MilkOligoThesaurus, a dataset of mammalian milk oligosaccharide synonyms

Mathilde Rumeau, François Fenaille, Agnès Girard, Valentin Loux, Mouhamadou Ba, Claire Nédellec, Louise Deleger, Robert Bossy, Sophie Aubin, Christelle Knudsen, et al.

Data in Brief 54 (2024) 110404

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

MilkOligoThesaurus, a dataset of mammalian milk oligosaccharide synonyms

Mathilde Rumeau[a], François Fenaille[b], Agnès Girard[c], Valentin Loux[d,e], Mouhamadou Ba[d,e], Claire Nédellec[e], Louise Deléger[e], Robert Bossy[e], Sophie Aubin[f], Christelle Knudsen[a], Sylvie Combes[a,*]

[a] GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France
[b] Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), MetaboHUB, 91191 Gif sur Yvette
[c] INRAE, LPGP, 35000 Rennes, France
[d] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE Bioinformatics Facility, Jouy-en-Josas, France
[e] Université Paris-Saclay, INRAE, MaIAGE, Jouy-en-Josas, France
[f] INRAE, DipSO, 42 rue Georges Morel, 49070 Beaucouzé, France

ARTICLE INFO

Article history:
Received 3 October 2023
Revised 22 February 2024
Accepted 5 April 2024
Available online 9 April 2024

Dataset link: MilkOligoThesaurus: A milk oligosaccharide thesaurus (HoloOLIGO project) (Original data)

Keywords:
Chemical nomenclature
Normalized milk oligosaccharide name
Milk oligosaccharide monoisotopic mass
Milk oligosaccharide monosaccharide composition
Oligosaccharide isomer name
Vocabulary extraction
Systematic names

ABSTRACT

There is a growing interest in milk oligosaccharides (MOs) because of their numerous benefits for newborns' and long-term health. A large number of MO structures have been identified in mammalian milk. Mostly described in human milk, the oligosaccharide richness, although less broad, has also been reported for a wide range of mammalian species. The structure of MOs is particularly difficult to report as it results from the combination of 5 monosaccharides linked by various glycosidic bonds forming structurally diverse and complex matrices of linear and branched oligosaccharides. Exploring the literature and extracting relevant information on MO diversity within or across species appears promising to elucidate structure-function role of MOs. Currently, given the complexity of these molecules, the main issues in exploring literature to extract relevant information on MO diversity within or across species relate to the heterogeneity in the way authors refer to these molecules. Herein, we pro-

* Corresponding author.
E-mail address: sylvie.combes@inrae.fr (S. Combes).
Social media: @vloux (V. Loux)

https://doi.org/10.1016/j.dib.2024.110404

vide a thesaurus (MilkOligoThesaurus) including the names and synonyms of MOs collected from key selected articles on mammalian milk analyses. MilkOligoThesaurus gathers the names of the MOs with a complete description of their monosaccharide composition and structures. When available, each unique MO molecule is linked to its ID from the NCBI PubChem and ChEBI databases. MilkOligoThesaurus is provided in a tabular format. It gathers 245 unique oligosaccharide structures described by 22 features (columns) including the name of the molecule, its abbreviation, the chemical database IDs if available, the monosaccharide composition, chemical information (molecular formula, monoisotopic mass), synonyms, its formula in condensed form, and in abbreviated condensed form, the abbreviated systematic name, the systematic name, the isomer group, and scientific article sources. MilkOligoThesaurus is also provided in the SKOS (Simple Knowledge Organization System) format. This thesaurus is a valuable resource gathering MO naming variations that are not found elsewhere for (i) Text and Data Mining to enable automatic annotation and rapid extraction of milk oligosaccharide data from scientific papers; (ii) biology researchers aiming to search for or decipher the structure of milk oligosaccharides based on any of their names, abbreviations or monosaccharide compositions and linkages.

Specifications Table

Subject	Biochemistry
Specific subject area	Mammalian milk analysis, milk oligosaccharide names, milk oligosaccharide monosaccharide composition, milk oligosaccharide isomer, milk oligosaccharide structure, biochemistry
Data format	Raw, Analyzed
Type of data	Table, Figure
Data collection	A total of 245 milk oligosaccharide names were collected from 11 selected scientific papers. To avoid confusion or misspellings in the naming of the MOs, selected papers had to meet the following inclusion criteria: to report an extensive list of milk oligosaccharides AND to report their monosaccharide composition AND to detail their isomeric structures using the Symbol Nomenclature for Graphical Representations of Glycans or Oxford symbol nomenclature. Descriptive information such as chemical formula and monoisotopic mass was obtained from chemical databases (ChEBi and PubChem) and linked with unique identifiers.
Data source location	GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France
Data accessibility	Repository name: https://entrepot.recherche.data.gouv.fr/dataverse/inrae Data identification number (DOI) https://doi.org/10.57745/RA5DAC Direct URL to data: https://doi.org/10.57745/RA5DAC

1. Value of the Data

- The first purpose of the thesaurus is to be used to automatically detect and normalize MO mentions across the literature. In order to gain knowledge on the role of MOs it is necessary to find, visualize and analyze MO patterns that are scattered in thousands of scientific papers.

MilkOligoThesaurus covers a wide range of MO names and synonyms, and associates them to a restricted and standardized terminology. Text-mining methods can use the standardization proposed by MilkOligoThesaurus to normalize the variability of MO mentions in natural language texts, thus facilitating data exchange and integration. The thesaurus is designed to be human and computer-readable, making it easy to share and use.

- Text mining is an automatic process that consists of automatic steps: 1) collecting relevant documents, 2) recognizing in the documents' content the named entities of MO type and all other relevant types (e.g. species, lactation stages), 3) normalizing the named entities by the reference, i.e. linking the MO mentions to the entries of the MilkOligoThesaurus. On one hand, the MilkOligoThesaurus is the reference for MO normalization, on the other hand it serves as a lexical resource to support named entity recognition and normalization. Several strategies were developed and investigated to exploit external lexical and semantic resources to improve machine learning models. These strategies include thematic masking [1], named entity recognition by distant supervision [2], and ontology-based normalization [3].
- The biological roles of MOs depend mainly on their structure. Numerous studies have been carried out to elucidate these structure–function relationships. However, the heterogeneity in the way authors have named and reported these molecules is the main pitfall for academic and industrial researchers in having a clear overview of the published results. MilkOligoThesaurus is a valuable resource that will help the scientific community to navigate and explore the various structural variations of MOs and easily identify unique biologically significant MOs refered to with variable names or identifiers.
- The thesaurus will enable academic and industrial researchers to identify the usual way of naming MOs and provide them with a direct link to their monosaccharide composition and isomeric structure, including glycosidic linkages.
- The format of MilkOligoThesaurus allows it to be regularly updated with the latest discoveries on milk oligosaccharides. In this way, the current version will continue to evolve alongside with the availability of new data.

2. Background

Depending on the author's scientific background, the names given to milk oligosaccharides (MOs) in scientific publications are very heterogeneous and no consensus has yet been reached. Some MOs have no or several designations. Thus, collecting MO characteristics is a time-consuming process. To date, only two databases are dedicated to oligosaccharides found in mammalian milk, the National Institute of Standards and Technology (NIST) mass spectral library of free oligosaccharides in milk of mammals [4] as well as the recently published MilkOligoDB [5]. The aim of these databases is to explore the MO structural diversity across mammals by direct analysis of mammalian milk samples for the NIST database or through the harvest of MOs from the literature for MilkOligoDB. Both are exhaustive MO databases but none of them provide neither normalized names nor synonyms for most of the molecules registered. Some MO names and their chemical description can be found in larger chemical databases such as ChEBI, PubChem or GlyGen. Still none of these databases provides (i) all the naming variations (ii) and classify the molecule as a constituent of milk. On the basis of these observations, we built MilkOligoThesaurus that gathers the naming variations found in the literature.

3. Data Description

The names given to milk oligosaccharides (MOs) in scientific publications are very heterogeneous, probably due to their structural diversity and complexity, and the broad scientific area they are related to. MilkOligoThesaurus lists the most commonly identified MOs by analyzing selected reference scientific articles. Accordingly, MilkOligoThesaurus aims to:

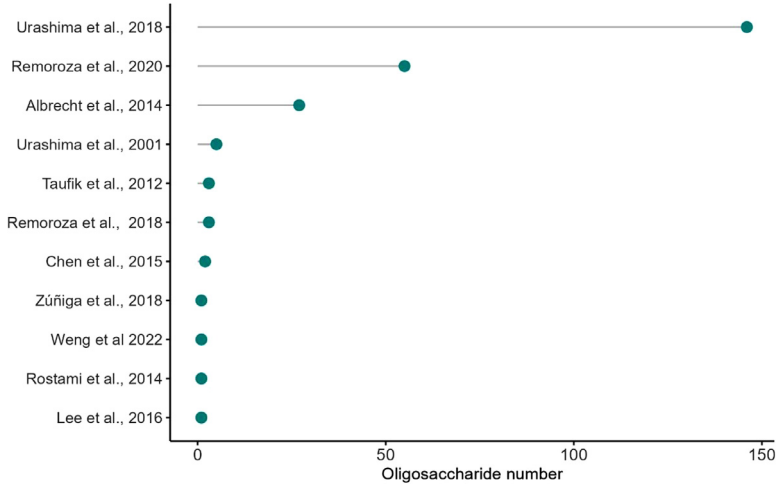

Fig. 1. Number of milk oligosaccharides extracted from the eleven sources used to build MilkOligoThesaurus.

- Associate each MO with its chemical properties (monoisotopic mass, formula, monosaccharide composition, and linkages)
- Propose a consensus name for MOs with heterogenous names found in the eleven selected articles and in the two chemical databases, or the MOs without names
- Index MO synonyms
- Build a FAIR (Findable, Accessible, Interoperable, Reusable) tool bound to evolve alongside with milk oligosaccharide knowledge [6].

MilkOligoThesaurus is accessible in the public data repository https://entrepot.recherche.data.gouv.fr/dataverse/inrae and registered with the following DOI: https://doi.org/10.57745/RA5DAC. Table 1 lists the sources used to extract the vocabulary (name and abbreviation) of the MOs. These sources were obtained from PubMed and Web of Science (WOS) databases. Four articles are reviews that summarize the structures of several milk oligosaccharides from different species including Human. The remaining articles are descriptive studies that analyze the MO composition of several individuals from different species. Each article provides an extensive list of milk oligosaccharides.

Fig. 1 presents the number of MO names extracted from the eleven sources used to build the thesaurus. Considering the redundancy of the information in the literature, a total of 11 articles was used to harvest the most common milk oligosaccharides. As shown in Fig. 1, the majority (95 %) of molecules have been extracted from three articles, illustrating the redundancy in the molecule names found in the set of papers. Fig. 2 presents the distribution of the number of names or abbreviations per oligosaccharide retrieved from the eleven sources, and from the databases ChEBI, PubMed and MilkOligoDB [5]. Of the 245 MOs, most had more than one unique name and more than one unique abbreviated name (127 MOs had three or more names or abbreviations Fig. 2). Nine had no name, one MO had only one denomination (B-tetrasaccharide with its IUPAC abbreviated form Gal(α1-3)Gal(β1-4)[Fuc(α1-2)]Glc).

The 245 oligosaccharides identified in mammalian milk retrieved from these sources were compiled in a 22-column tabular file. This tabular file is stored in the public data repository https://entrepot.recherche.data.gouv.fr/dataverse/inrae and registered with the following DOI: https://doi.org/10.57745/RA5DAC. Table 2 describes the content of each column of the tabular file.

In order to share this work and make it machine-actionable, the data were converted into the Simple Knowledge Organization System (SKOS) format and made available on the AgroPortal ontology repository https://agroportal.lirmm.fr/ontologies/MILKOLIGO/. We selected the features

Table 1
List of the eleven sources used to build MilkOligoThesaurus. These sources have been obtained from PubMed or Web of Science (WOS) databases. MO = milk oligosaccharide.

First author, year	DOI	Title *(targeted species)*	Type of study
Urashima et al., 2018	10.4052/tigg.1734.1SE	Human milk oligosaccharides as essential tools for basic and application studies on galectins *(Human)*	Review article
Remoroza et al., 2020	10.1021/acs.analchem.0c00342	Increasing the coverage of a mass spectral library of milk oligosaccharides using a hybrid-search-based bootstrapping method and milks from a wide variety of mammals *(cow, goat, asian buffalo, african lion)*	Research article
Albrecht et al., 2014	10.1017/S0007114513003772	A comparative study of free oligosaccharides in the milk of domestic animals *(cow, goat, sheep, pig, horse, dromedary camel)*	Research article
Urashima et al., 2001	10.1023/a:1014881913541	Oligosaccharides of milk and colostrum in non-human mammals *(brown capuchin, cow, buffalo, horse, goat, sheep, Ezo brown bear, Japanese black bear, Polar bear, white nosed coati, crabeater seal, hooded seal, elephant, rat, dog)*	Review article
Taufik et al., 2012	10.1007/s10719-012-9370-9	Structural characterization of neutral and acidic oligosaccharides in the milks of strepsirrhine primates: greater galago, aye-aye, Coquerel's sifaka and mongoose lemur *(greater galago, aye-aye, Coquerel's sifaka, mongoose lemur)*	Research article
Remoroza et al., 2018	10.1021/acs.analchem.8b01176	Creating a mass spectral reference library for oligosaccharides in human milk *(Human)*	Research article
Chen et al., 2015	10.1016/bs.accb.2015.08.002	Chapter Four - Human Milk Oligosaccharides (HMOS): Structure, Function, and Enzyme-Catalyzed Synthesis *(Human)*	Review article
Zúñiga, Monedero & Yebra, 2018	10.3389/fmicb.2018.01917	Utilization of host-derived glycans by intestinal *Lactobacillus* and *Bifidobacterium* species *(Human)*	Review article
Weng et al. 2022	10.1038/s41598-022-15140-7	Unusual free oligosaccharides in human bovine and caprine milk *(Human, cow, goat)*	Research article
Rostami et al., 2014	10.1371/journal.pone.0099824	Milk oligosaccharides over time of lactation from different dog breeds *(dog)*	Research article
Lee et al., 2016	10.1021/acs.jafc.6b02039	Rapid screening of bovine milk oligosaccharides in a whey permeate product and domestic animal milks by accurate mass database and tandem mass spectral library *(cow, buffalo, sheep)*	Research article

listed in Table 3. An intermediary tabular file was built to prepare the transformation into the SKOS standard. This intermediate tabular file and the following SKOS processed file (.rdf and .ttl) are available at https://entrepot.recherche.data.gouv.fr/dataverse/inrae and registered with the following DOI: https://doi.org/10.57745/RA5DAC .

Table 2
Description of the 22 columns of the MO thesaurus. (MO = milk oligosaccharide).

Column	Description
Entry_number	Unique identifier assigned to each MO
URI	Uniform Resource Identifier associates the scheme name http://opendata.inrae.fr/holooligo/ to the Entry_number
Normalized_MO_name	Consensus full name of the MO
Normalized_MO_name_source	Source from which the full MO name has been extracted
Abbreviated_normalized_MO_name	Consensus abbreviation of the MO name (shorthand MO annotation)
Abbreviated_normalized_MO_name_source	Source from which the abbreviation of the MO has been extracted
InChIKey	MO International chemical identifier, a textual identifier for chemical substances registered in the NCBI PubChem database. Note that not all molecules are registered in this database
CHEBI_ID	MO ID of the chemical ontology ChEBI. Note that not all molecules are registered in this database
Composition_Hex_HexNAc_Fuc_NeuAc_NeuGc Composition_Hex()HexNAc()dHex()NeuAc()NeuGc()	Two common ways to indicate the monosaccharide composition of MO with their respective number
Formula	The chemical formula of the molecule indicates the number of each type of atom of the MO
Monoisotopic_mass	MO molecular mass
Synonyms	There are several names for the same MO molecule. This column lists synonyms, and includes spelling differences.
Abbreviated_IUPAC_condensed_form IUPAC_condensed_form Abbreviated_IUPAC_extended_form IUPAC_extended_form	MO form according to nomenclature of the MO based on the Symbol Nomenclature for Glycans (SNFG) guidelines [7]. In abbreviated forms, symbols α and β have been replaced by a and b respectively
Isomer_group Isomer_group_abbreviation	MO isomers are oligosaccharides with the same monosaccharide composition but differences in a linkage or in the position of a fucose or sialyl acid. Generic terms are used to gather those isomeric forms with the same formula. Some oligosaccharide analytical techniques are unable to distinguish between isomers; thus authors indicate the presence of isomer groups.
Sourcing_MO_DOI	It contains the identifier of the sources from which the MO structure has been extracted
Author_year	First author and year of publication of the sources from which the MO structure has been extracted
PMID	PMID (PubMed Identifier) of the source from which the MO structure has been extracted

Hex: glucose or galactose, HexNAc N-acetylglucosamine or N- acetylgalactosamine, Fuc or dHex: fucose; NeuAc: N-acetylneuraminic acid; NeuGc: N-glycolylneuraminic acid.

Table 3
MO information collected in the SKOS properties.

SKOS properties	Tabulate file columns
URI	URI
skos:prefLabel@en	Normalized_MO_name
skos:altLabel@en(separator=";")	Abbreviated_normalized_MO_name Synonyms Abbreviated_IUPAC_condensed_form
skos:exactMatch	CHEBI_ID
skos:notation	InChIKey
dct:source	Normalized_MO_name_source Sourcing_MO_DOI Abbreviated_normalized_MO_name_source

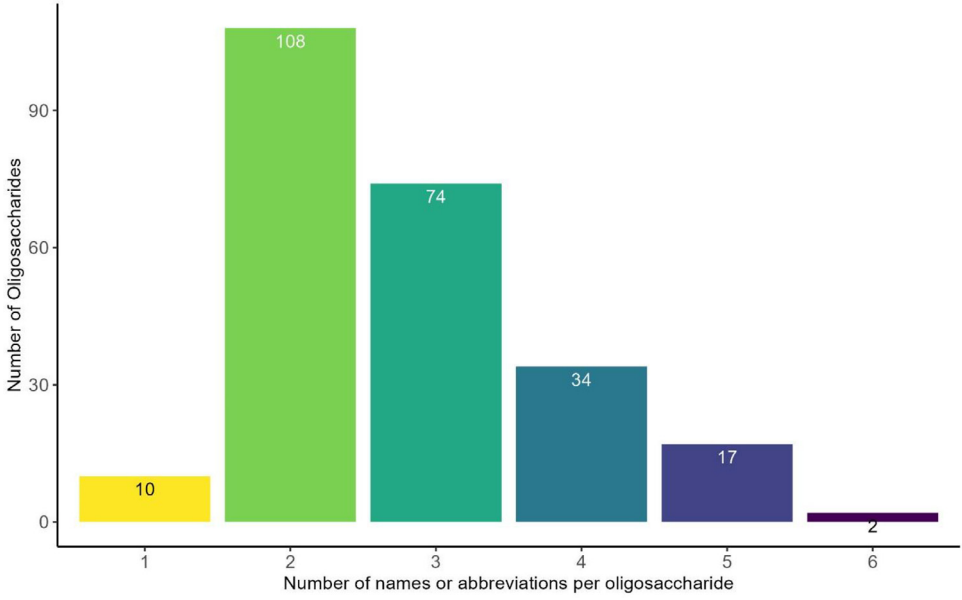

Fig. 2. Distribution of the number of names or abbreviations per oligosaccharide in MilkOligoThesaurus.

4. Experimental Design, Materials and Methods

The construction of MilkOligoThesaurus involved the following steps:

First, a query was formulated to retrieve articles about milk oligosaccharides from the PubMed and WOS databases. Among the results, articles were selected based on the following inclusion criteria. They had to:

- Deal with milk oligosaccharides (MOs) in mammalian species including Human.
- Provide an extensive list of MOs, with structural details provided to clearly identify the sequence and linkages of the constituting monosaccharides.
- Detail their isomeric structures using Symbol Nomenclature for Graphical Representations of Glycans [7] or the Oxford symbol nomenclature [8].

Second, the molecules described in the sources had to meet the following definition adapted from Chen et al. [9] for human MOs:

- MOs are built from five monosaccharides including d-glucose (Glc), d-galactose (Gal), N-acetyl-d-glucosamine (GlcNAc), l-fucose (Fuc), and a sialic acid either N-acetylneuraminic acid or N-glycolylneuraminic (NeuAc or NeuGc, respectively).
- MOs are extended from lactose (Galβ1–4Glc) with Glc at the reducing end with an existing mix of α and β anomers. While Gal and GlcNAc are always presented with β-d-glycosidic linkages, Fuc and NeuAc are always presented with α-l- and α- glycosidic linkages, respectively.

Considering the importance of their occurrence in MO scientific literature, we added some MOs with « deviant » structures according to Chen et al. [9]:

- Milk oligosaccharides containing a terminal N-acetylgalactosamine (GalNAc) such as Fucosyl-galactosaminyllactose.
- Structures in which the glucose or lactose at the reducing end have been replaced by N-acetyllactosamine such as 3-Fucosyllactosamine and Disialyllactosamine.

- Milk oligosaccharides including a succession of galactose monosaccharides (α- or β-linked) such as Isoglobotriose and 6′-Galactosyllactose

MilkOligoThesaurus also contains 2 oligosaccharides that are bound to lipids as they were found in their free form in several non-human milks (Asialo-GM1 and Asialo-GM2) [10].

Third, oligosaccharides were manually searched in two databases: NCBI PubChem (https://pubchem.ncbi.nlm.nih.gov/) and ChEBI (https://www.ebi.ac.uk/chebi/). If the oligosaccharide was found, relevant information was collected and included: monoisotopic mass, chemical formula, database ID and, if available, synonyms. Whenever the oligosaccharide was not registered in a database, these properties were obtained directly from the scientific articles. As a last resort, some properties could be inferred from the structure of the milk oligosaccharides and comparisons with similar milk oligosaccharides for which properties were available.

Fourth, the data obtained were gathered in the thesaurus. An important number of molecules were found in several of the selected articles, but data were not always consistent regarding the following properties: name, and abbreviated name. Thus, a comparison was made between the sources supported by the chemical database information when available. Consequently, three sources are available for each molecule, analogous or not: the source where the oligosaccharide structure was found (Sourcing_MO_DOI), the source we have chosen for the oligosaccharide name (Normalized_MO_name_source) as well as the source we have chosen for the oligosaccharide name abbreviation (Abbreviated_normalized_MO_name_source).

We encountered two situations: (i) MO had several names. We prioritized the name and abbreviation that allowed relative homogeneity within the same structural group (the other names being included in the “Synonyms” column). In some cases, although the molecule had a name, to make the thesaurus easier to read, we have suggested names considering the similarity of the structure in the same structural group. The names from the scientific source were thus added in the “Synonyms” column. (ii) A few molecules had no names. For some of them, we proposed names based on their structure and/or the names of the closest oligosaccharides. All proposed names are identified in the name and abbreviation source columns by the mention “Author_proposal” to ensure traceability.

The prime symbol (’) in the names of the oligosaccharides indicates that the monosaccharide residue is not linked on the reducing end monosaccharide but on the immediately following residue. The double prime symbol (”) indicates that the monosaccharide residue is linked to the residue immediately following the first prime residue (for example 2′-Fucosyllactose stands for Fuc(a1-2)Gal(b1-4)Glc; 3-Fucosyllactose stands for Gal(b1-4)[Fuc(a1-3)]Glc and 3″-Neu5Ac-6′-Neu5Ac-galactotriose stands for NeuAc(a2-3)Gal(b1-3)[NeuAc(a2-6)]Gal(b1-4)Glc) [11].

Finally, the thesaurus has been converted into the SKOS format using the SKOSplay! converter function (https://skos-play.sparna.fr/play/convert) following the recommendations of [12]. And the quality was checked using https://skos-play.sparna.fr/skos-testing-tool/. Results are made available on AgroPortal [13] (https://agroportal.lirmm.fr/ontologies/MILKOLIGO/), an ontology repository for the agronomy domain. The SKOS format is used to standardize the data and allocate an URI (Unique Ressource Identifier) to each concept (milk oligosaccharide) of the thesaurus. This makes MilkOligoThesaurus both machine-readable and humans-friendly, fundamental elements of FAIR principles [6].

Limitations

None.

Ethics Statement

The MilkOligoThesaurus construction process did not involve experiments on humans or animals, and no social media data was collected.

Data Availability

MilkOligoThesaurus: A milk oligosaccharide thesaurus (HoloOLIGO project) (Original data) (https://entrepot.recherche.data.gouv.fr/).

CRediT Author Statement

Mathilde Rumeau: Conceptualization, Methodology, Writing – original draft; **François Fenaille:** Data curation, Validation, Writing – review & editing; **Agnès Girard:** Methodology, Resources, Writing – review & editing; **Valentin Loux:** Conceptualization, Writing – review & editing; **Mouhamadou Ba:** Conceptualization, Writing – review & editing; **Claire Nédellec:** Conceptualization, Writing – review & editing; **Louise Deléger:** Conceptualization, Writing – review & editing; **Robert Bossy:** Conceptualization, Writing – review & editing; **Sophie Aubin:** Software, Resources, Writing – review & editing; **Christelle Knudsen:** Conceptualization, Writing – review & editing; **Sylvie Combes:** Conceptualization, Data curation, Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

This work was supported by the ANR (HoloOLIGO project, ANR-21-CE20-0045-01), the INRAE metaprogram (Holoflux). This work used resources developed as part of the ANR FooSIN project (ANR-19-DATA-0019-01). Funders had no role in the review process.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Borovikova, A. Ferré, R. Bossy, M. Roche, C. Nédellec, Could keyword masking strategy improve language model?, in: E. Métais, F. Meziane, V. Sugumaran, W. Manning, S. Reiff-Marganiec (Eds.) Natural Language Processing and Information Systems, Springer Nature Switzerland, Cham, 2023, pp. 271–284, doi:10.1007/978-3-031-35320-8_19.
[2] X. Wang, V. Hu, X. Song, S. Garg, J. Xiao, J. Han, ChemNER: fine-grained chemistry named entity recognition with ontology-guided distant supervision, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Dominican Republic, Association for Computational Linguistics, Online and Punta Cana, 2021, pp. 5227–5240, doi:10.18653/v1/2021.emnlp-main.424.
[3] A. Ferré, L. Deléger, R. Bossy, P. Zweigenbaum, C. Nédellec, C-Norm: a neural approach to few-shot entity normalization, BMC Bioinform. 21 (2020) 579, doi:10.1186/s12859-020-03886-8.
[4] C.A. Remoroza, T.D. Mak, M.L.A. De Leoz, Y.A. Mirokhin, S.E. Stein, Creating a mass spectral reference library for oligosaccharides in human milk, Anal. Chem. 90 (2018) 8977–8988, doi:10.1021/acs.analchem.8b01176.
[5] S.D. Durham, Z. Wei, D.G. Lemay, M.C. Lange, D. Barile, Creation of a milk oligosaccharide database, MilkOligoDB, reveals common structural motifs and extensive diversity across mammals, Sci. Rep. 13 (2023) 10345, doi:10.1038/s41598-023-36866-y.
[6] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding principles for scientific data management and stewardship, Sci. Data 3 (2016) 160018, doi:10.1038/sdata.2016.18.
[7] A. Varki, R.D. Cummings, M. Aebi, N.H. Packer, P.H. Seeberger, J.D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J.J. Prestegard, R.L. Schnaar, H.H. Freeze, J.D. Marth, C.R. Bertozzi, M.E. Etzler, M. Frank, J.F. Vliegenthart, T. Lütteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K.F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, S. Kornfeld, Symbol nomenclature for graphical representations of glycans, Glycobiology 25 (2015) 1323–1324, doi:10.1093/glycob/cwv091.

[8] D.J. Harvey, A.H. Merry, L. Royle, M.P. Campbell, R.A. Dwek, P.M. Rudd, Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds, Proteomics 9 (2009) 3796–3801, doi:10.1002/pmic.200900096.
[9] X. Chen, Chapter four - human milk oligosaccharides (HMOS): structure, function, and enzyme-catalyzed synthesis, in: Advances in Carbohydrate Chemistry and Biochemistry, Elsevier, 2015, pp. 113–190, doi:10.1016/bs.accb.2015.08.002.
[10] C.A. Remoroza, Y. Liang, T.D. Mak, Y. Mirokhin, S.L. Sheetlin, X. Yang, J.V. San Andres, M.L. Power, S.E. Stein, Increasing the coverage of a mass spectral library of milk oligosaccharides using a hybrid-search-based bootstrapping method and milks from a wide variety of mammals, Anal. Chem. 92 (2020) 10316–10326, doi:10.1021/acs.analchem.0c00342.
[11] A.D. McNaught, Nomenclature of carbohydrates (IUPAC Recommendations 1996), Pure Appl. Chem. 68 (1996) 1919–2008, doi:10.1351/pac199668101919.
[12] S. Aubin, J. Yon, 3 outils pour transformer du tabulé en SKOS, (2022). https://hal.science/hal-03852086.
[13] C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. Dzalé Yeumo, V. Emonet, J. Graybeal, M.-A. Laporte, M.A. Musen, V. Pesce, P. Larmande, AgroPortal: a vocabulary and ontology repository for agronomy, Comput. Electron. Agric. 144 (2018) 126–143, doi:10.1016/j.compag.2017.10.012.