BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Monitoring carbohydrate 3D structure quality with the *Privateer* database

Jordan S. Dialpuri, Haroldas Bagdonas, Lucy C. Schofield, Phuong Thao Pham, Lou Holland and Jon Agirre*

Full Research Paper

Open Access

Address:
York Structural Biology Laboratory, Department of Chemistry, University of York, UK

Email:
Jon Agirre* - jon.agirre@york.ac.uk

* Corresponding author

Keywords:
carbohydrates; database; *N*-glycans; *N*-glycosylation; polysaccharides; validation; website

Beilstein J. Org. Chem. **2024,** *20,* 931–939.
https://doi.org/10.3762/bjoc.20.83

Received: 30 January 2024
Accepted: 10 April 2024
Published: 24 April 2024

This article is part of the thematic issue "Chemical glycobiology".

Guest Editor: E. Fadda

© 2024 Dialpuri et al.; licensee Beilstein-Institut.
License and terms: see end of document.

Abstract

The remediation of the carbohydrate data of the Protein Data Bank (PDB) has brought numerous enhancements to the findability and interpretability of deposited glycan structures, yet crucial quality indicators are either missing or hard to find on the PDB pages. Without a way to access wider glycochemical context, problematic structures may be taken as fact by keen but inexperienced scientists. The *Privateer* software is a validation and analysis tool that provides access to a number of metrics and links to external experimental resources, allowing users to evaluate structures using carbohydrate-specific methods. Here, we present the *Privateer* database, a free resource that aims to complement the growing glycan content of the PDB.

Introduction

Carbohydrate modelling is an important but often cumbersome stage in the macromolecular X-ray structure solution workflow. The accurate modelling of glycoproteins and protein–carbohydrate complexes is pivotal in understanding the complex biochemical interactions that affect the physiological function of cells [1]. Any mechanistic analysis done with finely grained approaches such as QM/MM [2] relies heavily on the correctness of the starting coordinates. Despite this, carbohydrate models often contain modelling inconsistencies that cannot easily be attributed to known biochemical principles [3]. These inconsistencies cannot solely be attributed to model-building inexperience, as carbohydrate model building is an inherently difficult task, which in the past has been plagued with software related problems from incorrect libraries to incomplete support [4]. Carbohydrates are mobile, highly branched additions to the comparatively rigid protein framework; in macromolecular crystallography, this causes heterogeneity throughout the crystal lattice and, therefore, poorly resolved density regions, whereas

in electron cryo-microscopy different conformations and compositions are averaged out during image classification and volume reconstruction [5].

Owing to these difficulties, it is not uncommon to find problematic carbohydrate structures in the Protein Data Bank (PDB), from the initial works of Lütteke, Frank and von der Lieth [6,7], who identified numerous issues affecting nomenclature and linkages (estimated to affect 30% of the structures at the time), to the reports of surprising – or indeed glyco-chemically impossible – linkages in a glycoprotein as pointed out by Crispin and collaborators [8], and more recently the realisation that high-energy ring conformations, a rare event in six-membered pyranosides, were present in ca. 15% of the *N*-glycan components of glycoproteins in the PDB [3]. Many of these findings originated the development of new resources, including services and databases [9-13], and standalone software [14-18]. Among these, the *Privateer* software package has been a key tool for glycoprotein and protein–carbohydrate complex validation: *Privateer* analyses the conformational plausibility of each sugar model [3], checks that structures match the nomenclature used for deposition in the PDB [14], compares glycan compositions to known structures as reported by glycomics (e.g., GlyConnect [19]) and glyco-informatics (e.g., GlyTouCan [20]) databases and repositories [15], and checks how close the overall conformation of *N*-glycans comes to that of validated deposited structures [16].

The PDB-REDO [21] database is a separate resource, albeit linked to the PDB in that the entries that compound PDB-REDO are those original PDB crystallographic entries that included experimental data (i.e., reflection intensities or amplitudes); each entry includes a re-refined, sometimes even re-built to some extent, copy of the original model. These newer versions are produced with state-of-the-art methods, many of which were probably not available at the time of deposition; hence, the quality of the models is expected to improve. Because the methodology included in PDB-REDO had been affected by the lack of automatic support that plagued general purpose crystallographic model building and refinement software [4], carbohydrate-specific methods have been gradually introduced over the years [22,23].

Whilst *Privateer* has been a staple tool in carbohydrate validation, the results of *Privateer* have not been collated in such a way that allows for easy judgement of carbohydrate model quality in the PDB [24]. Providing users with metrics that allow them to make chemically sound conclusions about the model is an important facility, especially for novice users. To allow this to happen readily on PDB distribution sites, we present the *Privateer* database, a freely available, up-to-date collection of validation information for both the PDB and PDB-REDO [21] archives.

Results and Discussion

Format of the validation report

The JSON file deposited for each PDB entry follows a consistent format, as shown in Figure 1. At the top level, the file contains metadata about the validation report. This metadata provides the date that the validation report was generated as well as the availability of experimental data. It is helpful to have this information easily accessible as *Privateer* cannot calculate the real space correlation coefficient without experimental data; therefore, programmatic access to further validation metrics could be streamlined, knowing the information is not present.

Also at the top level of the validation report is the beginning of the carbohydrate information, listed as ‘glycans’ in the JSON format. Within this ‘glycan’ scope, information is segmented into glycan types, that is, ‘n-glycan’, ‘o-glycan’, ‘s-glycan’, ‘c-glycan’, and 'ligand'. Each of these glycan types contains an array of individual glycans of that type, and the format of the data inside each of these glycan types is identical.

The data contained in each glycan entry is shown in Table 1. Each entry contains information about the protein chain attachment, the number of sugars in the glycan, the WURCS2.0 code [25], the standard nomenclature for glycan SVG, and an array of sugar entries. The validation data calculated by *Privateer* for each sugar entry is shown in Table 2, and that for each linkage is shown in Table 3.

Visualising a validation report

While the database is available on GitHub for programmatic access, viewing a validation report entry in plaintext can be difficult, time-consuming and would certainly be a poor experience for the end user. To improve the utility of this database, we have provided a visualisation of the information contained within the validation report for both PDB and PDB-REDO databases, which is available alongside the *Privateer Web App* [26], https://privateer.york.ac.uk/database.

The first section of this visual report displays a global outlook on the validity of the model through two graphs. The first graph shows the conformational landscape for the pyranose sugars. For a sugar model to be deemed valid, the ring must be in the $^4C_1$ chair conformation. This can be measured through the Cremer–Pople parameters θ and ψ [27]. Theta angles of $0° < \theta < 360°$ indicate that the sugar may be in a higher-energy confirmation; therefore, caution should be placed on any conclusions drawn from the molecular model of the sugar. Also

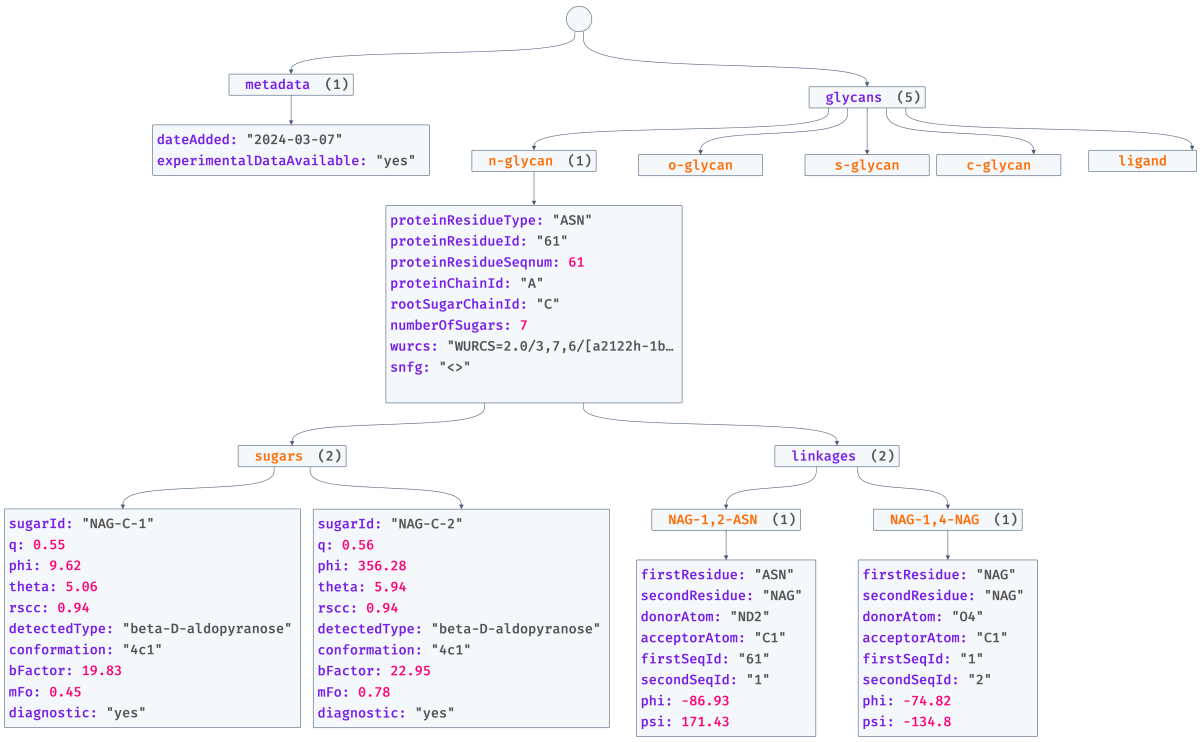

Figure 1: Format of a validation report in JSON format. At the top level of the tree, the report contains metadata about itself, such as the date the entry was added to the database and if experimental data is available. Also at the top level of the tree is the glycan information, separated into glycan types. Each glycan also contains a list of sugars, with a range of validation information and a list of linkage with torsion angle information. Tree visualisation was created with jsoncrack.com.

Table 1: Data contained within each glycan entry.

Key	Example	Type
proteinResidueType	ASN	string
proteinResidueId	61	string
proteinResidueSeqnum	61	number
proteinChainId	A	string
rootSugarChainId	C	string
numberOfSugars	7	number
wurcs	WURCS=2.0/3,7,6/…	string
snfg	<svg> … </svg>	string
sugars	see Table 2	array

Table 2: Data contained within each sugar entry.

Key	Example	Type
sugarID	NAG-D-1	string
q	0.54	number
phi	303.44	number
theta	6.45	number

Table 2: Data contained within each sugar entry. (continued)

rscc	0.922	number
detectedType	beta-D-aldopyranose	string
conformation	4c1	string
bFactor	22.367	number
mFo	0.421	number
diagnostic	yes	string

Table 3: Data contained within each linkage entry.

Key	Example	Type
firstResidue	NAG	string
secondResidue	NAG	string
donorAtom	O4	string
acceptorAtom	C1	string
firstSeqId	1	string
secondSeqId	2	string
phi	−54.91	number
psi	−108.47	number

in the first section of the visual validation report is a plot of the B-factor (temperature factor) versus the real space correlation coefficient (RSCC) (Figure 2). A well-refined, well-built model would be expected to have a B-factor that increases somewhat linearly as the RSCC decreases. Over-refined models may deviate from this trend and would be trivial to identify.

The validation report also displays a table (Figure 3) representing two-dimensional descriptions of each glycan in the

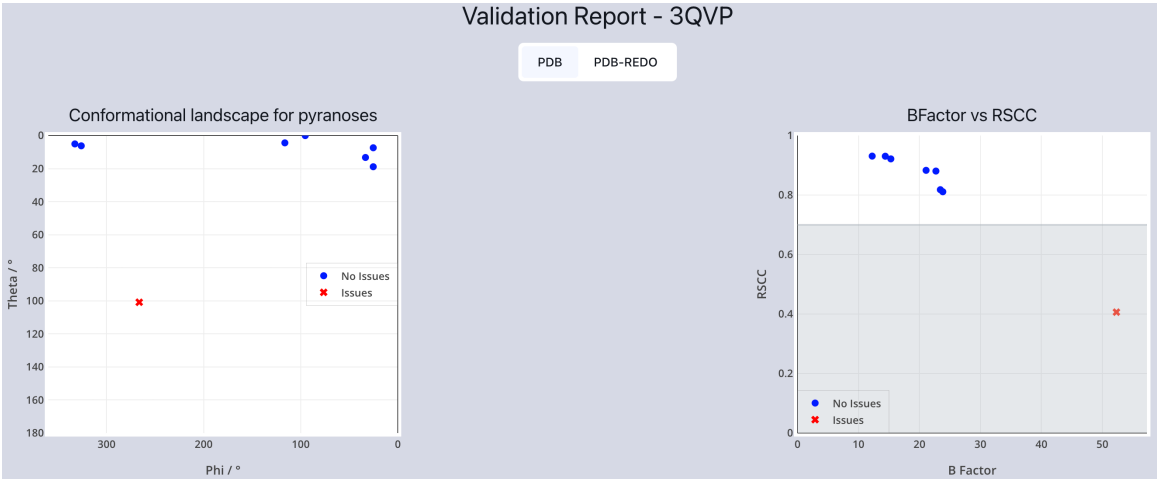

Figure 2: Left: Graphical representation of the conformational landscape of pyranose sugars. A well-modelled D-sugar would be expected to be in the lowest-energy conformation and have a theta angle close to 0° and would be indicated by a blue point; deviations from the ideal conformation are highlighted with a red cross. Right: Real space correlation coefficient plotted against the B-factor, which enables the refinement of the sugars to be assessed. A slight negative correlation would be expected for a well-refined model. Results taken from the *Privateer* database report for 3QVP [28].

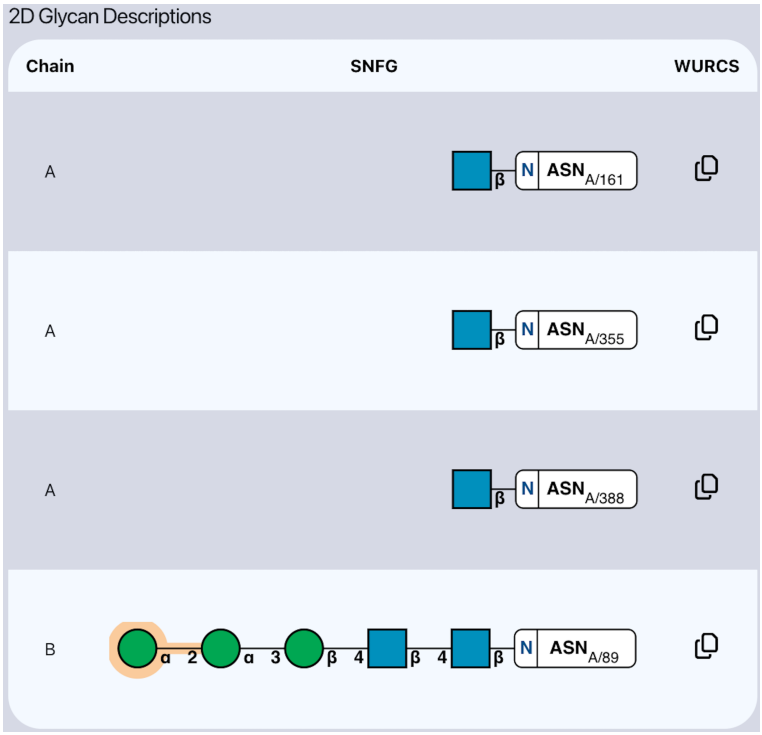

Figure 3: Table of two-dimensional Symbol Nomenclature for Glycan (SNFG) visualisations, which can allow for easy oversight of the validity of a particular glycan. Sugars that have issues identified by *Privateer* are highlighted in orange, and linkages that have unusual torsion angles are also highlighted in orange. The WURCS codes for each glycan are also available to copy to the clipboard. Table taken from the *Privateer* database report for 3QVP.

model. Each row in the table represents a unique glycan and includes the chain identifier, standard Symbol Nomenclature for Glycans (SNFG [29]) visualisation, and copyable WURCS [25] identifier. The SNFG displayed for each glycan paints a picture of how well built the glycan model is, as the metrics and validity conclusions calculated by *Privateer* are embedded within each shape and linkage of the diagram. For example, a shape with an orange highlight indicates something is abnormal about the ring's conformation, puckering, or monosaccharide nomenclature [30]. Similarly, a linkage with an orange highlight indicates that the torsion angles between the linkages are unexpected and require further inspection [16].

In addition to the SNFG, also displayed for each table entry is a copyable WURCS link, which encodes the complete glycan format in a linear code. The decision to present this information as a copyable link, as opposed to as plaintext is due to the inherent difficulty and unlikeliness for a human to read and understand the WURCS code. It is much more likely that the WURCS code would be copied and searched for in a glycomics database, hence we provide that functionality in a streamlined way.

The final section of the validation report includes all of the validation metrics calculated by *Privateer* and, most importantly, the diagnostic provided by *Privateer* (Figure 4). A 'yes' diagnostic indicates the conformation is correct for the glycosylation type (e.g., $^4C_1$ for GlcNAc in an *N*-glycan, $^1C_4$ for mannose in a *C*-glycan), has the correct anomer, and has an acceptable fit to density. This diagnostic indicates that the sugar is valid, whereas a diagnostic of 'check' indicates that *Privateer* has detected a potential inconsistency affecting ring conformation, which requires manual inspection. Finally, a 'no' diagnostic indicates that the sugar needs a more detailed manual inspection to correct any conformational issues, anomeric issues, or fitting issues.

Searching for entries in the *Privateer* database

Another interesting application of the collection of data available in the *Privateer* database is to visualise aggregated carbohydrate data from the PDB. Using the search interface on the *Privateer* database homepage, carbohydrate-containing PDB entries can easily be found and filtered. *Privateer* database entries for specific glycosylation types, namely, *N*-glycosylation, *O*-glycosylation, *S*-glycosylation, or *C*-glycosylation can be filtered quickly and easily. Additional filtering by linkage type is also possible, allowing niche glycosylation targets to be obtained. For example, filtering for *C*-glycans with a 'BMA-1,1-TRP' (the correct pair would be 'MAN-1,1-TRP', as the linkage in the modification is an alpha linkage) returns nine instances of incorrect sugar conformations in *C*-mannosylation found within the *Privateer* database in a table containing the frequency of the target linkage as well as a link to the *Privateer* database report page for target entry (Figure 5). This table view is also keyword or range-filterable at every data column, which allows for trivial searches of potentially interesting models.

Detailed monosaccharide validation data

Sugar ID	Conformation	Q	Phi	Theta	RSCC	B Factor	Detected Type	Type	Diagnostic
NAG-A-603	$^4c_1$	0.57	25.25	7.38	0.93	12.25	beta-D-aldopyranose	n-glycan	yes
NAG-A-604	$^4c_1$	0.56	326.04	6.21	0.88	21.09	beta-D-aldopyranose	n-glycan	yes
NAG-A-605	$^4c_1$	0.56	332.64	5.06	0.88	22.70	beta-D-aldopyranose	n-glycan	yes
NAG-B-1	$^4c_1$	0.55	116.33	4.41	0.93	14.40	beta-D-aldopyranose	n-glycan	yes
NAG-B-2	$^4c_1$	0.55	33.27	13.26	0.92	15.31	beta-D-aldopyranose	n-glycan	yes
BMA-B-3	$^4c_1$	0.52	25.27	18.87	0.81	23.83	beta-D-aldopyranose	n-glycan	yes
MAN-B-4	$^4c_1$	0.59	95.31	0.03	0.82	23.43	alpha-D-aldopyranose	n-glycan	yes
MAN-B-5	$^1s_5$	0.80	266.38	100.86	0.41	52.31	alpha-D-aldopyranose	n-glycan	check

Figure 4: Table of validation data for each sugar residue within PDB code 3QVP available in the visual validation report. The table contains all validation metrics calculated by *Privateer* including the Cremer–Pople puckering parameters, correlation coefficient, and, importantly, *Privateer* diagnostic, which can be used to identify the validity of each sugar. Table taken from the Privateer database report for 3QVP.

Figure 5: Table of available *Privateer* reports for the BMA-1,1-TRP linkage in *C*-glycans (*C*-mannosylation) sorted by the frequency (count) of the linkage in the deposited model. The table contains information of the carbohydrate type, PDB code, linkage, frequency, and resolution, as well as a link to the *Privateer* database report for each PDB entry.

Trends in the *Privateer* database

Using the *Privateer* database, global statistics throughout the PDB and PDB-REDO can be calculated with ease. Observing deposition trends in the PDB is often interesting as it can provide insight into the kinds of structures that are experimentally obtainable over time. With the *Privateer* database, trends in glycosylation deposition in the PDB over time can be measured, as shown in Figure 6. Importantly, as the *Privateer* database is completely recompiled every week, these trends remain consistent with the PDB. To allow for easy and up-to-date observation for anyone, compiled statistics are freely available alongside the *Privateer Web App*, https://privateer.york.ac.uk/statistics.

While simply looking at glycosylation over time using the *Privateer* database is possible, the validation reports calculated by *Privateer* contain a whole host of other interesting pieces of information. In an analogous way to looking at glycosylation over time, the type and validity of carbohydrates in the PDB can also be observed over time. The statistics page available alongside the *Privateer Web App* contains up-to-date plots of validation and conformational errors over time and resolution.

Conclusion

In conclusion, the new *Privateer* database encompasses the carbohydrate validation capabilities of *Privateer* in an easily accessible pre-prepared form. The database contains all validation metrics calculated by *Privateer* as well as highlighted SNFG diagrams in SVG format for easy third-party web use. Statistics are automatically computed weekly and are available alongside the database both on GitHub and the interactive web page.

Materials and Methods

The *Privateer* software package [14] was used to compute metrics and statistics for each entry in the PDB [24] or in PDB-REDO [21]. For each structure in the PDB, the carbohydrate-containing chains are first identified before being validated using the suite of validation tools available within *Privateer*. Using the Python bindings available within the latest versions of *Privateer*, a validation report can be generated for each carbohydrate in the molecular model. This report is put out in JSON format for easy consumption by web-based database frontends. The initial report generation was completed in parallel over 64 CPU cores in around 5 h. After the initial surveys through PDB and PDB-REDO, this process only needs

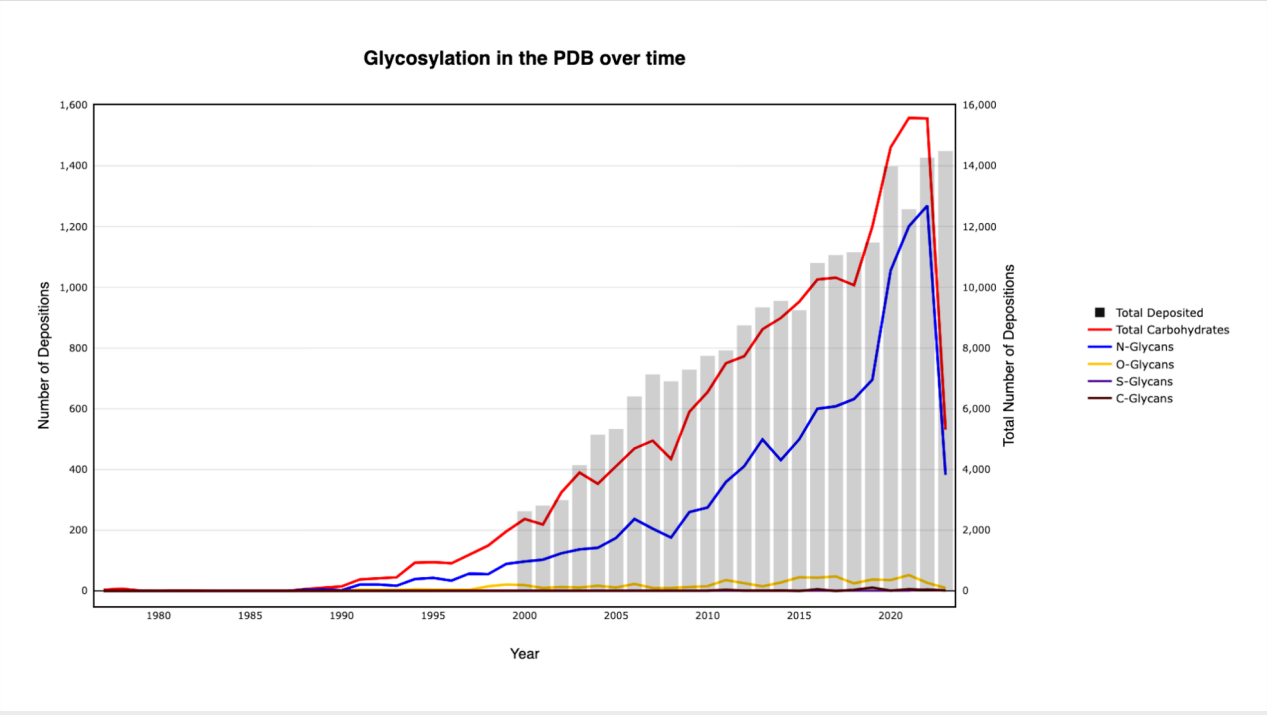

Figure 6: Plot showing trends in deposition in the PDB over time from 1975 to the present. Grey bars show the total deposited models into the PDB for all structural determination methods. Lines show glycosylation in the PDB over time, split into *N*-glycans, *O*-glycans, *S*-glycans, and *C*-glycans.

to be completed when new molecular models are deposited into the PDB, which occurs weekly. Although compiling validation reports for only new structures would be more efficient, this would fail to encompass changes in structures in historical entries, therefore the *Privateer* database is recompiled weekly.

The database, which receives any updates to the reports after recompilation is hosted on GitHub. The database is separated into PDB and PDB-REDO sections, which are in turn structured in the same format as the PDB archive, separated into folders by the middle two characters of the PDB four-letter code. For convenience, the presentation of the database is hosted alongside the *Privateer Web App* [26]; the database part can be accessed at https://privateer.york.ac.uk/database or by navigating to the database icon on the top right of the screen. The website is dynamic and compatible with desktop and laptop computers, plus tablets and smartphones.

Acknowledgements

We are grateful to the University of York IT Services and Darren Miller in particular for accommodating our needs and offering timely and excellent technical support. Lastly, we should like to acknowledge and highlight the contributions of Thomas Lütteke, Martin Frank, and the late Willy von der Lieth, pioneers of carbohydrate structure validation, whose research informed some of the methods showcased in the *Privateer* database.

Funding

Jordan Dialpuri is funded by the Biotechnology and Biological Sciences Research Council (BBSRC; grant No. BB/T0072221). Haroldas Bagdonas is funded by The Royal Society (grant No. RGF/R1/181006). Lucy Schofield is funded by STFC/CCP4 PhD studentship agreement 4462290 (York) / S2 2024 012 (STFC) awarded to Jon Agirre. Phuong Thao Pham is a self-funded PhD student. Lou Holland is funded by The Royal Society (URF\R\221006). Jon Agirre is a Royal Society University Research Fellow (awards UF160039 and URF\R\221006).

Author Contributions

Jordan S. Dialpuri: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; software; validation; visualization; writing – original draft; writing – review & editing. Haroldas Bagdonas: software. Lucy C. Schofield: conceptualization; software; visualization. Phuong Thao Pham: data curation. Lou Holland: software; validation; visualization. Jon Agirre: conceptualization; data curation; funding acquisition; investigation; project administration; software; supervision; validation; writing – original draft; writing – review & editing.

ORCID® iDs

Jordan S. Dialpuri - https://orcid.org/0000-0002-6205-2661
Haroldas Bagdonas - https://orcid.org/0000-0001-5028-4847
Lucy C. Schofield - https://orcid.org/0009-0001-2069-878X

Phuong Thao Pham - https://orcid.org/0000-0002-6205-1298
Lou Holland - https://orcid.org/0000-0002-3867-1833
Jon Agirre - https://orcid.org/0000-0002-1086-0253

Data Availability Statement

All source code is publicly available on GitHub (https://github.com/glyco-jones/privateer and https://github.com/Dialpuri/PrivateerDatabase). The *Privateer* database is available at https://privateer.york.ac.uk/database and calculated statistics are available at https://privateer.york.ac.uk/statistics. Both pages will remain automatically updated with respect to the source code on GitHub.

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: https://doi.org/10.3762/bxiv.2024.6.v1

References

1. Brockhausen, I.; Schutzbach, J.; Kuhns, W. *Acta Anat.* **1998,** *161,* 36–78. doi:10.1159/000046450
2. Calvelo, M.; Males, A.; Alteen, M. G.; Willems, L. I.; Vocadlo, D. J.; Davies, G. J.; Rovira, C. *ACS Catal.* **2023,** *13,* 13672–13678. doi:10.1021/acscatal.3c02378
3. Agirre, J.; Davies, G.; Wilson, K.; Cowtan, K. *Nat. Chem. Biol.* **2015,** *11,* 303. doi:10.1038/nchembio.1798
4. Agirre, J. *Acta Crystallogr., Sect. D: Struct. Biol.* **2017,** *73,* 171–186. doi:10.1107/s2059798316016910
5. Atanasova, M.; Bagdonas, H.; Agirre, J. *Curr. Opin. Struct. Biol.* **2020,** *62,* 70–78. doi:10.1016/j.sbi.2019.12.003
6. Lütteke, T.; Frank, M.; von der Lieth, C.-W. *Nucleic Acids Res.* **2005,** *33,* D242–D246. doi:10.1093/nar/gki013
7. Lütteke, T.; Frank, M.; von der Lieth, C.-W. *Carbohydr. Res.* **2004,** *339,* 1015–1020. doi:10.1016/j.carres.2003.09.038
8. Crispin, M.; Stuart, D. I.; Jones, E. Y. *Nat. Struct. Mol. Biol.* **2007,** *14,* 354. doi:10.1038/nsmb0507-354a
9. Frank, M.; Lütteke, T.; von der Lieth, C.-W. *Nucleic Acids Res.* **2007,** *35,* 287–290. doi:10.1093/nar/gkl907
10. von der Lieth, C.-W.; Freire, A. A.; Blank, D.; Campbell, M. P.; Ceroni, A.; Damerell, D. R.; Dell, A.; Dwek, R. A.; Ernst, B.; Fogh, R.; Frank, M.; Geyer, H.; Geyer, R.; Harrison, M. J.; Henrick, K.; Herget, S.; Hull, W. E.; Ionides, J.; Joshi, H. J.; Kamerling, J. P.; Leeflang, B. R.; Lütteke, T.; Lundborg, M.; Maass, K.; Merry, A.; Ranzinger, R.; Rosen, J.; Royle, L.; Rudd, P. M.; Schloissnig, S.; Stenutz, R.; Vranken, W. F.; Widmalm, G.; Haslam, S. M. *Glycobiology* **2011,** *21,* 493–502. doi:10.1093/glycob/cwq188
11. Lütteke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C.-W. *Glycobiology* **2006,** *16,* 71R–81R. doi:10.1093/glycob/cwj049
12. Toukach, P. V.; Egorova, K. S. *Nucleic Acids Res.* **2016,** *44,* D1229–D1236. doi:10.1093/nar/gkv840
13. Böhm, M.; Bohne-Lang, A.; Frank, M.; Loss, A.; Rojas-Macias, M. A.; Lütteke, T. *Nucleic Acids Res.* **2019,** *47,* D1195–D1201. doi:10.1093/nar/gky994
14. Agirre, J.; Iglesias-Fernández, J.; Rovira, C.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D. *Nat. Struct. Mol. Biol.* **2015,** *22,* 833–834. doi:10.1038/nsmb.3115
15. Bagdonas, H.; Ungar, D.; Agirre, J. *Beilstein J. Org. Chem.* **2020,** *16,* 2523–2533. doi:10.3762/bjoc.16.204
16. Dialpuri, J. S.; Bagdonas, H.; Atanasova, M.; Schofield, L. C.; Hekkelman, M. L.; Joosten, R. P.; Agirre, J. *Acta Crystallogr., Sect. D: Struct. Biol.* **2023,** *79,* 462–472. doi:10.1107/s2059798323003510
17. Emsley, P.; Crispin, M. *Acta Crystallogr., Sect. D: Struct. Biol.* **2018,** *74,* 256–263. doi:10.1107/s2059798318005119
18. Atanasova, M.; Nicholls, R. A.; Joosten, R. P.; Agirre, J. *Acta Crystallogr., Sect. D: Struct. Biol.* **2022,** *78,* 455–465. doi:10.1107/s2059798322001103
19. Alocci, D.; Mariethoz, J.; Gastaldello, A.; Gasteiger, E.; Karlsson, N. G.; Kolarich, D.; Packer, N. H.; Lisacek, F. *J. Proteome Res.* **2019,** *18,* 664–677. doi:10.1021/acs.jproteome.8b00766
20. Fujita, A.; Aoki, N. P.; Shinmachi, D.; Matsubara, M.; Tsuchiya, S.; Shiota, M.; Ono, T.; Yamada, I.; Aoki-Kinoshita, K. F. *Nucleic Acids Res.* **2021,** *49,* D1529–D1533. doi:10.1093/nar/gkaa947
21. Joosten, R. P.; Long, F.; Murshudov, G. N.; Perrakis, A. *IUCrJ* **2014,** *1,* 213–220. doi:10.1107/s2052252514009324
22. van Beusekom, B.; Lütteke, T.; Joosten, R. P. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2018,** *74,* 463–472. doi:10.1107/s2053230x18004016
23. van Beusekom, B.; Wezel, N.; Hekkelman, M. L.; Perrakis, A.; Emsley, P.; Joosten, R. P. *Acta Crystallogr., Sect. D: Struct. Biol.* **2019,** *75,* 416–425. doi:10.1107/s2059798319003875
24. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. *Nucleic Acids Res.* **2007,** *35,* D301–D303. doi:10.1093/nar/gkl971
25. Matsubara, M.; Aoki-Kinoshita, K. F.; Aoki, N. P.; Yamada, I.; Narimatsu, H. *J. Chem. Inf. Model.* **2017,** *57,* 632–637. doi:10.1021/acs.jcim.6b00650
26. Dialpuri, J. S.; Bagdonas, H.; Schofield, L. C.; Pham, P. T.; Holland, L.; Bond, P. S.; Sánchez Rodríguez, F.; McNicholas, S. J.; Agirre, J. *Acta Crystallogr., Sect. F: Struct. Biol. Commun.* **2024,** *80,* 30–35. doi:10.1107/s2053230x24000359
27. Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975,** *97,* 1354–1358. doi:10.1021/ja00839a011
28. Kommoju, P.-R.; Chen, Z.-w.; Bruckner, R. C.; Mathews, F. S.; Jorns, M. S. *Biochemistry* **2011,** *50,* 5521–5534. doi:10.1021/bi200388g
29. Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; Varki, A.; Woods, R. J.; The SNFG Discussion Group. *Glycobiology* **2019,** *29,* 620–624. doi:10.1093/glycob/cwz045
30. Agirre, J.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D. *Curr. Opin. Struct. Biol.* **2017,** *44,* 39–47. doi:10.1016/j.sbi.2016.11.011

License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (https://www.beilstein-journals.org/bjoc/terms), which is identical to the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:
https://doi.org/10.3762/bjoc.20.83