



HAL
open science

The occurrence of seven-fold helical molecular conformation in cellulose-phosphoric acid complex

Jia Hui Lim, Isabelle Morfin, Isabelle Jeacomine, Alfred French, Yoshiharu Nishiyama, Yu Ogawa

► **To cite this version:**

Jia Hui Lim, Isabelle Morfin, Isabelle Jeacomine, Alfred French, Yoshiharu Nishiyama, et al.. The occurrence of seven-fold helical molecular conformation in cellulose-phosphoric acid complex. *Cellulose*, 2023, 30 (13), pp.8063-8073. 10.1007/s10570-023-05346-x . hal-04209944

HAL Id: hal-04209944

<https://hal.science/hal-04209944>

Submitted on 18 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **The occurrence of seven-fold helical molecular conformation in cellulose-**
2 **phosphoric acid complex.**

3

4 Jia Hui Lim,¹ Isabelle Morfin,² Isabelle Jeacomine,¹ Alfred D. French,³ Yoshiharu
5 Nishiyama,¹ Yu Ogawa^{1,*}

6

7 1 Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

8 2 Univ. Grenoble Alpes, CNRS, LiPhy, 38000 Grenoble, France

9 3 Southern Regional Research Center, U.S. Department of Agriculture, New
10 Orleans, LA 70124, USA

11 * corresponding author: yu.ogawa@cermav.cnrs.fr

12

13 **Abstract**

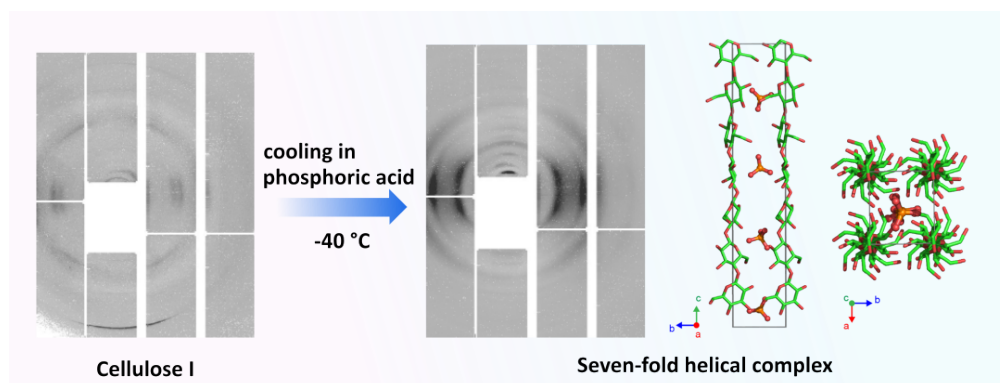
14 Phosphoric acid is widely used for the swelling and hydrolysis of cellulose. The
15 detailed description of molecular interactions between cellulose and phosphoric acid
16 is essential for understanding and controlling these processes. Here, to obtain
17 structural insights into the swelling behavior, we investigated the structural evolution
18 of cellulose swollen in the concentrated phosphoric acid solution using X-ray fiber
19 diffraction and solid-state NMR spectroscopy. We observed the formation of a
20 crystalline complex of cellulose and phosphoric acid at -40 °C, where cellulose
21 molecules adopt a seven-fold helical conformation. This structure is the second known
22 cellulose-acid crystalline complex and the first cellulosic crystal consisting of seven-
23 fold helical chains. Our observation highlights the conformational flexibility of cellulose
24 molecules in the solvated states and the strong influence of cellulose-acid interactions
25 on the packing and conformation of cellulose molecules.

26

27 **Keywords:** Cellulose; Phosphoric acid; Crystallosolvate; seven-fold helix

28

29 **Graphical abstract**



30

31

32 Introduction

33 Cellulose forms crystalline complexes with various polar solvent molecules.
34 The crystal structure analysis of these solvated crystals, or crystallosolvates, has
35 provided valuable insights into the molecular interactions between cellulose and
36 solvent molecules and the conversion mechanisms between cellulose crystalline
37 allomorphs. The crystalline complexes with alkalis and amines are among the most
38 extensively studied cellulose crystallosolvates as intermediate structures in
39 allomorphic conversions (Sobue et al. 1939; Lee et al. 1983; Okano and Sarko 1984,
40 1985; Porro et al. 2007; Wada et al. 2009). These molecules are the components of
41 historic cellulose solvents such as aqueous alkali solutions and cupriethylenediamine
42 hydroxide solutions. In recent years, new cellulose crystallosolvates were reported
43 with industrially relevant solvent molecules such as N-methyl morpholine-N-oxide
44 (NMMO) and ionic liquids (Song et al. 2018; Nishiyama et al. 2019; Endo et al. 2020).

45 Mineral acids, such as sulfuric and phosphoric acid, are essential reagents for
46 processing cellulose materials. They are widely-used hydrolysis and swelling agents
47 of cellulose. Concentrated mineral acid solutions are also direct solvents of cellulose.
48 Despite their importance, the complexation of cellulose and acid molecules has not
49 received attention compared to the solvent molecules mentioned above. This situation
50 hinders us from understanding the molecular-level details of the interactions between
51 cellulose and acids. We recently reported the cellulose-sulfuric acid complex as the
52 first cellulose-acid crystallosolvate (Li et al. 2022). The complexation of cellulose and
53 sulfuric acid occurs only at low temperatures, and cellulose molecules adopt a rare
54 five-fold helical conformation (five glucose residues in two turns, or $n=2.5$) in the
55 complex structure. Such a structure was not without precedent; earlier work with the
56 gelling molecule xanthan had shown that its cellulose backbone could apparently form

57 both single- and double-helices with each strand having the $n=2.5$ conformation
58 (French and Johnson, 2004; Moffat et al. 2016; Morris 2019). The complexation was
59 likely driven by the deficiency of water around sulfuric acid molecules that recruit
60 cellulose hydroxyl groups to form a stable hydrate-like structure (Li et al. 2022). This
61 finding prompted us further to explore the complexation behavior of cellulose and
62 mineral acids. In this study, we extended a similar approach to concentrated
63 phosphoric acid.

64 Phosphoric acid is widely used as a swelling agent of cellulose. It dissolves
65 cellulose at room temperature, and subsequent regeneration of cellulose yields
66 amorphous cellulose. Such amorphous cellulose, so-called phosphoric acid swollen
67 cellulose or PASC, is the most widely used form of amorphous cellulose and a
68 standard substance for assays of cellulolytic enzyme activity (Zhang et al. 2006; Hall
69 et al. 2010). While sulfuric acid is more common in literature, phosphoric acid is also
70 used to hydrolyze cellulose to prepare cellulose oligomers and nanocrystals (Isogai
71 and Usuda 1991; Camarero Espinosa et al. 2013; Billès et al. 2016). Furthermore, like
72 sulfuric acid, concentrated phosphoric acid solution can dissolve cellulose at high
73 concentrations. Cellulose molecules form a liquid crystalline phase in the resulting
74 highly concentrated dope (Boerstoel et al. 2001). The anisotropic molecular
75 organization in the dope can be exploited to obtain highly oriented cellulose II fibers.
76 These fibers, called Fiber B, exhibit improved mechanical properties compared to the
77 other commercially regenerated cellulose fibers (Northolt et al. 2001).

78 Given its importance in the processing of cellulose, it is crucial to better
79 understand the molecular interactions between cellulose and phosphoric acid. Here,
80 we investigated the structure of cellulose swollen in a concentrated phosphoric acid
81 solution at low temperatures. Using X-ray fiber diffraction and ^{13}C solid-state nuclear
82 magnetic resonance (NMR) spectroscopy, we observed the formation of a
83 crystallosolvate of cellulose and phosphoric acid at $-40\text{ }^\circ\text{C}$. Interestingly, the structure
84 has an exceptionally long fiber repeat of 37 \AA , making it the first example of seven-
85 fold helical molecular conformation in cellulosic crystals. This extended form, seven-
86 fold helix is unlike the collapsed seven-fold V_7 amylose structure (Nishiyama et al.
87 2010).

88

89 **Material and Methods**

90

91 **X-ray scattering measurement**

92 X-ray scattering measurements were carried out at the D2AM beamline at the
93 European Synchrotron Radiation Facility (ESRF) using an X-ray beam of 16 KeV and
94 hybrid pixel detectors (D5 and WOS). Bundles of well-aligned flax fibers were packed
95 in glass capillaries and immersed in 83 wt% H₃PO₄ aqueous solution at room
96 temperature. The specimens were then mounted on a low-temperature specimen
97 holder and placed under a vacuum. The diffraction patterns were acquired at
98 temperatures ranging from room temperature to -40°C with an exposure time of 3
99 seconds. The background scattering from the solvent in the patterns was subtracted
100 numerically using in-house programs.

101 **Solid-state ¹³C NMR spectroscopy**

102 Solid-state ¹³C CP/MAS NMR analysis was performed on swollen and dry regenerated
103 flax fibers using a Bruker Avance III 400 spectrometer (100 MHz, ¹³C). For the swollen
104 samples, spectra were obtained at room and low temperatures (23, -20, and -40°C).
105 Approximately 50 mg of flax fibers were swollen in 83 wt% H₃PO₄ solution and allowing
106 them to swell for 10 minutes. The fibers were then mixed with ca. 10 mg of glass wool
107 and packed into the zirconia rotor. The specimen temperature was controlled using a
108 Bruker BCU-II cooling unit. The spinning rate was set at 9 kHz with 2 ms contact time.

109 **Model construction and analysis**

110 Models of cellulose molecules in seven-fold conformations were built using a
111 glucose residue of the corner chain of cellulose I_β. Each residue was rotated about the
112 chain axis by 51° (7₁), 103° (7₂), 154° (7₃), 206° (7₄), 257° (7₅), and 309° (7₆) and shifted
113 along the chain axis by 5.29 Å compared to the adjacent residue. The models were
114 visualized and analyzed using Pymol software (Schrödinger and DeLano 2020). The
115 glycosidic torsion angles of the generated helices were compared with the crystal
116 structure conformations of cellulose allomorphs and di- and oligosaccharide crystals
117 from the Protein Data Base with cellobiose linkage as well as the conformational
118 energy map of cellobiose based on SMD continuum solvation and B3LYP/6-31+G(d)
119 energies (French et al., 2012, 2021).

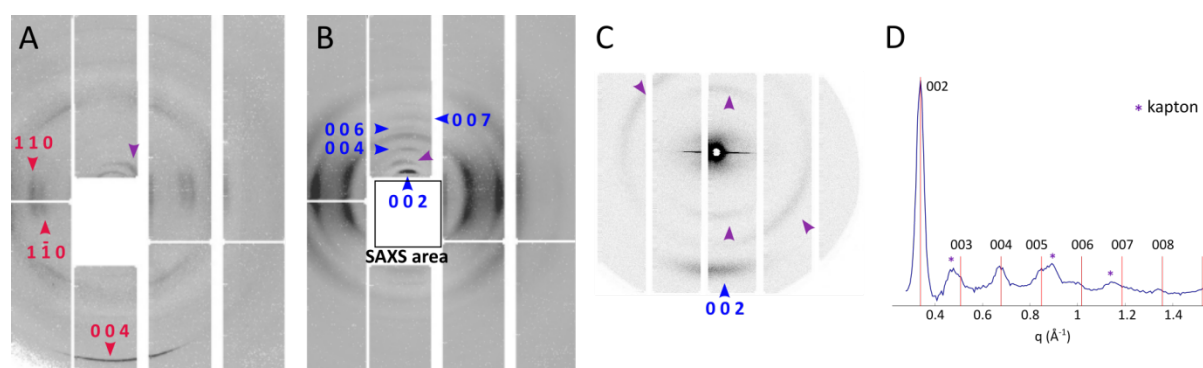
120

121 **Results and Discussion**

122 **Crystal structure of cellulose-phosphoric acid complex**

123 We performed in situ low-temperature X-ray scattering measurements to
 124 monitor the changes in the crystal structure of flax cellulose immersed in concentrated
 125 H_3PO_4 aqueous solution. We chose a concentration of 83 wt% since more
 126 concentrated solutions could dissolve cellulose fiber at room or low temperature.
 127 Furthermore, this concentration is commonly used in the preparation of amorphous
 128 cellulose.

129 Fig. 1A shows the X-ray fiber diffraction diagram of flax fiber immersed in 83
 130 wt% H_3PO_4 at room temperature (20 °C) for 15 minutes. The room-temperature
 131 diagram shows characteristic reflections of cellulose I_β , i.e., equatorial 1 -1 0 and 1 1
 132 0, and meridional 0 0 4. Thus, the room-temperature immersion in 83 wt% H_3PO_4 did
 133 not affect the crystal structure of flax cellulose for at least 15 min.



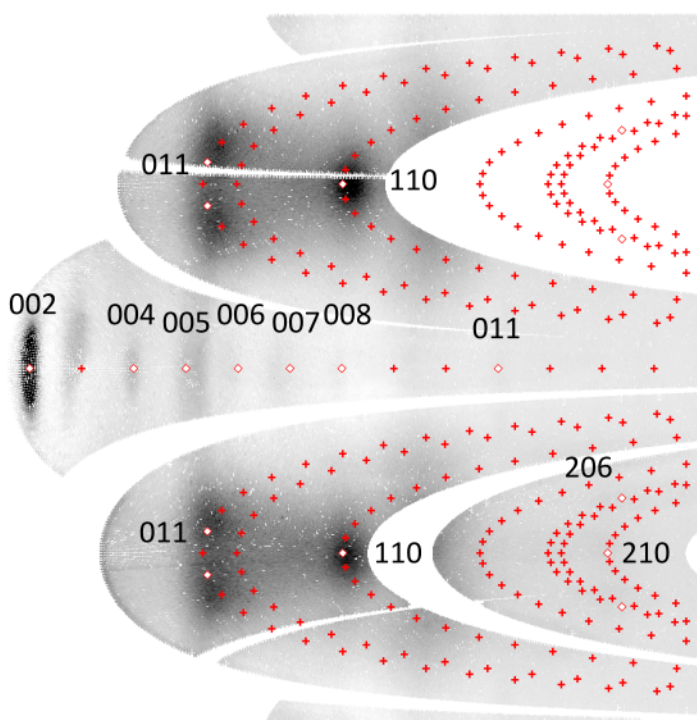
134
 135 **Fig. 1** X-ray fiber diffraction patterns of flax cellulose immersed in 83 wt% phosphoric
 136 acid solution at (A) room temperature and (B, C) -40 °C. Panel (C) corresponds to the
 137 small angle area indicated in panel (B). Reflections indicated by purple arrowheads
 138 correspond to the Kapton window used in the low-temperature vacuum cell. (D)
 139 Meridional profile of the diffraction diagram in (B). Reflection indices are shown above
 140 peaks.

141
 142 When cooled down to -40 °C (Fig. 1B and 1C), the diagram showed diffraction
 143 features different from cellulose I_β , indicating the crystalline complex formation of
 144 cellulose and phosphoric acid. The diffraction pattern of the cellulose-phosphoric acid
 145 complex is poorly resolved, with only a few strong reflections in the observed wide-
 146 angle area. On the meridian of the fiber diagram, a new set of reflection spots
 147 appeared at $d = 18.5, 9.3, 6.2, 5.3 \text{ \AA}$ (Fig. 1D). No other reflection appears on the
 148 meridian in the smaller angle region (Fig. 1C). This meridional pattern indicates that

149 the fiber repeat of the complex crystal is 37 Å, increased from that of cellulose I_β (10
150 Å). The aforementioned meridional reflections were thus indexed, from the utmost to
151 the outermost, as 0 0 2, 0 0 4, 0 0 6, and 0 0 7, respectively (Fig. 1D). The long 37-Å
152 fiber repeat suggests cellulose molecules likely adopt a seven-fold helical
153 conformation in the complex. The rise per residue, *h*, is thus 5.28 Å for this helix, the
154 longest among the known cellulosic crystals (cf. 5.19 Å for cellulose I_β).

155 We refined the unit cell based on 11 independent reflections indicated in Fig.
156 2. The refined unit cell is orthorhombic with the cell parameters of *a* = 6.19 Å, *b* = 6.95
157 Å, and *c* = 37.05 Å. In Fig. 2, the reflection positions predicted from the unit cell are
158 superimposed on the diffraction pattern remapped in the polar coordinate. They are in
159 good agreement with the observed reflection positions. The unit cell volume is 1594
160 Å³. The volumes of each component molecule are 164 Å³ for glucose residue based
161 on cellulose I_β crystal structure (Nishiyama et al. 2002), 75 Å³ for phosphoric acid
162 based on its hemihydrate crystal structure (Smith et al. 1955), and 30 Å³ for water.
163 Thus, the unit cell contains only one cellulose chain composed of seven glucose
164 residues (1148 Å³). If the molar ratio between phosphoric acid and water remains
165 unchanged at 1:1 in the 83 wt% solution and the complex, the unit cell may
166 accommodate four phosphoric acid and four water molecules (420 Å³). This
167 stoichiometric ratio makes the sum of the component volume, 1568 Å³, nearly equal
168 to the unit cell volume. The strong 0 0 2 and 0 0 4 intensities, corresponding to
169 periodicities of 18.5 and 9.3 Å along the fiber axis, support the even number of
170 phosphoric acid molecules in the unit cell.

171



172

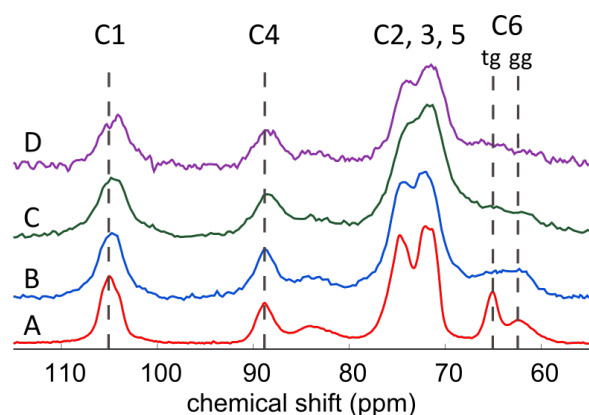
173 **Fig. 2** X-ray diffraction pattern of flax cellulose immersed in 83 wt% H₃PO₄ in polar
 174 coordinates. Reflection positions predicted based on the refined orthorhombic unit cell
 175 parameters are indicated with red crosses. 11 reflections used for the unit cell
 176 refinement are indicated with white circles.

177

178 While the calculated diffraction positions fit well with the experimental ones, the
 179 real unit cell is bigger than the one-chain orthorhombic unit cell. No two-fold rotation
 180 axis can be conceived along either unit-cell axis due to the low symmetry of cellulose
 181 molecule in the seven-fold structure. This unit cell thus does not satisfy the symmetry
 182 conditions of the orthorhombic systems. The crystallographically correct unit cell
 183 should have each axis at least twice as long as the current proposed values.
 184 Nevertheless, we believe that the molecular packing and structural details in the
 185 complex can be discussed based on this small one-chain cell since it is likely an
 186 asymmetric unit of the real unit cell.

187 To confirm the seven-fold helical molecular conformation, we conducted the
 188 solid-state NMR spectroscopy of the cellulose-phosphoric acid complex. We
 189 measured CP/MAS ¹³C NMR spectra of cellulose swollen in 83 wt% H₃PO₄ solution at
 190 three temperature conditions, i.e., room temperature (23 °C), -20 °C, and -40 °C.

191



192

193 **Fig. 3** CP/MAS ^{13}C NMR spectra of flax cellulose under different conditions. (a) intact
 194 flax cellulose and (b-d) flax cellulose swollen in 83 wt% H_3PO_4 and measured at (b)
 195 room temperature, (c) $-20\text{ }^\circ\text{C}$, and (d) $-40\text{ }^\circ\text{C}$.

196

197 As shown in Fig. 3, the spectra of swollen samples (Fig. 3b-d) are different from
 198 that of intact flax cellulose (Fig. 3a), especially in the C6 region (60-68 ppm) (Horii et
 199 al. 1983). The sharp signal at 66 ppm of cellulose I_β , corresponding to the *tg* conformer,
 200 disappeared in all the swollen samples. The swelling in phosphoric acid thus resulted
 201 in the loss of the original cellulose I_β crystal structure regardless of the temperature.
 202 The room-temperature swollen sample had a broad C6 signal at 62 ppm
 203 corresponding to the *gg* conformer, while even more broader peaks were observed at
 204 $-20\text{ }^\circ\text{C}$ and $-40\text{ }^\circ\text{C}$ (Fig. 3c and 3d).

205 The loss of cellulose I_β crystal structure at room temperature contradicts the X-
 206 ray diffraction result at room temperature (Fig. 1A). This discrepancy likely arose from
 207 the duration of swelling. We carried out the X-ray scattering measurement at room
 208 temperature after 15 min of swelling. In comparison, the sample preparation time and
 209 measurements of solid-state NMR experiments were ca. 30 min and 5 hours,
 210 respectively. Furthermore, the swelling of flax fiber was laterally constrained by the
 211 glass capillary in the X-ray scattering experiments, which may have further slowed
 212 down the swelling process at room temperature.

213 Compared to the drastically altered C6 signals, the changes in the other carbon
 214 signals are less significant. The C1 and C4 upfield shifted by 1 and 0.5 ppm,
 215 respectively, at $-40\text{ }^\circ\text{C}$ compared to those of cellulose I_β . These upfield shifts suggest

216 the conformational change of the cellulose molecular chain around the glycosidic
217 linkage at the low temperature. The observation is consistent with the occurrence of
218 the seven-fold helical chain conformation in the complex predicted based on the X-ray
219 diffraction results. The change in the C1 and C4 chemical shifts is less pronounced at
220 the higher temperature.

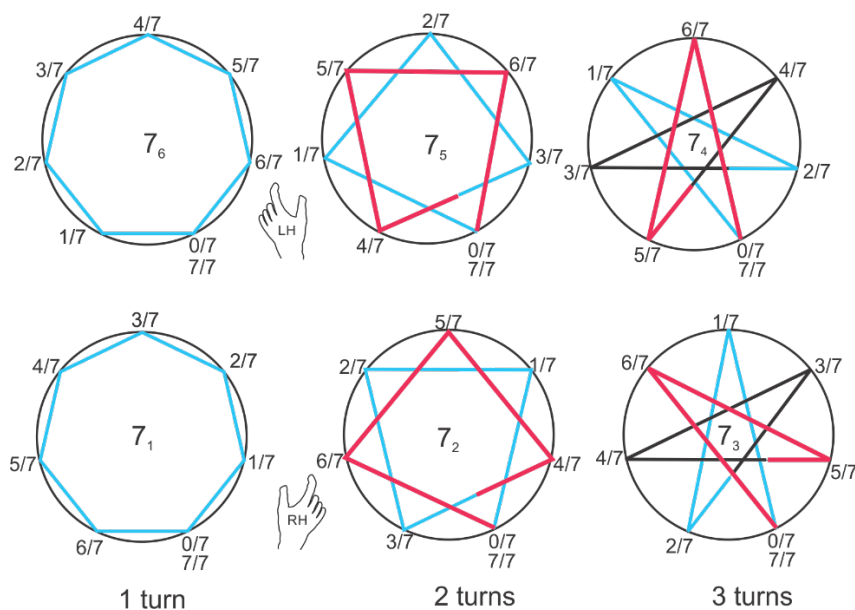
221 The chemical shifts of C1 and C4 are known susceptible to the conformational
222 changes of glycosidic linkage of β -1,4 (e.g. cellulose and xylan) and α -1,4 linked
223 glycans (e.g. amylose) (Veregin et al. 1987; Porro et al. 2007; Falcoz-Vigne et al.
224 2017). The magnitude of the observed C1 upfield shift is similar to those of the
225 solvation-induced conformation changes in cellulose and xylan crystals. For instance,
226 the C1 signal shifts up-field by ca. 1 ppm in Na-cellulose II with three-fold helical
227 cellulose chains compared to cellulose I with two-fold helices. On the other hand, the
228 same conformational change gives a larger up-field shift of the C4 signal, by 9 ppm,
229 in the transition from cellulose I to Na-cellulose II (Porro et al. 2007).

230 The small C4 upfield shift likely indicates the similarity of the molecular
231 conformation in the complex to the two-fold one in cellulose I. The seven-fold
232 symmetry allows six possible helices: 7_1 , 7_2 , 7_3 , 7_4 , 7_5 , and 7_6 helices. The nominal
233 rotation angles between adjacent glucose residues are 51° (7_1), 103° (7_2), 154° (7_3),
234 206° (7_4), 257° (7_5), 309° (7_6), respectively. The first three are right-handed, whereas
235 the last three are considered left-handed because a left-handed helical thread that
236 connects equivalent points on helices with a rotation angle greater than 180° has a
237 shorter path by proceeding in a left-handed manner. The 7_3 and 7_4 helices thus have
238 high conformational similarities to the two-fold helices of native cellulose, which have
239 a 180° rotational angle between residues.

240

241

242



243

244 Fig. 4. The various 7-fold helices viewed down the helix axis. Blue colors
 245 indicate that the lines that connect equivalent points are in the first helix turn, the red
 246 lines (for the 2- and 3-turn repeats) are in the last turn (closest to the viewer) and the
 247 black lines (for the 3-turn repeat structures) are in the second turn.

248

249 Fig. 4 illustrates the possible seven-fold helices with circles representing the
 250 projected helical thread that connects equivalent points (French and Johnson, 2009).
 251 Straight lines connect arbitrarily assigned equivalent points. It is initially helpful to think
 252 of the equivalent points as the linkage oxygen atoms. In that case, the lines can
 253 represent the O1—O4 virtual bonds of the glucose residue. The numbers around the
 254 circles indicate the fractional height along the z-axis for the structure repeat,
 255 regardless of the number of turns. For the right-handed helix in the lower left corner,
 256 7_1 , with seven residues in one turn, the height of the equivalent points starts at 0 and
 257 increases toward the viewer after a counter-clockwise rotation of about 51.4° to the
 258 second point at $1/7$ of the helix repeat. A second 51° rotation gives a height of $2/7$ of
 259 the repeat and so-forth until one complete turn brings the equivalent point back to the
 260 $7/7$ point where the helix generation process begins for a second turn.

261 For the 7_6 helix above it, also with seven residues in one turn, clockwise
 262 rotations of 51.4° each produce the analogous left-handed helix. If a rotation of 309°
 263 had been applied in a counter-clockwise sense from the $0/7$ point, it would have
 264 brought the equivalent point to the $6/7$ point on the right-handed lower drawing. But

265 that is in the exact same place as the 1/7 point on the left-handed helix and the path
266 is much shorter. This is why this left-handed helix is said to have 7_6 helix symmetry.

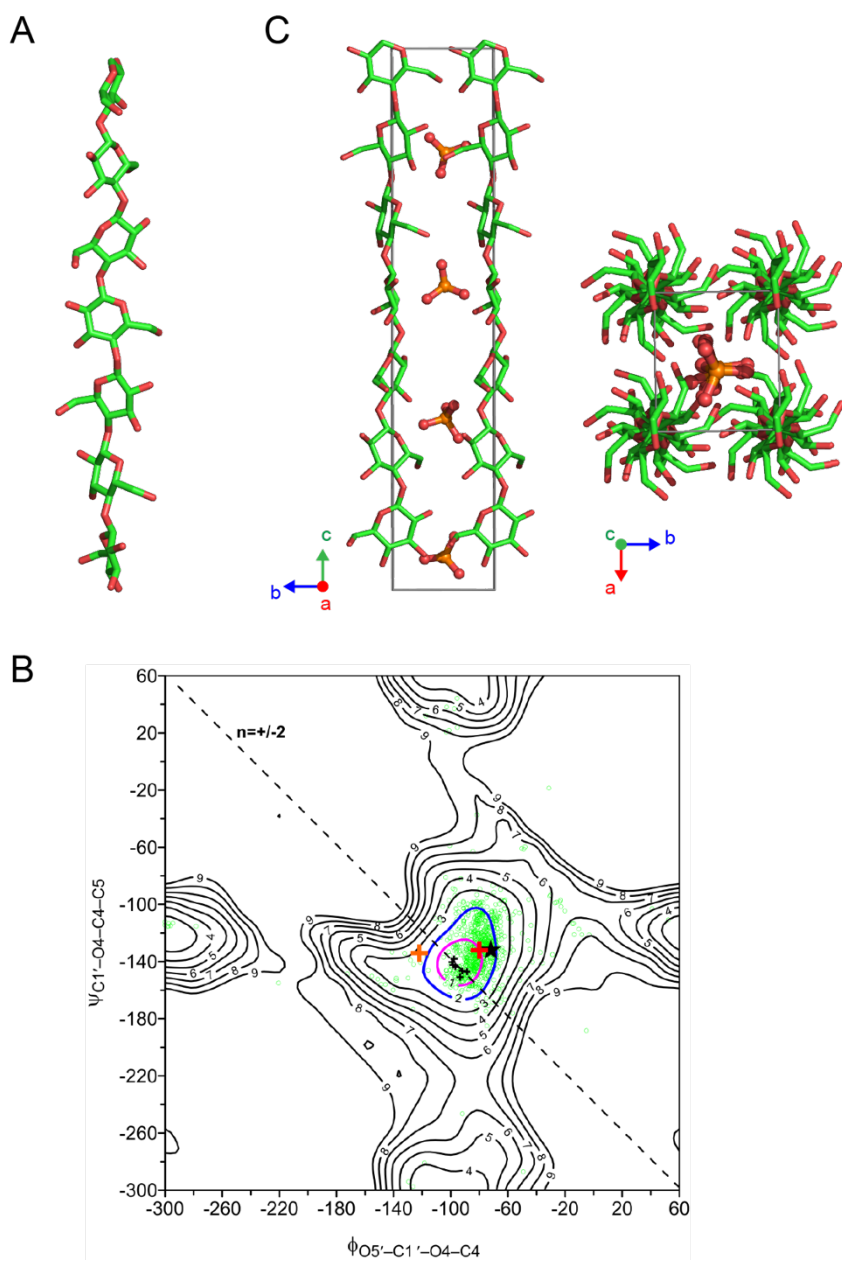
267 Looking at the 7_5 helix in the upper middle, the equivalent points are spaced at
268 102.9° clockwise increments. Thus 3.5 residues will be needed to complete the first
269 turn, the path indicated by the blue line that connects equivalent points. The path of
270 these lines connecting equivalent points for the second turn of the helix is indicated by
271 the red lines, taking again 3.5 residues to complete the second turn. Here, the 1/7
272 point indicates that the equivalent point has moved towards the viewer by one seventh
273 of the two-turn structure repeat distance.

274 The upper right diagram depicts the 7_4 helix. To get to the 1/7 point, a clockwise
275 rotation of 154.3° is needed, indicated by the blue line starting from the 0/7 point. Each
276 turn will have $(7 \div 3 =) 2.3333$ of the connecting lines. The height of the first turn is
277 reached after two lines connecting equivalent points plus one-third of the next line.
278 The second 360° turn of the helix includes the remaining black two-thirds of the O2—
279 O3 virtual bond, all of the O3—O4 virtual bond, and two-thirds of the O4—O5 virtual
280 bond, indicated by the black lines. That is followed by the red set of 2.3333 lines that
281 bring the third sequence to the structure repeating point of the three-turn helix (7/7).

282 A few more observations complete these explanations. If a structure has only
283 two residues per turn, the distance between equivalent points is the same regardless
284 of whether a clockwise or counter-clockwise rotation is used to reach the second point
285 180° away. A slight twist in either direction will create a helix that has a shorter distance
286 to the next equivalent point in either the clockwise or the counter-clockwise direction.
287 If the distance is shorter, then there will be more residues per turn. That is why there
288 are no helices with descriptions of fewer than two residues per turn. Also, it is important
289 to remember that the helical description applies to any given point, such as the C6
290 atom, and its equivalent points in other members of the helix. For this reason, a given
291 rotation would apply to all the atoms, and the chirality of the glucose residue itself
292 would be maintained, even though the sense, or chirality of the helix changes from
293 right- to left-handed. These distinct helical conformations, including the seven-fold
294 helix in the present structure and the five-fold helix in the sulfuric acid complex are
295 among the different reasons why it is not useful to describe cellulose as having a
296 cellobiose repeating unit (French 2017).

297 We built a model of the 7_3 and 7_4 helical cellulose chains using the corner chain
298 glucose residue of cellulose I $_{\beta}$ (Fig. 5a) (Nishiyama et al. 2002). The two glycosidic

299 torsion angles, ϕ (O5-C1-O1-C4) and ψ (C1-O1-C4-C5), of the 7_4 helix are -80° and -
300 132° , respectively, notably close to $\phi = -89^\circ$ (center), -99° (corner) and $\psi = -147^\circ$, -
301 142° in cellulose I_β . The 7_3 helix has a larger discrepancy in these torsion angles from
302 the two-fold helix, with $\phi = -122^\circ$ and $\psi = -134^\circ$. Thus, the 7_4 helix only slightly twists
303 away from the two-fold helix. This conformation is energetically stable due to its
304 similarity to the 2_1 helical conformation, as shown in the conformational energy map
305 based on the B3LYP density functional theory calculations of cellobiose (Fig. 4b)
306 (French et al. 2012, 2021). Its ϕ/ψ point (red cross in Fig. 5b) is within the 1 kcal/mol
307 contour relative to the global minimum. This contour circumscribes the known neat
308 (un-complexed) cellulose allomorph conformations (black crosses). On the other hand,
309 the 7_3 helix corresponds to an energy of about 2.7 kcal/mol. Unlike the chosen 7_4
310 model, which lies roughly in the center of the observed cellodextrin conformations from
311 the Protein Data Bank (green squares), the 7_3 helix model is in a very sparsely
312 populated area (French et al. 2012). Compared to the five-fold helix found in the
313 sulfuric acid complex, the 7_4 helix is more energetically stable. The most plausible
314 helical conformation in the sulfuric acid complex is a 5_3 helix with $\phi = -72^\circ$ and $\psi = -$
315 132° , giving an energy of ca. 1.8 kcal/mol (black star in Fig 4b).
316



317

318 **Fig. 5** (a) Cellulose chain in 7₄ helical conformation. (b) Conformational energy map
 319 of cellobiose based on SMD continuum solvation and B3LYP/6-31+G(d) energies. Red
 320 and orange crosses correspond to the current structures with 7₄ and 7₃ helical
 321 conformations. The black star corresponds to the 5₃ helical conformation in the sulfuric
 322 acid complex. The black cross and green squares are for conformations of cellulose
 323 allomorphs and for di- and oligosaccharide crystals from the Protein Data Base with
 324 cellobiose linkage, respectively. The dotted diagonal line represents two-fold screw
 325 conformations. (c) Molecular packing model of the cellulose-phosphoric acid complex.

326

327 Based on the above-mentioned consideration of the conformational energy and
328 the solid-state NMR results, the 7_4 helix is the most likely conformation formed in the
329 phosphoric acid complex. We thus built a molecular packing model of the cellulose-
330 phosphoric acid complex composed of 7_4 cellulose helices, as shown in Fig. 5c. In this
331 model, cellulose chains are at the corners of the unit cell, leaving room for phosphoric
332 acid molecules at the center of the cell. Four phosphoric acid molecules stack along
333 the fiber axis in this central space. Considering the drastic change in the C6 signal in
334 the solid-state NMR spectra, the hydroxymethyl group should directly interact with the
335 phosphoric acid molecule. On the other hand, the complex may retain the
336 intramolecular O3-H...O5 hydrogen bond, the common feature in the neat cellulose
337 allomorphs, considering its conformational similarity with the 2_1 helix. The distance
338 between O3 and O5 in the 7_4 helix is 2.84 Å, only 0.1 Å longer than that in cellulose I_β
339 (Nishiyama et al., 2002).

340 The complexation with phosphoric acid caused unexpected changes in the
341 molecular conformation of cellulose. This complex is the first crystal structure
342 consisting of β -1,4 glycan chains in a seven-fold helical conformation. Together with
343 the five-fold helical conformation observed in the cellulose-sulfuric acid complex, the
344 seven-fold helix in this complex indicates the strong influence of the cellulose-acid
345 interactions on the packing and conformation of cellulose molecules. The long pitch
346 per glucose residue is another characteristic of this complex structure. In most
347 cellulose crystallosolvates, this value decreases compared to cellulose I, presumably
348 due to the solvation-induced structural relaxation. The stretched molecular
349 conformation may result from the spatial coordination of cellulose and phosphoric acid
350 molecules. A further crystallographic analysis based on an improved diffraction data
351 set is necessary to elucidate such finer structural details.

352 As aforementioned, the diffraction pattern of the phosphoric acid complex was
353 poorly resolved, and the radial width of reflections remained approximately the same
354 as the original cellulose I_β . These observations are in striking contrast with the
355 cellulose-sulfuric acid complex, whose radial width of reflections were significantly
356 narrower than the original cellulose I_β . We attributed this apparent increase of
357 crystallite size in the sulfuric acid complex to the increase in the interchain distance of
358 cellulose due to the incorporation of sulfuric acid molecules in the crystalline lattice (Li
359 et al. 2022). Although its unit cell is significantly smaller than the sulfuric acid complex,

360 the interchain distance also increased in the phosphoric acid complex compared to
361 cellulose I_β. This lack of increase in the crystallite size implies the partial amorphization
362 or disorganization of the cellulose crystal domain during the swelling process in
363 phosphoric acid. While the preparation conditions of the complexes are nearly
364 identical for these two acid species, the swelling and complexation behaviors are
365 distinct between sulfuric and phosphoric acids.

366

367 **Conclusion**

368 We reported the formation of the cellulose-phosphoric acid crystalline complex
369 based on X-ray diffraction measurement and solid-state NMR spectroscopy at -40 °C.
370 This complex is the second known cellulose-acid complex and the first crystal structure
371 consisting of cellulose molecules in a seven-fold helical conformation. The occurrence
372 of the seven-fold conformation in this complex structure highlights the conformational
373 flexibility of cellulose in its solvated states. The range of structures includes the Na-
374 cellulose II structure with $n=3$, the sulfuric acid complex with a 5_3 helical conformation
375 ($n=2.5$), the present phosphoric acid structure with a 7_4 helical conformation ($n=$
376 2.333), and cellulose itself with $n=2$. We are currently conducting a further
377 investigation of the complexation behaviors of cellulose and mineral acids and a more
378 detailed crystallographic analysis of these complexes. Such efforts will improve our
379 understanding of the molecular interactions between cellulose and acids, which is
380 crucial for the optimal processing of cellulosic biomass.

381

382 **Competing interests**

383 The authors have no relevant financial or non-financial interests to disclose.

384

385 **Consent for publication**

386 All authors have provided their consent for publication.

387

388 **Availability of data**

389 All data generated or analyzed during this study are included in this article.

390

391 **Authors' contributions**

392 I.M. and Y.O. conceptualized the study. J.H.L., I.M., I.J., Y.N., and Y.O. performed
393 the experiments. J.H.L., Y.N., A.D.F., and Y.O. analyzed the data. J.H.L., A.D.F.,
394 and Y.O. wrote the manuscript. All authors read and approved the final manuscript.

395 **Acknowledgment and funding**

396 We acknowledge Agence Nationale de la Recherche (ANR grant number: ANR-21-
397 CE29-0016-1 and ANR-11-EQPX-0010 under the "Investissements d'Avenir"
398 program) and Glyco@Alps (ANR-15-IDEX-02) for the financial support and the
399 NanoBio-ICMG platform (FR 2607) for granting access to the NMR facility. ESRF is
400 acknowledged for the provision of beamtime (experiment number A02-1-902, D2AM
401 beamline). We acknowledge Benjamin Richter and Yves Watier at the Sample
402 Environment Service of ESRF, who designed the temperature-controlled specimen
403 holder.

404 **Ethics approval and consent to participate**

405 Not applicable

406

407 **References**

- 408 Billès E, Onwukamike KN, Coma V, et al (2016) Cellulose oligomers production and
409 separation for the synthesis of new fully bio-based amphiphilic compounds.
410 *Carbohydr Polym* 154:121–128. <https://doi.org/10.1016/j.carbpol.2016.07.107>
- 411 Boerstoeel H, Maatman H, Westerink JB, Koenders BM (2001) Liquid crystalline solutions of
412 cellulose in phosphoric acid. *Polymer* 42:7371–7379. [https://doi.org/10.1016/S0032-3861\(01\)00210-5](https://doi.org/10.1016/S0032-3861(01)00210-5)
- 414 Camarero Espinosa S, Kuhnt T, Foster EJ, Weder C (2013) Isolation of Thermally Stable
415 Cellulose Nanocrystals by Phosphoric Acid Hydrolysis. *Biomacromolecules* 14:1223–
416 1230. <https://doi.org/10.1021/bm400219u>
- 417 Endo T, Yoshida S, Kimura Y (2020) Self-Assembly and Complexation of Cellulose/Ionic Liquid
418 at High Cellulose Concentration: Anion Dependence. *Cryst Growth Des* 20:6267–
419 6271. <https://doi.org/10.1021/acs.cgd.0c00346>
- 420 Falcoz-Vigne L, Ogawa Y, Molina-Boisseau S, et al (2017) Quantification of a tightly adsorbed
421 monolayer of xylan on cellulose surface. *Cellulose* 24:3725–3739.
422 <https://doi.org/10.1007/s10570-017-1401-z>
- 423 French AD, Johnson GP (2004) What crystals of small analogs are trying to tell us about
424 cellulose structure. *Cellulose* 11:5-22.
425 <https://doi.org/10.1023/B:CELL.0000014765.94239.fe>
- 426 French AD, Johnson GP (2009) Cellulose and the twofold screw axis: modeling and
427 experimental arguments. *Cellulose* 16:959-973. [https:// DOI 10.1007/s10570-009-9347-4](https://doi.org/10.1007/s10570-009-9347-4)
- 428
- 429 French AD (2017) Glucose, not cellobiose, is the repeating unit of cellulose and why that is

430 important. *Cellulose* 24:4605–4609 <https://doi.org/10.1007/s10570-017-1450-3>
431 French AD, Johnson GP, Cramer CJ, Csonka GI (2012) Conformational analysis of cellobiose
432 by electronic structure theories. *Carbohydr Res* 350:68–76.
433 <https://doi.org/10.1016/j.carres.2011.12.023>
434 French AD, Montgomery DW, Prevost NT, Edwards JV, Woods RW (2021) Comparison of
435 cellulooligosaccharide conformations in complexes with proteins with energy maps for
436 cellobiose. *Carbohydr Polym* 264:118004
437 Hall M, Bansal P, Lee JH, et al (2010) Cellulose crystallinity – a key predictor of the enzymatic
438 hydrolysis rate. *FEBS J* 277:1571–1582. [https://doi.org/10.1111/j.1742-](https://doi.org/10.1111/j.1742-4658.2010.07585.x)
439 [4658.2010.07585.x](https://doi.org/10.1111/j.1742-4658.2010.07585.x)
440 Horii F, Hirai A, Kitamaru R (1983) Solid-state ¹³C-NMR study of conformations of
441 oligosaccharides and cellulose. *Polym Bull* 10:357–361.
442 <https://doi.org/10.1007/BF00281948>
443 Isogai A, Usuda M (1991) Preparation of low-molecular weight celluloses using phosphoric
444 acid. *J Jpn Wood Res Soc Jpn*
445 Lee DM, Blackwell J, Litt MH (1983) Structure of a cellulose II-hydrazine complex.
446 *Biopolymers* 22:1383–1399. <https://doi.org/10.1002/bip.360220510>
447 Li W, Ogawa Y, Perez J, et al (2022) Fivefold Helical Cellulose Trapped in a Sulfuric Acid
448 Framework. *Cryst Growth Des* 22:20–25. <https://doi.org/10.1021/acs.cgd.1c00664>
449 Moffat J, Morris VJ, Al-Assaf S, Gunning AP (2016) Visualisation of xanthan conformation by
450 atomic force microscopy. *Carbohydr Polym* 148:380–389.
451 <https://doi.org/10.1016/j.carbpol.2016.04.078>
452 Morris ER (2019) Ordered conformation of xanthan in solutions and “weak gels”:
453 Single helix, double helix – or both? *Food Hydrocoll* 86:18–25.
454 <https://doi.org/10.1016/j.foodhyd.2017.11.036>
455 Nishiyama Y, Mazeau K, Morin M, Cardoso MB, Chanzy H, Putaux J-L (2010) Molecular and
456 crystal structure of 7-fold V-amylose complexed with 2-propanol. *Macromolecules*,
457 43(20):8628-8636
458 Nishiyama Y, Asaadi S, Ahvenainen P, Sixta H (2019) Water-induced crystallization and nano-
459 scale spinodal decomposition of cellulose in NMMO and ionic liquid dope. *Cellulose*
460 26:281–289. <https://doi.org/10.1007/s10570-018-2148-x>
461 Nishiyama Y, Langan P, Chanzy H (2002) Crystal Structure and Hydrogen-Bonding System in
462 Cellulose I β from Synchrotron X-ray and Neutron Fiber Diffraction. *J Am Chem Soc*
463 124:9074–9082. <https://doi.org/10.1021/ja0257319>
464 Northolt MG, Boerstoeel H, Maatman H, et al (2001) The structure and properties of cellulose
465 fibres spun from an anisotropic phosphoric acid solution. *Polymer* 42:8249–8264.
466 [https://doi.org/10.1016/S0032-3861\(01\)00211-7](https://doi.org/10.1016/S0032-3861(01)00211-7)
467 Okano T, Sarko A (1984) Mercerization of cellulose. I. X-ray diffraction evidence for
468 intermediate structures. *J Appl Polym Sci* 29:4175–4182.
469 <https://doi.org/10.1002/app.1984.070291247>
470 Okano T, Sarko A (1985) Mercerization of cellulose. II. Alkali–cellulose intermediates and a
471 possible mercerization mechanism. *J Appl Polym Sci* 30:325–332.
472 <https://doi.org/10.1002/app.1985.070300128>
473 Porro F, Bédoué O, Chanzy H, Heux L (2007) Solid-State ¹³C NMR Study of Na–Cellulose
474 Complexes. *Biomacromolecules* 8:2586–2593. <https://doi.org/10.1021/bm0702657>
475 Schrödinger L, DeLano W (2020) PyMOL. <http://www.pymol.org/pymol>
476 Smith JP, Brown WE, Lehr JR (1955) Structure of Crystalline Phosphoric Acid. *J Am Chem Soc*

477 77:2728–2730. <https://doi.org/10.1021/ja01615a013>

478 Sobue H, Kiessig H, Hess K (1939) Das System Cellulose–Natriumhydroxyd–Wasser in
479 Abhängigkeit von der Temperatur. *Z Für Phys Chem* 43B:309–328.
480 <https://doi.org/10.1515/zpch-1939-4324>

481 Song G, Yu J, Ding M, Zhang J (2018) A Novel Cellulose/Ionic Liquid Complex Crystal. *Cryst*
482 *Growth Des* 18:4260–4264. <https://doi.org/10.1021/acs.cgd.8b00754>

483 Veregin RP, Fyfe CA, Marchessault RH, Taylor MG (1987) Correlation of ¹³C chemical shifts
484 with torsional angles from high-resolution, ¹³C-C.P.-M.A.S. N.M.R. studies of
485 crystalline cyclomalto-oligosaccharide complexes, and their relation to the structures
486 of the starch polymorphs. *Carbohydr Res* 160:41–56. [https://doi.org/10.1016/0008-
487 6215\(87\)80302-6](https://doi.org/10.1016/0008-6215(87)80302-6)

488 Wada M, Heux L, Nishiyama Y, Langan P (2009) The structure of the complex of cellulose I
489 with ethylenediamine by X-ray crystallography and cross-polarization/magic angle
490 spinning ¹³C nuclear magnetic resonance. *Cellulose* 16:943–957.
491 <https://doi.org/10.1007/s10570-009-9338-5>

492 Zhang Y-HP, Cui J, Lynd LR, Kuang LR (2006) A Transition from Cellulose Swelling to Cellulose
493 Dissolution by o-Phosphoric Acid: Evidence from Enzymatic Hydrolysis and
494 Supramolecular Structure. *Biomacromolecules* 7:644–648.
495 <https://doi.org/10.1021/bm050799c>