5**, 1565–1571.
- Okuda, T., Yoshida, T. & Hatano, T. (1982). *J. Chem. Soc. Perkin Trans. 1*, pp. 9–14.
- Okuda, T., Yoshida, T. & Hatano, T. (1995). *Progress in the Chemistry of Organic Natural Products 66*, edited by W. Herz, G. W. Kirby, R. E. Moore, W. Steglich & Ch. Tamm, pp. 1–117. Wien, New York: Springer-Verlag.
- Okuda, T., Yoshida, T. & Nayeshiro, H. (1977). *Chem. Pharm. Bull.* **25**, 1862–1869.
- Rigaku Corporation (1990). *AFC Software*. Rigaku Corporation, Tokyo, Japan.
- Quideau, S. & Feldman, K. S. (1996). *Chem. Rev.* **96**, 475–503.
- Sayer, J. M., Yagi, H., Wood, A. W., Conney, A. H. & Jerino, D. M. (1982). *J. Am. Chem. Soc.* **104**, 5562–5564.
- Sheldrick, G. M. (1990). *Acta Cryst. A* **46**, 467–473.
- Sheldrick, G. M. (1993). *SHELXL93. Program for the Refinement of Crystal Structures*. University of Göttingen, Germany.
- Strumpel, M., Schmidt, H.-J., Luger, P. & Paulsen, H. (1984). *Carbohydr. Res.* **125**, 185–201.
- Wood, A. W., Huang, M.-T., Chang, R. L., Newmark, H. L., Lehr, R. E., Yagi, H., Sayer, J. M., Jerina, D. M. & Conney, A. H. (1982). *Proc. Natl. Acad. Sci. USA*, **79**, 5513–5517.