Chemistry Databases

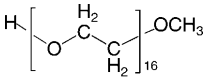
DOI: 10.1002/anie.200904422

Polymer Backbone Conformation—A Challenging Task for Database Information Retrieval

Dieter Seebach, Engelbert Zass, W. Bernd Schweizer, Amber J. Thompson, Alister French, Benjamin G. Davis,* Gwenda Kyd, and Ian J. Bruno**

crystal structure analysis · polyethylene glycol · polymers · structure databases

At the beginning of 2009, in an *Angewandte* Communication, the preparation of monodisperse oligomers of ethylene oxide (up to 48 units), the isolation of single crystals of hexadecaethylene glycol monomethyl ether **1**, and its X-ray crystal structure were reported.[1]



1

We believe that this is a widely relevant result, in view of the practice of attaching polyethylene glycol (PEG) chains (usually as a Gaussian distribution of chain lengths) to therapeutic proteins, peptides, and small molecules (PEGylation) to influence stability, immunogenicity, solubility, pharmacokinetics, and mode of action.[2] The construction of such large monodisperse oligoethylene glycol derivatives requires substantial effort as is true in other cases (cf. oligoethylenes,[3] oligo(ε-aminocaproic acid),[4] oligo(β-hydroxybutanoic acid)[5a]). It is also a breakthrough to have, for the first time, a single-crystal X-ray structure of an oligoethylene glycol derivative consisting of *M*- and *P*-3_{10}-helical building blocks.[1] However, readers will have gained the impression that this is the first crystallographic demonstration of the backbone structure of $(CH_2CH_2O)_n$ chains; this is false, for the reasons outlined herein.

Our common ongoing interests in the conformation of PEG[1] and of oligoethylene glycol conjugates[5b] have led to continued examination of the pertinent literature. We detail here how and why original and subsequent searches missed reports from the 1960s and 1970s on the subject.[6] This, in turn, will highlight some of the current challenges for chemical and crystallographic database construction, interrogation, and information retrieval. Our purpose here is, firstly, to bring the community's attention to this early work regarding PEG crystal structures determined from fibers[6a,b] and freeze-dried[6c] particles, and secondly, to point out problems that can be encountered in modern electronic data retrieval, especially for certain extended polymeric structures, as well as suggest strategies for improved structure retrieval.

It is known that X-C-C-Y systems (X,Y = electronegative heteroatoms) often prefer synclinal over antiplanar conformations;[7] this "*gauche* effect"[8] leads naturally to a helical backbone conformation in PEG. There are scattered statements in textbooks about the "well-known" helical backbone of PEG,[9–11] but how does one find the corresponding original literature?

Since *Beilstein* does not really cover polymers, *Chemical Abstracts* databases are the first choice. Prior to the submission of the original paper, searches using *SciFinder Scholar* on oligoethylene glycols,[1] both on structure and topic (and mutually refined), had failed to reveal reference [6a]. A "Topic Search" in *SciFinder Scholar* with the term "backbone of polyethylene glycol" retrieved 859 references (June 8, 2009); refinement with the term "helix (helical)" reduced this number to seven, none of which was directly relevant; likewise, a refinement with "gauche" failed. A comprehensive compound search gave more than 116000 references, and with the limitation "crystal structure" there were still 657 references; after refinement with "backbone" this was reduced to three, which did not include reference [6a].

In a second approach, the electronic version of *Ullmann*[12] was consulted. A full-text search for "polyethylene glycol" yielded 135 hits (i.e. chapters in *Ullmann*). Restricting the search to "chapter" or "section titles", however, gave no hits. Restricting the full-text search to the topic "polymers and plastics" gave 26 results. The relevance ranking in *Ullmann* placed the chapter on "Polyoxyalkylenes" as 16th on the results list. On the second page of the latter chapter, Table 2 gives structural data of PEG with three references. None of these were directly relevant, but the second one (about

[*] Prof. Dr. D. Seebach, Dr. W. B. Schweizer
Laboratorium für Organische Chemie, Department für Chemie und Angewandte Biowissenschaften, Eidgenössische Technische Hochschule Zürich, HCI Hönggerberg H331
Wolfgang-Pauli-Strasse 10, 8093 Zürich (Switzerland)
E-mail: seebach@org.chem.ethz.ch

Dr. E. Zass
Informationszentrum Chemie, Biologie, Pharmazie
Eidgenössische Technische Hochschule Zürich
HCI Hönggerberg J57.5, Wolfgang-Pauli-Strasse 10
8093 Zürich (Switzerland)

A. J. Thompson, A. French, Prof. Dr. B. G. Davis
Department of Chemistry, University of Oxford
Mansfield Road, Oxford OX1 3TA (UK)

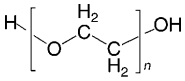
G. Kyd, I. J. Bruno
Cambridge Crystallographic Data Centre (CCDC)
12 Union Road, Cambridge CB2 1EZ (UK)

WILEY InterScience DISCOVER SOMETHING GREAT

© 2009 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Raman spectra)[13] quoted in its introduction the paper by Tadokoro et al.[6a] Thus, a data-retrieval specialist was needed to find the original paper.

The initial fragment search in the Cambridge Structural Database (with ConQuest[14,15]) carried out prior to the original publication, used the fragment "C-C-O-C-C-O". A very large number of hits were obtained that were dominated by the prevalence of crown ethers. Thus, the strategy employed to refine the search for polymeric structures increased the length of the "$(C\text{-}C\text{-}O\text{-})_n$" motif until a manageable number of hits was reached at $n=7$; in the current version of the CSD[15] this produces a total of 141 hits. Although these include a number of extended species, none are of free, unmodified PEG, but rather derivatives or complexes with metal ions. In the light of the failure to find the paper by Tadokoro et al.,[6a] this original search was recently revisited. Altering the search string from "$(C\text{-}C\text{-}O\text{-})_7$" to "$O\text{-}(C\text{-}C\text{-}O\text{-})_6$" and combining this with "only allowed elements C H O", gave 35 hits (mostly cyclic crown ethers). These include a follow-up paper (1973) by Takahashi and Tadokoro[6b] about an improved structural analysis of polyethylene glycol (**2**) from stretched-fiber X-ray diffraction. Adding the non-cyclic bond flag on the fragment eventually yielded reference [6b] and a hexaethylene glycol cyclodextrin cocrystal as the only hits.



2

The reason for the encountered difficulties is the complexity of polymeric structures. The current practice when polymeric structures are added to the CSD is to include all atoms in the asymmetric unit plus the minimum additional atoms needed to fully represent the chemical structure. The bonds to additional atoms are assigned using the CCDC's "polymeric bond type". For organic polymers in particular, two chemically equivalent structures may have different asymmetric units that result in, for example, "polymeric bonds" to carbon in one case and "polymeric bonds" to oxygen in another. In addition, the length of the asymmetric unit may also vary leading to chains of different length being stored in the CSD. This variability can affect the success of a user's search strategy. The fiber structure of Takahashi and Tadokoro[6b] has an asymmetric unit of $CH_2\text{-}(OCH_2CH_2)_6OCH_2$, which is represented as $H_2C\sim CH_2\text{-}(OCH_2CH_2)_6OCH_2\sim CH_2$ (where $\sim$ is the CCDC's "polymeric bond type"). One might imagine that this structure should have been located using the extended "$(C\text{-}C\text{-}O\text{-})_7$" query. However, the polymeric bond type ($\sim$) unfortunately masks the underlying nature of the bond, with the result that the structure was not found by this query. This is further highlighted by the fact that there is, in fact, another structure of PEG, obtained from freeze-dried particles, in the CSD where the asymmetric unit is just OCH_2CH_2;[6c] this was not located by any of the searches described above because all the query fragments used exceeded the length of the asymmetric unit of this structure. The CCDC is currently developing new methods for searching polymeric crystal structures that will make it easier for these to be located in the CSD.

Finally, we want to compare the structures of oligomer **1** and polymer **2**. In the polymer structure reported by Takahashi and Tadokoro the PEG backbone folds to a 7_2-helix[6a] (Figure 1a), corresponding to a $3_{10,5}$-helix in Bragg's nomenclature). In contrast, as a result of extended end-to-end

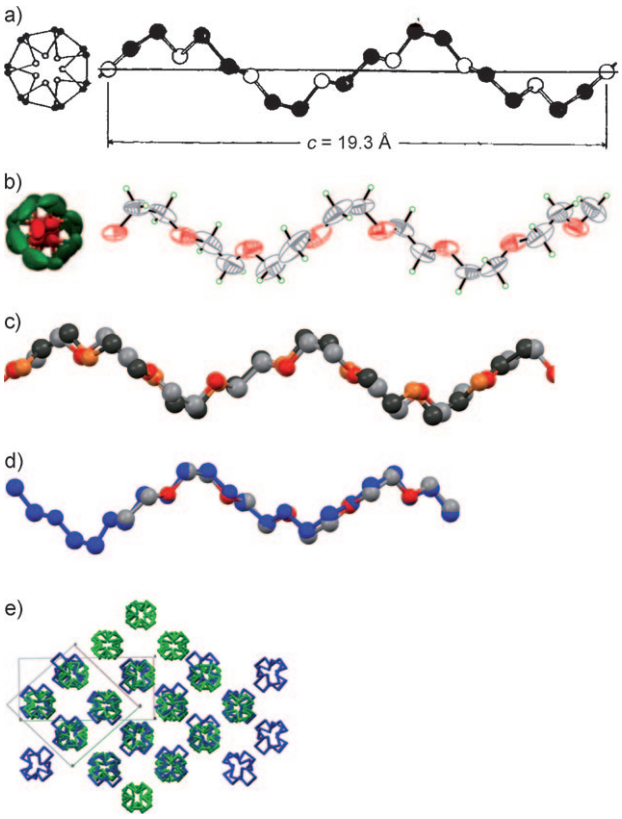


Figure 1. The helical structure of the $(CH_2\text{-}CH_2\text{-}O)_n$ chain. a) Taken from Ref. [6a]; b) taken from Ref. [1]; c) overlay of the backbone in hexadecaethylene glycol monomethyl ether (**1**, red/gray, from the coordinates in the CSD, CCDC No. 707050[1]) and in polyethylene glycol (**2**, orange/black, constructed with coordinates obtained from the geometry reported in Ref. [6a]); d) overlay of the backbones of **1** (red/gray)[1] and **2**[6b] (blue, CSD code WIMYIA); [6b] e) overlay of the structures of the 16mer **1** (green) [1] and of the polymer **2** (blue);[6b] the view down the *c* axis clearly shows the similarity of the packing. All overlays were produced with the *Mercury* program.[16]

packing, and a period of the PEG molecule that is not commensurate with the period of the helices or the lattice, a continuous solution is reported for the single-crystal structure of the hexadecaethylene glycol monomethyl ether (**1**).[1] Thus, the asymmetric unit of this species contains just eight CH_2CH_2O moieties, and the 16mer molecule **1** is actually seen as an infinite string in this X-ray structure, with the analysis leading to a 3_{10}-helical backbone (Figure 1b). Overlaying the individual helices with the Mercury[16] program (Figure 1c and d) demonstrates that they are almost superimposable. Furthermore, overlaying a $2\times2\times2$ "supercell" (Figure 1e) shows that the packing is also similar. The structure of the methyl ether **1** is (as discussed in the Supporting Information in Ref. [1]) a crude, average model, compressed to the smallest possible asymmetric unit, leading to a structure of high symmetry, which accounts for the main difference between **1** and **2**.[17]

 © 2009 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

The case we have presented shows that neither state-of-the-art data-retrieval methods (used by authors) nor the peer-review system (used by editors) can guarantee that seminal contributions are not overlooked in today's publishing of scientific data. Furthermore, we have outlined strategies that may be adopted for more successful searching of polymeric crystallographic data.

Received: August 7, 2009
Published online: November 4, 2009

[1] A. C. French, A. L. Thompson, B. G. Davis, *Angew. Chem.* **2009**, *121*, 1274; *Angew. Chem. Int. Ed.* **2009**, *48*, 1248.
[2] G. Pasut, A. Guiotto, F. M. Veronese, *Expert Opin. Ther. Pat.* **2004**, *14*, 859.
[3] K. S. Lee, G. Wegner, *Makromol. Chem. Rapid Commun.* **1985**, *6*, 203.
[4] M. Rothe in *Chemistry and Physics of Macromolecules: Final Report of the Sonderforschungsbereich Chemie und Physik der Makromoleküle* (Eds.: E. W. Fischer, R. C. Schulze, H. Sillescu), Verlag Chemie, Weinheim, **1991**, p. 30.
[5] a) U. D. Lengweiler, M. G. Fritz, D. Seebach, *Helv. Chim. Acta* **1996**, *79*, 670; D. Seebach, M. G. Fritz, *Int. J. Biol. Macromol.* **1999**, *25*, 217; b) "β-Peptide Conjugates: Syntheses, CD and NMR Investigations of β/α-Chimeric Peptides, of a DPA-β-Decapeptide and of a PEGylated-β-Heptapeptide": J. Gardiner, R. I. Mathad, B. Jaun, J. V. Schreiber, O. Flögel, D. Seebach, *Helv. Chim. Acta* **2009**, *92*, in print.
[6] a) H. Tadokoro, Y. Chatani, T. Yoshihara, S. Tahara, S. Murahashi, *Makromol. Chem.* **1964**, *73*, 109; b) Y. Takahashi, H. Tadokoro, *Macromolecules* **1973**, *6*, 672; c) Y. Takahashi, I. Sumita, H. Tadokoro, *J. Polym. Sci. Polym. Phys. Ed.* **1973**, *11*, 2113.
[7] See textbooks on stereochemistry. For a recent full paper on the subject see: I. V. Alabugin, T. A. Zeidan, *J. Am. Chem. Soc.* **2002**, *124*, 3175, and references therein.
[8] Not to be confused with the *gauche* or anomeric effect in molecules containing (:X–C–Y:) moieties, such as oligo- and polymethylene oxide: E. Juaristi, G. Cuevas *The Anomeric Effect*, CRC, Boca Raton, **1995**; S. Eppacher, G. Giester, J. W. Bats, C. R. Noe, *Helv. Chim. Acta* **2008**, *91*, 581.
[9] J. Dale, *Stereochemistry and Conformational Analysis*, Universitetsforlaget, Oslo, **1975**; *Stereochemie und Konformationsanalyse*, Verlag Chemie, Weinheim, **1978**, p. 98: *"Dieser… Helixtyp ist von einer…Form des Polyethylenglykols…her bekannt" (no reference!)*.
[10] E. Juaristi, *Introduction to Stereochemistry & Conformational Analysis*, Wiley, New York, **1991**, p. 301: *"…in polyoxyethylene the OCH_2CH_2O gauche arrangement is preferred" (reference to secondary literature only)*.
[11] A *Google Scholar* search ("polyethylene glycol structure") produced as first hit a 1962 paper that describes an investigation of PEG by polarized IR spectroscopy, with the following sentence in the Abstract: *"The most likely model (TGT) contains seven repeating units and two helical turns per fiber period of 19.25 Å"*: T. Miyazawa, K. Fukushima, Y. Ideguchi, *J. Chem. Phys.* **1962**, *37*, 2764. We thank Professor J. D. Dunitz for this information.
[12] *Ullmann's Encyclopedia of Industrial Chemistry*: http://mrw.interscience.wiley.com/emrw/9783527306732/home/.
[13] J. L. Koenig, A. C. Angood, *J. Poly. Sci. Polym. Phys. Ed.* **1970**, *8*, 1787.
[14] F. H. Allen, *Acta Crystallogr. Sect. B* **2002**, *58*, 380; I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson R. Taylor, *Acta Crystallogr. Sect. B* **2002**, *58*, 389.
[15] Version 5.30; November 2008, including updates 1–3.
[16] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, P. A. Wood, *J. Appl. Crystallogr.* **2008**, *41*, 466–470.
[17] Although the structure of **2** obtained by fiber diffraction is very similar to that of **1**, the structure of **2** obtained with freeze-dried particles is not helical, but, rather, a zig-zag chain (like the backbone of polyethylene); this highlights that structures can be source-dependent (cf. polymorphic modifications).

 © 2009 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim