

DETECTION OF MONOSACCHARIDE TYPES FROM COORDINATES

MASANORI ARITA^{1,2,3}
arita@k.u-tokyo.ac.jp

TOSHIAKI TOKIMATSU¹
tokimatsu@cb.k.u-tokyo.ac.jp

¹*Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo and PRESTO JST, 5-1-5 CB05 Kashiwanoha, Kashiwa, 277-8561 Japan*

²*Plant Science Center, RIKEN, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, 230-0045 Japan*

³*Institute of Advanced Biosciences, Keio Univ., 14-1 Baba-cho, Tsuruoka, 997-0035 Japan*

Almost half of biological molecules (proteins and metabolites) are extrapolated as glycosylated within cells. Detection of glycosylation patterns and of attached sugar types is therefore an important step in future glycomics research. We present two algorithms to detect sugar types in Haworth projection, *i.e.*, from x - y coordinates. The algorithms were applied to the database of flavonoid and identified backbone-specific biases of sugar types and their conjugated positions. The algorithms contribute not only to bridge between polysaccharide databases and pathway databases, but also to detect structural errors in metabolic databases.

Keywords: algorithm, flavonoid, monosaccharide, stereochemistry

1. Introduction

Glycosylation is a major post-transcriptional and post-translational modification for biological molecules. Two thirds of all proteins are extrapolated as glycosylated to become functional [2], and the same is true for many secondary metabolites. For example, among 6,850 flavonoid species structurally identified to date, as much as 50 % were found in glycosylated form [16]. Despite such universality, a common, computational notation for oligo- and polysaccharides that can be used for research articles and databases has been missing in glycomics. The situation clearly decelerated the development and integration of carbohydrate information resources when compared against other molecular information such as genome and protein sequences. A notable exception is the Glycan database, which uses a graphical layout for polysaccharide molecular structures [12].

Although the structure of carbohydrates is much more complicated than linear DNAs and protein sequences, linear systematic codes have been already proposed [3,5]. In time, a particular code would become the universal standard, and the next problem is how to translate currently available carbohydrate information in different formats into such a systematic code. Obviously manual conversion is time-consuming and error-prone, and there is an urgent demand for a computational solution. The purpose of this contribution is to provide an automated technique that can convert conventional descriptions into a systematic code, in preparation for a possible standard description of saccharide structures. The crucial step is the recognition of major drawing styles to describe the stereochemistry of cyclic monosaccharides; oligo- and polysaccharides are their

repetitions and can be treated similarly. Moreover, less than 5-carbon or more than 6-carbon monosaccharides such as tetrose and heptose are rare as natural modifiers. For this reason we focus here on the recognition of 5- or 6-carbon cyclic monosaccharides. In drawing molecular structures, frequently used styles are the following: Mills depiction, Haworth projection, and its variation for three-dimensional view. The guidelines of each drawing style are formally recommended by IUPAC-IUBMB as Nomenclature of Carbohydrates 2-Carb-5 [9].

The paper is organized as follows. Since depiction scheme of sugar structures is not well known, we introduce the major drawing styles and type of sugars after this introduction. In Chapter 2, the formal detection scheme is presented. The procedure was implemented in Java and was tested on our Flavonoid databases [16]. The result is introduced in Chapter 3, followed by the conclusion and future work of this study.

1.1. The Fischer projection

The standard method in teaching stereochemistry of monosaccharides is the Fischer projection. It represents every stereocenter as a cross. The horizontal line represents bonds extending above the paper plane, and the vertical line represents bonds extending below the plane (Figure 1). The IUPAC numbering of 6 carbons in hexose starts from top (C-1) to bottom (C-6) direction. This projection is used in textbooks only, but we introduce it for easier understanding of the following other descriptions.

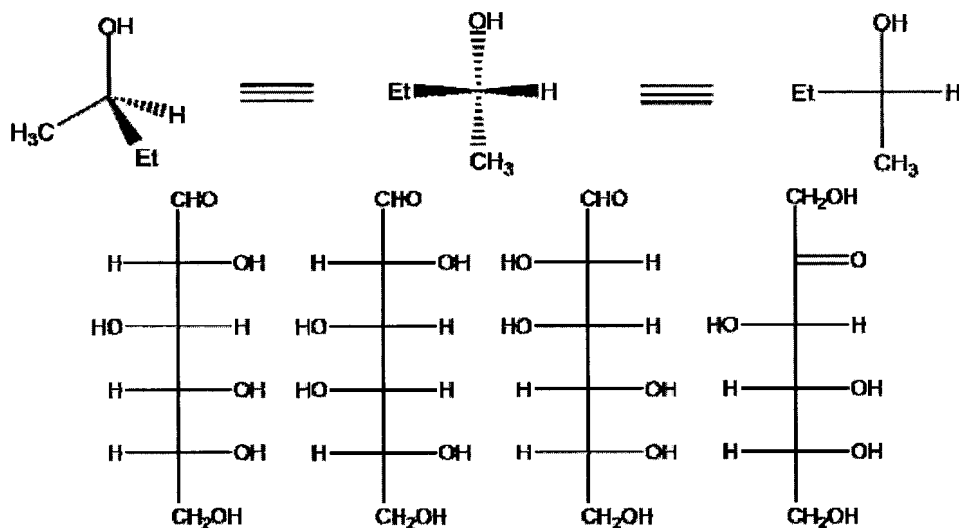


Figure 1. Fischer projection and four hexose examples. From the left, D-glucose, D-galactose, D-mannose and D-fructose. The topmost carbons are C-1, and bottommost are C-6. The figures are taken from <http://www.metabolome.jp/doc/lectures/biochem/sugar/>.

1.2. The Mills depiction

The Mills method is widely used in organic chemistry and molecular biology, including web-based databases such as the Kyoto Encyclopedia of Genes and Genomes database (KEGG) [11], PubChem database [15], and metabolomics web-services such as Biological Magnetic Resonance Data Bank (BMRB) [4]. In the method, the ring of monosaccharide is set on the paper plane; thickened black bonds denote chemical substituents projected above the plane, and dashed bonds, beneath the plane (Figure 2). Note that this method requires up/down information for bonds *in addition to* the x - y coordinates for atoms.

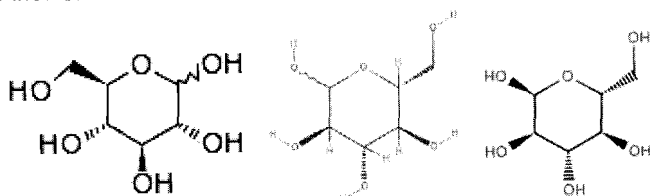


Figure 2. Mills depiction of the same D-glucose from the ligand section of the KEGG database (left), PubChem database (center), and the metabolomics section of the BMRB (right).

1.3. The Haworth projection

In glyco- and plant biology, a stereoscopic depiction is far more preferred. In Haworth projection, the sugar ring is placed almost perpendicular to the paper plane, and is viewed from above so that closer atoms and bonds are drawn below the farther components. The orientation usually (but not always) conforms to a clockwise numbering of the IUPAC ring atoms. In Figure 3, for example, CH₂OH group above the ring is C-6. Oxygen is usually placed behind at the right-hand side. The bonds are not necessarily thickened as in Figure 3, and indeed Chemical Abstract Service does not use thickening in its bond description [6].

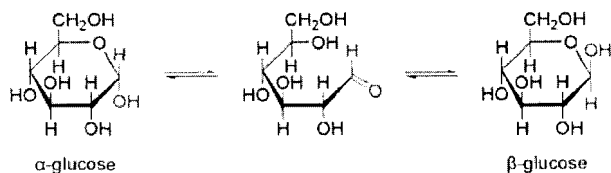


Figure 3. Conversion between *alpha*-D-glucose and *beta*-D-glucose in the Haworth projection. The figure was taken from Wikipedia (Japanese version).

Although the original Haworth projection depicts the ring as planar, it is a highly skewed representation of the original molecular structure. For more precise description, its conformational variant is used to show the boat or chair form. Chemical substituents are

called *equatorial* and *axial* when they extend on the plane of the ring, or perpendicular to the ring plane, respectively (Figure 4-a).

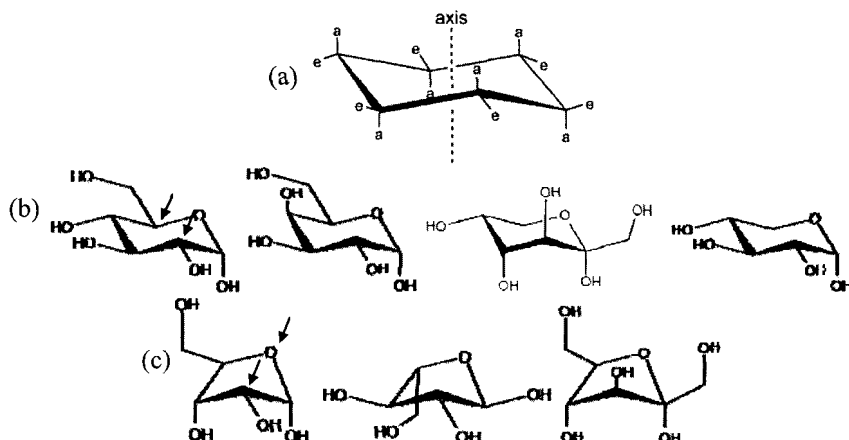


Figure 4. (a) Equatorial and axial directions of the ring, shown with *e* and *a* respectively. (b) Conformations of pyranose sugars. Chair conformation of *alpha*-D-glucose, *alpha*-D-galactose, *alpha*-D-fructose, and *alpha*-D-xylose. In *beta* forms, the hydroxyl group at the lower right corner will become equatorial. We call the two closest carbon atoms (shown with arrows for glucose) as 'concave positions'. (c) Conformations of furanose sugars. From the left, *alpha*-D-ribose, *alpha*-L-arabinose, and *alpha*-D-fructose. For furanose, one carbon and ring oxygen (shown with arrows for ribose) are the concave positions. Note the change between fructopyranose in (b) and fructofuranose in (c), both of which have the same parity. Figures are taken from the website <http://www.metabolome.jp/doc/lectures/biochem/sugar/>.

Large substituents are thermodynamically more stable when they are equatorial, because of more space in the equatorial positions. By far the most naturally abundant three monosaccharides (glucose, galactose and mannose) have all but one substituent in the equatorial positions in their *beta* forms.

1.4. Types of monosaccharides

Most monosaccharides exist as equilibrium mixtures of *alpha*- and *beta*-forms. If we ignore *alpha* and *beta* configuration, four chiral positions for hexoses and three for pentoses exist, theoretically. Each chiral configuration can be represented by a boolean parity corresponding to the Cahn-Ingold-Prelog priority rule (R/S notation) or the D/L nomenclature based on the stereochemistry of glyceraldehydes (visit Wikipedia for these basic rules at <http://en.wikipedia.org/wiki/Stereochemistry>). Chiral configurations of different monosaccharides are shown in Table 1. The molecular structures for all these monosaccharides in Fischer and Haworth projections are available at our website

<http://www.metabolome.jp/doc/lectures/biochem/sugar/> in PDF. In the next section, we introduce the procedure to recover these parities from x - y coordinates.

Table 1. Stereo parities of four chiral positions (from C5 to C2) for D-hexose and D-pentose sugars. Corresponding parity for L-sugars can be obtained by flipping all 1 and 2. The parity 1/2 corresponds to right/left orientation of the Fischer projection from the bottom up (see also Figure 1). Nonchiral positions are skipped and the keto group is represented by ‘_’ to distinguish hexoses and pentoses. This parity description from the bottom-up is our original notation.

Sugar Type				
aldohexose (8 patterns)	1111 allose	1121 glucose	1211 gulose	1221 galactose
aldohexose (8 patterns)	1112 altrose	1212 idose	1122 mannose	1222 talose
ketohexose (4 patterns)	111_ psicose	112_ fructose	121_ sorbose	122_ tagatose
aldopentose (4 patterns)	111 ribose	121 xylose	112 arabinose	122 lyxose
ketopentose (2 patterns)	11_ ribulose	12_ xylulose		

2. Algorithms for detecting stereo parities of cyclic monosaccharides

We assume that x - y coordinates and the topology of molecular structure (*i.e.* atoms and their connections) are available. Such structural information includes MOL format and PDB format. SMILES format does not conform to this criterion because it has no coordinates. We also assume that the cyclic parts in the structure are already detected by a standard method in computational chemistry [7].

The detection procedure differs for the Mills depiction, the Haworth projection, and its conformational variant. Since the Mills depiction requires parity information in addition to x - y coordinates, we skip its detection procedure here. The parity information essentially corresponds to the 1/2 codes in Table 1, although a cumbersome parity conversion may be needed depending on the file format of molecular structures. For Mills depiction, therefore, there is no essential need for considering coordinates for sugar detection. We focus here on the remaining two depictions. First, to distinguish the Haworth projection and its conformational variant, distances between all ring atoms that are not directly bonded are computed. In the Haworth projection, there is no single shortest distance because the ring structure is symmetric. An unsymmetrical case is the conformational variant, and we call the closest atom pair ‘concave positions’ as will be explained in Section 2.3.

2.1. General Strategy

Some hexose sugars may form both pyranose (6-member) and furanose (5-member) rings. We use the notation ‘C- x ’ ($1 \leq x \leq 6$) to refer to the IUPAC carbon positions. Since C-1 carbon (the terminal carbon at the most oxidized side) may be either inside or outside of the ring atoms, our detection algorithm starts from C-5 carbon backward. (Note that C-6 carbon is never included in the ring component for hexoses, and there is no C-6 for pentoses. Starting from C-5 is thus simpler than from C-1 for computational recognition.) To designate ring positions we introduce the following functions.

$\text{prev}(\text{C-}x)$:= the previous atom in the ring in the IUPAC numbering system;
 $\text{succ}(\text{C-}x)$:= the next atom in the ring in the IUPAC numbering system;
 $\text{angle}(x, y, p)$:= the angle between chemical bond $p-x$ and $p-y$ in radian. Its return value is between $-\pi$ and π .

2.2. Detecting the Haworth projection

We introduce the basic procedure using a computer program pseudo code in `courier` font. Abstract functions are shown in English (times-roman). The actual program code must detect the type of chemical substituents (*e.g.* hydroxyl, keto, amino or other groups) to achieve the final identification of sugars.

Algorithm 2.2

```
int[] parity = new int[4];
q = (Is pentose) ? C-4 : C-5;
for (Carbon position p iterated from q to C-2)
{
  br = Chemical substituent at p;
  pv = prev(p);
  nx = succ(p);
  parity[p] = (angle(pv, br, p) * angle(pv, nx, p) < 0)
    ? 1 : 2;
  if (Is p on the left of pv)
    parity[p] = (parity[p] == 1) ? 2 : 1;
}
parity[q] = (parity[q] == 1) ? 2 : 1;
If (parity[q] == 2)
  for (Carbon position p iterated from q to C-2)
    parity[p] = (parity[p] == 1) ? 2 : 1;
```

Observation: The algorithm 2.2 correctly computes the stereo parities introduced in Table 1 for both hexoses and pentoses.

Proof: Irrespective of the orientation of the Haworth projection, *i.e.*, either clockwise or anticlockwise numbering of ring atoms, the conditional at the line (1) returns 1 when the chemical substituent (br) is above the plane of the ring in its upper half of the depiction, and when br is below the plane of the ring in its lower half. Therefore if all substituents

are below the plane, the algorithm computes the `parity = { 1, 1, 1, 1 }` in the `for` loop after correction of the line (2). Only C-5 must be treated differently. Since the true `prev` position of C-5 is C-6, not the ring oxygen as we defined in the algorithm, the computed result must be flipped as in the line (3). For example, the algorithm produces `parity = { 1, 1, 2, 1 }` for D-glucose in Figure 3, which is the same parity as in Table 1. Finally, the line (4) takes care of all corresponding L-forms.□

Corollary: The algorithm 2.2 is valid even when the ring is in all other, rotated or flipped orientations.

Proof: It is not hard to see the parity will not change even when the ring is rotated. Let us consider the flipped situation. Although the line (2) will reverse the parity, since all chemical substituents become upside down, both changes cancel out to produce the same parity. Since the conditional at the line (1) returns the same result even when the ring is flipped over, the final result remains the same.□

2.3 Detecting the conformational variant

The algorithm 2.2 does not work properly for the conformational variant of the Haworth projection such as in Figure 4-b (please manually simulate the algorithm to understand). Since chemical substituents are drawn to show equatorial/axial positions, they must be detected in relation with the plane of the ring, not with the adjacent ring atoms. In the conformational description, we specify two ring atoms as *concave positions* for hexoses and pentoses (Figure 4). These positions are placed in the middle column of the description and bonds extending from them are always equatorial to the ring plane. To use such positions we introduce the following functions. Note that a pentose can take both pyranose and furanose (6- and 5-member rings) forms.

```
concave(C-x) := return (whether C-x is at the middle position of the ring);
abs(x) := return (x >= 0) ? x : -x;
```

Algorithm 2.3

```
int[] parity = new int[4];
q = If pentose then C-4 else C-5;
for (Carbon position p iterated from q to C-2)
{
  br = Chemical substituent at p;
  pv = prev(p);
  nx = succ(p);
  if (concave(p))
    parity[p] = (angle(pv, br, p) * angle(pv, nx, p) > 0)
      ? 1 : 2; (1)
  else {
    cv = adjacent concave atom, either pv or nx; (2)
    parity[p] = (abs(angle(cv, br, p)) < 2π/3) ? 2 : 1 (3)
  }
}
```

}

$r =$ In pyranose, thirdly scanned atom; in furanose, secondly scanned atom. (4)

$\text{parity}[r] = (\text{parity}[r] == 1) ? 2 : 1;$

If ($\text{parity}[q] == 2$) (5)

for (carbon position p iterated from q to C-2)

$\text{parity}[p] = (\text{parity}[p] == 1) ? 2 : 1;$

Observation: The algorithm 2.3 correctly computes the stereo parities introduced in Table 1 for hexoses and pentoses.

Proof: The algorithm without the line (4) computes the parity 1 for equatorial, and 2 for axial positions. Note that *beta*-D-glucose (parity 1121) has all its substituents in equatorial position. Similarly, *beta*-D-xylose (parity 121) also has equatorial substituents only (Figure 4). In order to produce the flipped parity, 2 in this case, the line (4) is introduced. We cannot simply use C-3 position instead of r in the line (4) because hexoketose can form both pyranose and furanose (see fructose in Figure 4, for example). The position to be flipped corresponds to the lower left corner of the ring in the configurations of Figure 4, which depends on the size of rings, not on being hexose or pentose. Configurations for all other pentoses and hexoses can be rationalized in relation with the case of glucose, xylose and fructose.□

Corollary: The algorithm 2.3 is valid even when the ring is in all other, rotated or flipped orientations.

Proof: Let us consider *beta*-D-glucose and *beta*-D-xylose again. The algorithm 2.3 recognizes equatorial positions as parity 1 no matter how the structure is rotated or flipped. Because the position to be corrected in line (4) is invariant of rotation or flipping, the final result remains the same.□

Note: The algorithm 2.3 does **not** distinguish D and L series in hexoses. Although the line (5) can deal with L-forms of pentoses, whose ring orientation does not change between naturally occurring D and L series, the L-forms of hexoses change their ring conformation as in Figure 5. Since our algorithm does not consider whether the ring oxygen is at a concave position or not, it can not distinguish D and L series of hexoses. Checking whether the oxygen is concave or convex is necessary (and sufficient) for distinguishing D and L hexoses in Figure 5, but its process becomes highly complex when the ring may be rotated.

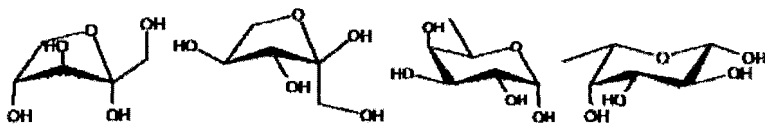


Figure 5. Difference of D and L series. From the left, *alpha*-D-xylulose, *alpha*-L-xylulose (pentose), *alpha*-D-fucose, and *alpha*-L-fucose (hexose).

3. Application to structural data in databases

Flavonoid is a class of secondary metabolites with C6-C3-C6 skeleton derived from phenylpropanoid-acetate pathway [1,8]. Beneficial effect of fruits and vegetables is often attributed to the anti-inflammatory and antioxidant activities of flavonoid, and it has drawn much attention in industry and academia. As shown in the Introduction section, 50 % of registered molecules are glycosides. Since sugars must be attached stepwise by special enzymes called glycosidases, there is a bias for the type of attached monosaccharides in each plant species. Such bias should reflect the evolutionary history of glycosidases, and therefore, of plant families. The algorithms in Section 2 were implemented in Java and were applied to total 6850 natural flavonoid species accessible from our Flavonoid Viewer software at

<http://www.metabolome.jp/software/FlavonoidViewer/>[†].

This dataset was manually collected from literatures including handbooks of flavonoid and research papers [1,6,8] in collaboration with Kanaya Laboratory in Nara Institute of Science and Technology. The curated structural information is also accessible from the KNApSACk database of plant secondary metabolites at <http://kanaya.aist-nara.ac.jp/KNApSACk/>.

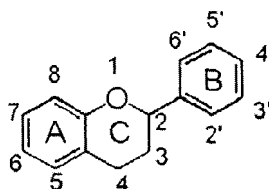


Figure 6. The backbone structure of flavonoid. For isoflavonoid, the phenol ring B is attached at the position 3, not 2. For neoflavonoid, the phenol ring B is attached at the position 4.

In our Flavonoid Viewer software, all molecules are classified into the following 9 categories according to their structure and biosynthetic origin: 1) chalcone and aurone, 2) flavanone, 3) flavone, 4) dihydroflavonol, 5) flavonol, 6) flavan, 7) anthocyanin, 8) isoflavonoid, and 9) neoflavonoid. For each of these categories, we identified monosaccharide types that are directly attached to their backbone structure.

3.1. Pattern of glycosylation

Major monosaccharides directly attached to the flavonoid backbone are listed in Table 2. From the Table, we can immediately tell the uniqueness of anthocyanins. Most common are their 3-O-glucosides, 3-O-galactosides, and 5-O-glucosides. The rest of the

[†] The software is freely accessible, but it is unpublished and the structural data are currently not downloadable in a bulk. Users can search and view, however, all structural data.

categories are much less glycosylated, but still show category-dependent characters. In flavones, 6-C-glucosides and 8-C-glucosides are observed. These molecular species (e.g. isovitexin C-glucosides) are mainly identified in caryophyllaceae (carnation family), but also occur in many higher plants such as fabaceae (bean family), gentianaceae (gentian family), or passifloraceae (passion flower family). The C-glycosylation should occur after oxidation of flavanones because the C-glucosides are not common in flavanones. Another character is the abundance of 7-O-arabinosides. These observations coincide with a previous report [14]. The characteristic of flavonols is 3-O- and 7-O-rhamnosylation. The abundance of 3-glycosides in flavans and anthocyanins in Table 2 suggests either these modifications occur prior to reduction of flavonols into flavans, or the substrate specificity of 3-O-rhamnosidase is not tight (the latter is the current consensus of experts). Such observations have been partially described previously [10], but this is the first solid statistical result obtained from around 7,000 natural molecular structures identified to date. In summary, our algorithms produced statistics only from molecular structures, and the obtained results coincided with previous observations by experts.

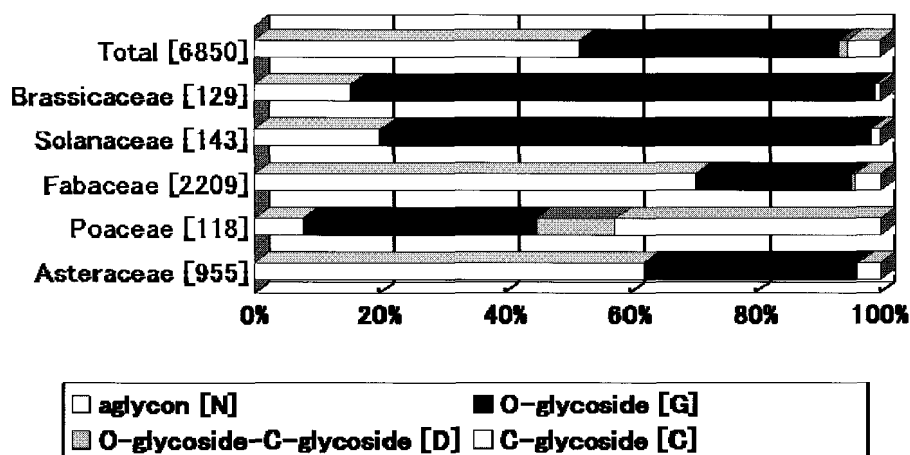


Figure 7. The glycosidation patterns for major plant families. Aglycon means backbone structure without sugars, and O-glycoside-C-glycoside means structure with both O- and C-glycosides. See main text for plant family names.

3.2. Pattern of Glycosylation in Plant Families

Using the species-molecule relationships obtained from the KNApSAcK database for plant secondary metabolites [13], we can obtain the distribution of flavonoid categories in plant families. As shown in Figure 7, we could confirm the following observations: 1) C-glycosides are abundant in poaceae (rice family), whereas O-glycosides are major in other representative plant families such as asteraceae (chrysanthemum family),

solanaceae (eggplant family), brassicaceae (cabbage family), and fabaceae (bean family);
 2) More than 95 % of isoflavonoids occur in fabaceae (statistical data not shown).

Table 2. Summary of the types of monosaccharides directly attached to the backbone in Figure 6. Components of less than 2% are ignored. The value in parentheses is the number of molecules currently registered in the software. All structures are accessible on the Internet. Abbreviations: Glc ... D-glucose; Gal ... D-galactose; Rha ... D-rhamnose.

Flavonoid Category	Monosaccharide Type at Each Position
chalcone and aurone (690)	Position 5: Glc 2% Position 7: Glc 6%; Gal 3%
flavanone (698)	Position 4': Gal 2% Position 5: Glc 2% Position 7: Glc 8%; Gal 4%
flavone (1475)	Position 4': Gal 4% Position 6: Glc 9%; Gal 3% Position 7: Glc 14%; Gal 4%; L-arabinose 2% Position 8: Glc 6%; Gal 3%
dihydroflavonol (272)	Position 3: Rha 6%; Glc 4% Position 7: Glc 4%
flavonol (1942)	Position 4': Glc 2% Position 3: Glc 21%; Gal 11%; Rha 10% Position 7: Glc 10%; Rha 3%; Gal 2%
flavan (289)	Position 3: Glc 2%; Rha 2% Position 5: Glc 5% Position 7: Glc 5%
anthocyanin (486)	Position 3': Gal 2%; Glc 2% Position 5': Gal 3% Position 3: Glu 64%; Gal 20%; Rha 2% Position 5: Glu 33%; Gal 8% Position 7: Gal 5%; Glc 4%
isoflavonoid (916)	Position 7: Glc 9% Position 8: Glc 2%
neoflavonoid (82)	Position 5: Glc 7%; Gal 3%

4. Conclusion

In the past decade we have seen an almost explosive expansion of biological databases. We demonstrated here that characteristics of plant families can be obtained from unbiased statistic of species-molecule information computationally obtained from well-curated structural databases. Database accuracy is of crucial importance. In this perspective, our computational procedure contributes not only to sort out polysaccharide structures but also to detect input errors (e.g. molecular names or structures) in glycomics databases. We propose that all polysaccharide information in databases be checked computationally to improve its data quality and to transfer its value to other databases with pathway and other metabolic information. Our future work is therefore to extend the structural recognition system to major acylation patterns in secondary metabolites such as acetyl, caffeic, p-coumaric, ferulic, gallic, malic, and malonic acids.

Acknowledgments

We thank Kazuhiro Suwa for the implementation of Flavonoid Viewer, and Yukiko Nakanishi and Yukiko Fujiwara for curating structural data of flavonoids. We also thank Yoko Shinbo and Prof. Shigehiko Kanaya for providing us the structure data of flavonoids and species-molecule relationships from the KNApSACk database. This work is a part of joint research with Kanaya Laboratory, and is supported by Grant-in-Aid for Scientific Research on Priority Areas "Systems Genomics" from the Ministry of Education, Culture, Sports, Science and Technology of Japan:

References

- [1] Andersen, O.M., Markham, K.R. (eds.) *Flavonoids: chemistry, biochemistry and applications*. CRC Press, 2006.
- [2] Apweiler, R. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim. Biophys. Acta* 1473:4–8, 1999
- [3] Banin, E., Neuberger, Y., Altshuler, Y., Halevi, A., Inbar, O., Dotan, N. and Avinoam, D. A novel linear code nomenclature for complex carbohydrates, *Trends Glycosci. Glycotechnol.* 14(77):127-137, 2002.
- [4] BMR Data Bank: <http://www.bmrb.wisc.edu/metabolomics/>
- [5] Bohne-Lang, A., Lang, E., Forster, T., and von der Lieth, C. W. LINUCS: linear notation for unique description of carbohydrate sequences, *Carbohydr. Res.* 336: 1-11, 2001.
- [6] Chemical Abstract Service: <http://www.cas.org/>
- [7] Gasteiger, J. (ed.) *Handbook of Chemoinformatics: From Data to Knowledge Vol. 1*, John Wiley & Sons Inc. 2003.
- [8] Harborne, J. B. and Baxter, H. (eds.) *The handbook of natural flavonoids*. John Wiley & Sons, 1999.
- [9] IUPAC-IUBMB Nomenclature of Carbohydrates 2-Carb-5: <http://www.chem.qmul.ac.uk/iupac/2carb/05.html>
- [10] Iwashina, T. The structure and distribution of the flavonoids in plants. *J. Plant Res.* 113, 287-299, 2000.
- [11] KEGG Database: <http://www.genome.ad.jp/kegg/>
- [12] KEGG Glycan database: <http://www.genome.jp/kegg/glycan/>
- [13] KNApSACk database <http://kanaya.aist-nara.ac.jp/KNApSACk/>
- [14] Mastenbroek, O., Prentice, H.C., Kamps-Heinsbroek, R., van Brederode, J., Niemann, G. J., van Nigtevecht, G. Geographic trends in flavone-glycosylation genes and seed morphology in European *Silene pratensis* (Caryophyllaceae) *Plant Systematics and Evolution* 141(3-4), 257-271, 1983.
- [15] PubChem Database: <http://pubchem.ncbi.nlm.nih.gov/>
- [16] Tokimatsu, T. and Arita, M. unpublished results. Review in Japanese is available as Tokimatsu, T. and Arita, M. Viewing flavonoid through metabolic maps. *Saibou Kougaku*, 25(12) 1388-1393, 2006.