Available online at www.sciencedirect.com
ScienceDirect
Tetrahedron Letters
Tetrahedron Letters 48 (2007) 2311–2313

QUIRAL: a computer program for the synthesis of chiral molecules from sugars

Jean-Marc Nuzillard* and Arnaud Haudrechy

Institut de Chimie Moléculaire de Reims, UMR CNRS, Université de Reims, BP 1039, F-51687 REIMS Cedex, France

Received 12 December 2006; revised 24 January 2007; accepted 29 January 2007
Available online 4 February 2007

Abstract—The QUIRAL computer program analyzes the 3D structure of a target organic molecule to find which sugar(s) can be used as a starting material for its synthesis. The program also proposes schemes for the preparation of rare or unavailable sugars whose chiral centers fit with those of the target molecule. Castanospermine, an anti-HIV natural compound is chosen as an example to illustrate what the QUIRAL program achieves.

The idea of computer assisted retrosynthetic analysis emerged several decades ago, but it still has little impact on organic chemists everyday life.The first computer program was proposed by Corey as early as 1969,[1] according to the guidelines defined by Vléduts.[2] This pioneering work was followed by others (Hendrickson,[3] Ugi,[4] Gasteiger,[5] see also reviews in Ref. 6). The most successful program was undoubtedly 'Logic and Heuristics Applied to Synthetic Analysis' (LHASA) that was developed at Harvard University by Corey.[7]

The synthesis of enantiopure molecules remains a challenge in organic synthesis. The main strategies in this field rely either on asymmetric catalysis or on the use of compounds derived from the natural chiral pool. In the latter approach, asymmetric centers from carefully selected starting materials are reused to elaborate those in the target molecules. Among naturally chiral compounds, carbohydrates are of particular importance because they generally present adjacent asymmetric centers that may be 'recycled' in target molecules that also contain adjacent asymmetric centers. This approach minimizes the waste of high value carbon atoms. Hanessian created the Chiron software which relates a complex structure to simpler chiral precursors, in order to guide chemists in the synthesis of enantiopure molecules.[8] We wrote the QUIRAL program with the same goal in mind so that a chemist may more quickly identify to which sugar(s) a synthetic target is structurally related. We chose to focus on sugars as starting materials, and gave ways to identify inversion of configurations (inter-relations between sugar families), keeping in mind the practical feasability.

In order to be as accurate as possible, a list of definitions is necessary. A $C^{\#}$ atom is a carbon bound to two carbon atoms, a hydrogen atom, and an X atom, X being oxygen, nitrogen, sulfur, or any halogen. A Q-sugar is an aldose molecule that contain from $p = 3$ to 6 carbon atoms and therefore, $n = p - 2$ adjacent $C^{\#}$ atoms. The aldehyde carbon of a Q-sugar is its A-end (anomeric end) and the primary alcohol is the NA-end (non anomeric end). In order to avoid confusion with conventional R and S labeling, T and U have been introduced to qualify the configuration of the $C^{\#}$ atoms (Fig. 1). In Q-sugars, R chiral centers are of the T type and S chiral centers are of the U type. The configuration of the chiral atoms in a Q-sugar is simply summarized by a list of its 'T or U's', arbitrarily starting from the $C^{\#}$ that is bonded to the A-end. With this convention, the reference of D-glucose is 'TUTT'.

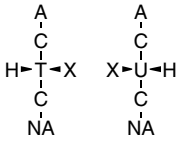


Figure 1. Definition of the T and U configuration of $C^{\#}$ atoms, as used by the QUIRAL program. In aldoses, T and U configurations are like R and S, respectively.

Keywords: Carbohydrates; Chiral pool; Computer; Retrosynthesis; Organic Synthesis.
* Corresponding author. Tel.: +33 3 26 91 82 10; fax: +33 3 26 91 35 96; e-mail: jm.nuzillard@univ-reims.fr

0040-4039/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetlet.2007.02.001

Q-reactant $\xrightarrow{\text{Q-reaction}}$ Q-subtarget $\xrightarrow{\text{Q-reaction}}$ Q-target $\Leftrightarrow$ Target

Scheme 1. General scheme for the creation of adjacent asymmetric centers in a given target molecule.

A target molecule contains one or more clusters of adjacent $C^{\#}$ atoms. By definition, a cluster is either linear or monocyclic. A linear cluster has two ends, one qualified as the A-end and the other one as NA-end, without any consideration of oxidation state. In a cyclic cluster, two adjacent $C^{\#}$ atoms are arbitrarily transformed in an A-end and an NA-end. A Q-target of a target molecule is a Q-sugar whose reference exactly matches the one of a cluster of $C^{\#}$ atoms within this target. The set of possible Q-targets is restricted to pentoses and hexoses.

A Q-reaction transforms a Q-sugar into another Q-sugar. In a general reaction scheme (Scheme 1), a Q-reactant is transformed into a Q-target, possibly with Q-subtarget(s) as intermediate(s), to obtain the desired asymmetric configurations in the target. Q-reactants and Q-subtargets are also Q-sugars. The list of Q-reactions includes: the inversion of a single $C^{\#}$ (e.g., Mitsunobu reaction), the inversion of two adjacent $C^{\#}$s (epoxide formation and opening), the exchange of the A-end and NA-end, the cleavage of one of the ends, and the addition of a carbon atom at one of the ends. For the latter reaction, the possible generation of a new $C^{\#}$ with either a T or a U configuration is considered. A Q-scheme is a succession of one or more Q-reactions that transforms a Q-reactant into a Q-target.

The QUIRAL program reads a 3D MDL (http://www.mdli.com) MOL file of the target molecule that is conveniently created by ChemDraw 3D (http://www.cambridgesoft.com). It identifies the clusters of $C^{\#}$ atoms and finds the list of possible Q-targets. When an X atom that is bonded to a $C^{\#}$ is not an oxygen, its configuration is automatically inverted because the introduction of the X atom is assumed to require a substitution reaction that inverts configuration. A particular Q-target may be very expensive or, more generally, not suitable to start with in a practical synthesis. Q-schemes that contain up to three Q-reactions are proposed by QUIRAL. Q-subtargets are indicated for Q-schemes that contain more than a single Q-reaction. The graphical user interface of the QUIRAL program allows the user to select the desired MOL file of a target, to rotate the molecule on screen, to select and visualize the position of a proposed Q-target within the target and to retrieve possible Q-schemes that lead to this Q-target.

Castanospermine **1** was chosen to exemplify what the QUIRAL program can do. The castanospermine target contains a linear cluster made of five $C^{\#}$ atoms. This cluster may conformationally be superimposed with four aldopentoses and three aldohexoses, as summarized in Table 1.

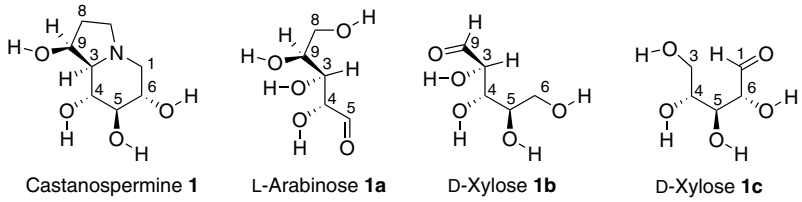
Table 1. Description of the Q-targets found by the QUIRAL program within castanospermine **1**

Q-target	A-end	$C^{\#}$s	NA-end	Reference
L-Xylose	8	9–3–4	5	TTU
L-Arabinose	5	4–3–9	8	TTU (≡ **1a**)
D-Xylose	9	3–4–5	6	TUT (≡ **1b**)
	1	6–5–4	3	TUT (≡ **1c**)
L-Xylose	6	5–4–3	9	UTU
	3	4–5–6	1	UTU
D-Gulose	8	9–3–4–5	6	TTUT
L-Glucose	6	5–4–3–9	8	UTUU
D-Idose	9	3–4–5–6	1	TUTU
	1	6–5–4–3	9	TUTU

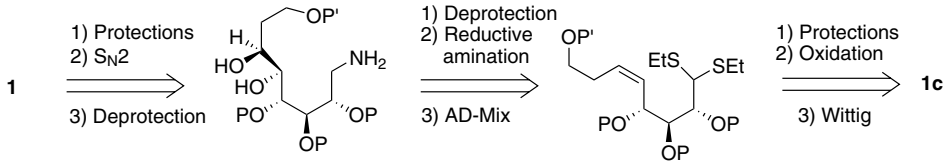
When analyzing the Q-targets suggested by QUIRAL for a possible retrosynthesis of castanospermine **1**, few chemists would choose unnatural L-glucose as the starting material inspite of its stereochemical resemblance with 4 of the 5 chiral centers in the target molecule. Clearly L-arabinose or D-xylose (Scheme 2) are the most cost effective chiral starting materials in Table 1 (atom numbers are those from the MOL file of **1**. The drawings of the Q-targets **1a**–**c** suggest how their $C^{\#}$ superimpose with those in **1**).

The major problem with L-arabinose **1a** as starting material is the controlled creation of asymmetric centers at contiguous C-5 and C-6 positions, along with the ring closures around the nitrogen atom.

D-Xylose may also be used for the synthesis of **1**, starting either from Q-target **1b** or **1c**. As in the case of L-arabinose **1a**, the challenges in using **1b** are the controlled creation of asymmetric centers at the C-9 and C-6 positions, along with the ring closures around the nitrogen atom. The **1c** Q-target seems analogous to **1a**, as it also requires the controlled creation of asymmetric centers at contiguous C-3 and C-9 positions, with bond formations with the nitrogen atom. However, reductive amination on the reducing end of **1c** is an attractive way to add a nitrogen functionality. A subsequent attack at the mesylate-activated C-3 and C-8 positions would constitute an elegant approach to castanospermine **1** (Scheme 3). Interestingly, among the many reported



Scheme 2. The structure of castanospermine **1** and of synthetically pertinent Q-targets.



Scheme 3. A possible retrosynthetic scheme for castanospermine **1** starting from D-xylose **1c**.

syntheses of **1**, a similar approach by Mulzer in 1992 follows the **1c** strategy.[9]

Other target molecules may not lead to the identification of at least one easily accessible sugar. In such situations, QUIRAL can be used to find Q-schemes that lead to the proposed Q-target(s) in one to three Q-reactions.

We hope that this simple example of castanospermine **1** shows that QUIRAL is an interesting tool for the retrosynthetic analysis of natural products in which many contiguous asymmetric centers are present.

Acknowledgment

We thank Dr. K. Plé for helpful discussions.

Supplementary data

The QUIRAL program can be downloaded from http://www.univ-reims.fr/LSD/JmnSoft/Quiral. Screenshots of QUIRAL in action are available free of charge via Internet at http://www.acs.org. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tetlet.2007.02.001.

References and notes

1. Corey, E. J.; Wipke, W. T. *Science* **1969**, *166*, 178.
2. Vléduts, G. E. *Inf. Stor. Retr.* **1963**, *1*, 117.
3. Hendrickson, J. B. *ACS Symp. Ser.* **2002**, *823*, 127; Hendrickson, J. B. *CHEMTECH* **1998**, *28*, 35; Hendrickson, J. B. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 1286.
4. Ugi, I.; Bauer, J.; Blomberger, C.; Brandt, J.; Dietz, A.; Fontain, E.; Gruber, B.; von Scholley-Pfab, A.; Senff, A.; Stein, N. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 3.
5. Ihlenfeldt, W. D.; Gasteiger, J. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2613.
6. Todd, M. H. *Chem. Soc. Rev.* **2005**, *34*, 247; Barone, R.; Chanon, M.. In *Chemoinformatics. From data to knowledge*; Gasteiger, J., Engel, T., Eds.; Wiley-VCH: Weinheim, 2003; Vol. 4, pp 1428–1456.
7. Corey, E. J.; Long, A. K.; Lotto, G. I.; Rubenstein, S. D. *Recl. Trav. Chim., Pays-Bas* **1992**, *111*, 304; Corey, E. J.; Long, A. K.; Rubenstein, S. D. *Science* **1985**, *228*, 408.
8. Hanessian, S.; Botta, M.; Larouche, B.; Boyaroglu, A. *J. Comput. Chem. Inf. Sci.* **1992**, *32*, 718.
9. Mulzer, J.; Dehmlow, H. *J. Org. Chem.* **1992**, *57*, 3194.