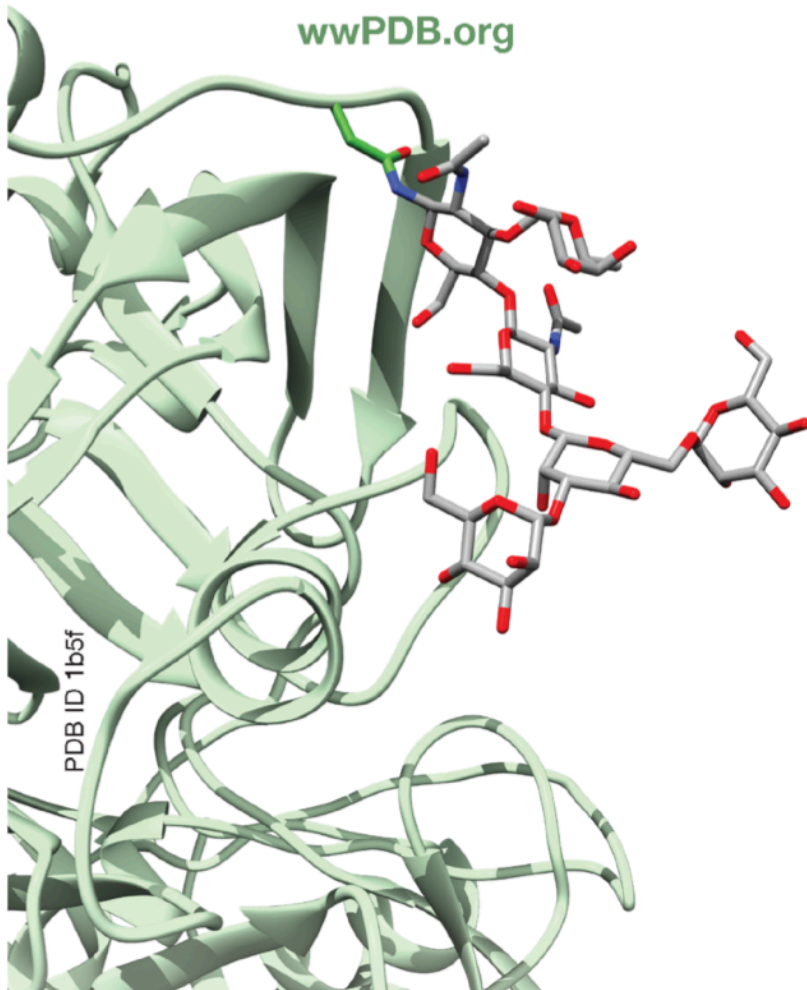


○横地 政志¹、岩田 武史¹、池川 恭代¹、Gert-Jan Bekker¹、宮ノ入 洋平¹、児嶋 長次郎²、
藤原 敏道¹、栗栖 源嗣¹

1. 大阪大学蛋白質研究所 (PDBj)、2. 横浜国立大学大学院工学研究院

PDBエントリーの糖鎖データ表現方法の改善



糖鎖：

1. 糖鎖は、蛋白質、核酸に並ぶ第3のポリマー
2. 多くの単糖がグリコシド結合を介して結合し、枝分かれした複雑な構造体
3. 糖鎖は、蛋白質や脂質と結合し、生理機能を発揮
4. 結晶試料中に、構造異性体（アノマー）、鎖状構造が共存

従来のリガンドに準じた表現（単糖ごとに別のAsym IDを与え、結合を逐一記述する方式）を改め、糖鎖構造の多様性の背後にある糖鎖化学を考慮して、糖鎖化学コミュニティの協力のもと、PDBエントリーの糖鎖構造データの表現方法を改善

<https://www.wwpdb.org/documentation/carbohydrate-remediation>

<https://www.wwpdb.org/news/news?year=2020#5f17634f0b582626e650e48b>

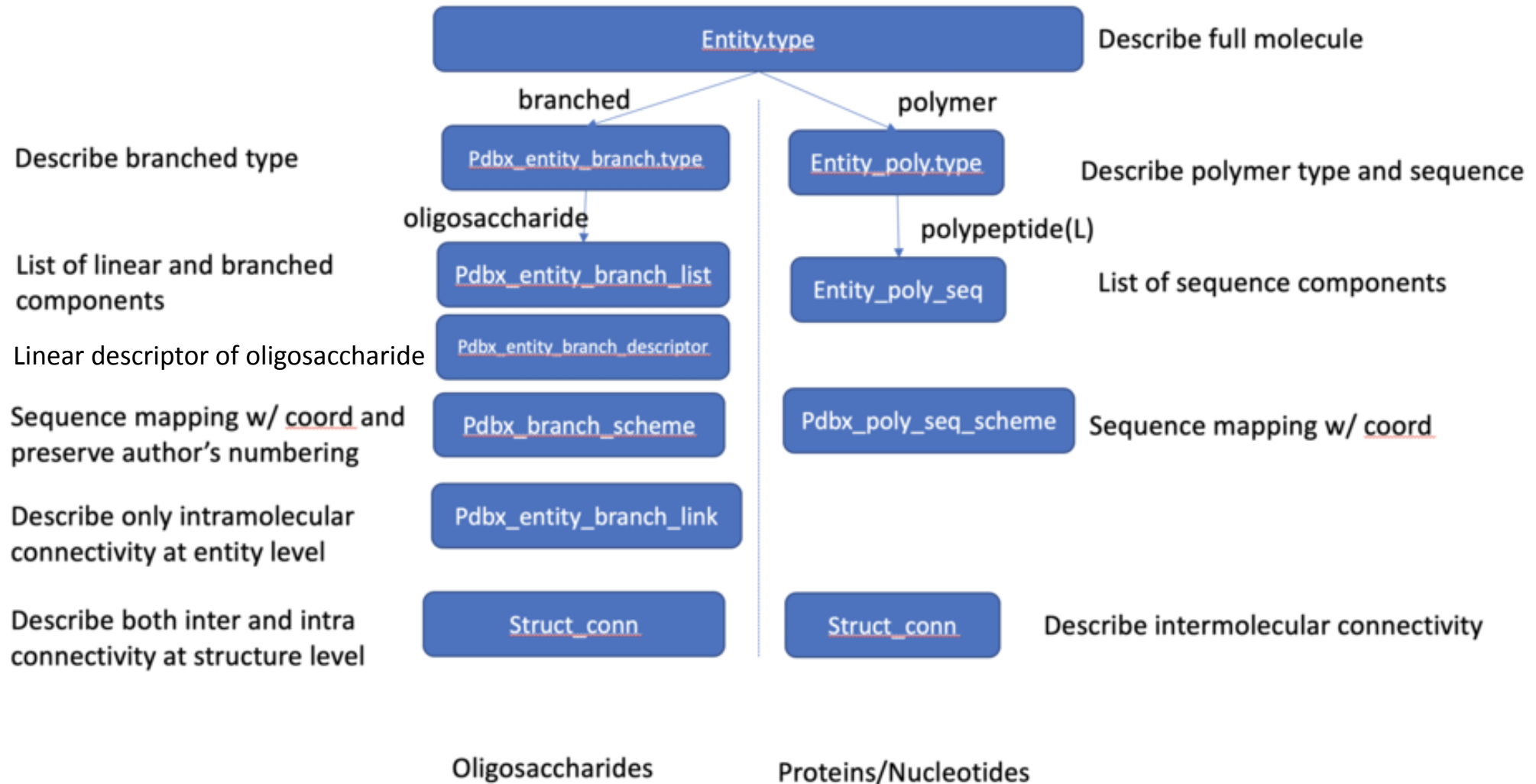
PDBエントリーの糖鎖データ表現方法の改善

糖鎖データ表現方法：

1. IUPAC/IUBMB勧告に従って各単糖の化合物辞書（CCD）を標準化
2. 糖鎖（分岐ポリマー）の統一表現方法の導入
3. 糖鎖科学コミュニティで受け入れられている糖鎖の線形記述子の採用
4. 糖鎖内の各グリコシド結合を、その他の糖鎖修飾結合の記述を分離して記述

PDBx/mmCIFフォーマットによる糖鎖表現：

分岐ポリマー（糖鎖）の記述と既存のポリマー（蛋白質／核酸）の記述の比較



PDBx/mmCIFフォーマットの糖鎖記述例（１）

```

loop_
  _entity.id
  _entity.type
  _entity.src_method
  _entity.pdbx_description
  _entity.formula_weight
  _entity.pdbx_number_of_molecules
  _entity.pdbx_ec
  _entity.pdbx_mutation
  _entity.pdbx_fragment
1 polymer    man 'FUCOLECTIN-RELATED PROTEIN'          66769.758 1  ? YES 'CATALYTIC MODULE, RESIDUES 31-589'
2 branched   nat "alpha-L-fucopyranose-(1-2)-beta-D-galactopyranose-(1-4)-[alpha-L-fucopyranose-(1-3)]2-acetamido-2-deoxy-beta-D-glucopyranose" 681.7    1  ?? ?
3 water      nat water                                18.015   334 ?? ?
#
_pdbx_entity_branch.entity_id 2
_pdbx_entity_branch.type      oligosaccharide
#
loop_
  _pdbx_entity_branch_list.entity_id
  _pdbx_entity_branch_list.comp_id
  _pdbx_entity_branch_list.num
  _pdbx_entity_branch_list.hetero
2 NAG 1  n
2 GAL 2  n
2 FUC 3  n
2 FUC 4  n
#

```

PDBx/mmCIFフォーマットの糖鎖記述例（2）

糖鎖内のグリコシド結合の記述： その他の糖鎖修飾結合の記述（リガンドと同様）：

```
loop_
_pdbx_entity_branch_link.link_id
_pdbx_entity_branch_link.entity_id
_pdbx_entity_branch_link.entity_branch_list_num_1
_pdbx_entity_branch_link.comp_id_1
_pdbx_entity_branch_link.atom_id_1
_pdbx_entity_branch_link.leaving_atom_id_1
_pdbx_entity_branch_link.atom_stereo_config_1
_pdbx_entity_branch_link.entity_branch_list_num_2
_pdbx_entity_branch_link.comp_id_2
_pdbx_entity_branch_link.atom_id_2
_pdbx_entity_branch_link.leaving_atom_id_2
_pdbx_entity_branch_link.atom_stereo_config_2
_pdbx_entity_branch_link.value_order
_pdbx_entity_branch_link.details
1 2 2 GAL C1 O1 1 NAG O4 HO4 sing ?
2 2 3 FUC C1 O1 2 GAL O2 HO2 sing ?
3 2 4 FUC C1 O1 1 NAG O3 HO3 sing ?
#
```

```
_struct_conn.id
_struct_conn.conn_type_id
_struct_conn.pdbx_leaving_atom_flag
_struct_conn.ptnr1_label_comp_id
_struct_conn.ptnr1_label_asym_id
_struct_conn.ptnr1_label_seq_id
_struct_conn.ptnr1_label_atom_id
_struct_conn.pdbx_ptnr1_label_alt_id
_struct_conn.pdbx_ptnr1_PDB_ins_code
_struct_conn.pdbx_ptnr1_standard_comp_id
_struct_conn.pdbx_ptnr1_leaving_atom_id
_struct_conn.pdbx_ptnr1_atom_stereo_config
_struct_conn.ptnr1_symmetry
_struct_conn.ptnr2_label_comp_id
_struct_conn.ptnr2_label_asym_id
_struct_conn.ptnr2_label_seq_id
_struct_conn.ptnr2_label_atom_id
_struct_conn.pdbx_ptnr2_label_alt_id
_struct_conn.pdbx_ptnr2_PDB_ins_code
_struct_conn.pdbx_ptnr2_standard_comp_id
_struct_conn.pdbx_ptnr2_leaving_atom_id
_struct_conn.pdbx_ptnr2_atom_stereo_config
_struct_conn.ptnr2_symmetry
_struct_conn.pdbx_role
_struct_conn.pdbx_dist_value
_struct_conn.pdbx_value_order
covale1 covale Y ASN A 93 ND2 ? ? ? OXT N 1_555 NAG C 801 C1 ? ? ? 01 R 1_555 N-Glycosylation 1.438 sing
covale2 covale Y ASN A 206 ND2 ? ? ? OXT N 1_555 NAG D 804 C1 ? ? ? 01 R 1_555 N-Glycosylation 1.415 sing
covale3 covale Y ASN B 93 ND2 ? ? ? OXT N 1_555 NAG E 801 C1 ? ? ? 01 R 1_555 N-Glycosylation 1.434 sing
#
```

PDBx/mmCIFフォーマットの糖鎖記述例（3）

```

loop_
_pdbx_entity_branch_descriptor.ordinal
_pdbx_entity_branch_descriptor.entity_id
_pdbx_entity_branch_descriptor.descriptor
_pdbx_entity_branch_descriptor.type
_pdbx_entity_branch_descriptor.program
_pdbx_entity_branch_descriptor.program_version
1 2 "LFucpa1-2DGalpb1-4[LFucpa1-3]DGlcpNAcb1-ROH"
    "Glycam Condensed Sequence" GMML 1.0
2 2
"WURCS=2.0/3,4,3/[a2122h-1b_1-5_2*NCC/3=0][a1221m-1a_1-5][a2112h-1b_1-5]/1-2-3-2/a3-b1_a4-c1_c2-d1" WURCS
    PDB2Glycan 1.1
3 2
"[] [b-D-GlcpNAc]{[(3+1)][a-L-Fucp]}{[(4+1)][b-D-Galp]}{[(2+1)][a-L-Fucp]}{}}}" LINUCS
    PDB-CARE ?

```

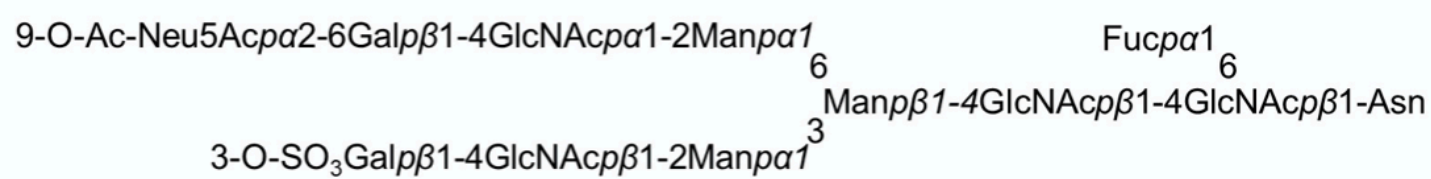
糖鎖科学コミュニティで一般的に受け入れられている線形記述子を採用したことにより、糖鎖構造を一意に特定できるようになりました。これには、米国ジョージア大学のGMMLが生成した短縮型IUPAC、日本の野口研究所のPDB2GlycanのWURCS、ドイツのpdb-careのLINUCSが含まれます。この糖鎖構造の表現（例えば、WURCS線形記述子）によって、PDBエントリーに含まれる糖鎖構造とGlyTouCanに収録された糖鎖情報を関連づけることが可能になりました。



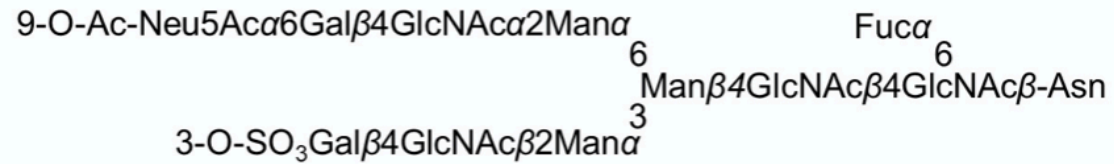
<https://glytoucan.org>

糖鎖の図示的表現 (SNFG: Symbol Nomenclature for Glycans)

FULL REPRESENTATION



SIMPLIFIED REPRESENTATION



SNFG REPRESENTATION

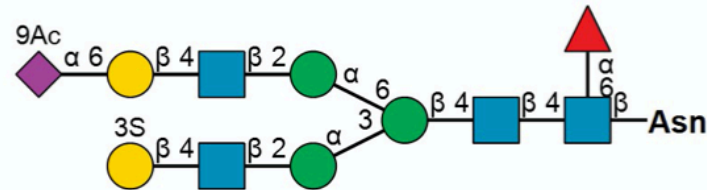


Table 1. Monosaccharide symbol nomenclature

SHAPE	White (Generic)	Blue	Green	Yellow	Orange
Filled Circle	 Hexose	 Glc	 Man	 Gal	 Gul
Filled Square	 HexNAc	 GlcNAc	 ManNAc	 GalNAc	 GulNAc
Crossed Square	 Hexosamine	 GlcN	 ManN	 GalN	 GulN
Divided Diamond	 Hexuronate	 GlcA	 ManA	 GalA	 GulA

Table. 1は一部抜粋

<https://www.ncbi.nlm.nih.gov/glycans/snfg.html>

糖鎖の図示的表現 (SNFG: Symbol Nomenclature for Glycans)

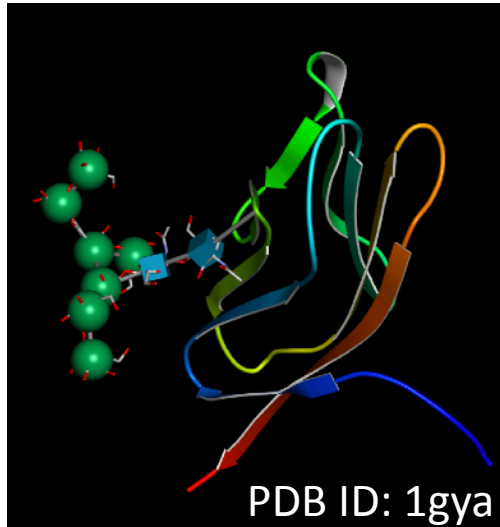
wwPDB検証レポート (PDB ID: 1b5f)

- Molecule 3 is an oligosaccharide called alpha-D-mannopyranose-(1-3)-beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-[alpha-L-fucopyranose-(1-3)]2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	E	5	60	34	2	24	0	0	0

分子ビューアーMolMilの3D-SNFG表現



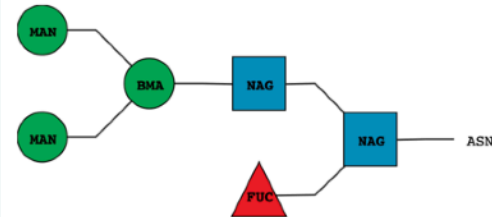
PDBjのエントリーページ (PDB ID: 1b5f)

G	alpha-D-mannopyranose-(1-3)-[alpha-D-mannopyranose-(1-6)]beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-[alpha-L-fucopyranose-(1-3)]2-acetamido-2-deoxy-beta-D-glucopyranose	branched	1057.0	1	PDB
H	alpha-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-[alpha-L-fucopyranose-(1-3)]2-acetamido-2-deoxy-beta-D-glucopyranose	branched	732.7	1	PDB
water	water		18.0	528	Chemie (HOH)

配列ビューア

Display chain:

G (alpha-D-mannopyranose-(1-3)-[alpha-D-mannopyranose-(1-6)]beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-[alpha-L-fucopyranose-(1-3)]2-acetamido-2-deoxy-beta-D-glu ...



非対称単位の内容

タンパク質・核酸鎖	鎖の数	4
	分子量合計	70969.2
糖鎖	分子数	4
	分子量合計	3417.2
全て*	分子量合計	74386.4

*水分子は含んでいません。

PDB/RDFアーカイブに糖鎖を記述、glycoinfo.orgと連携

```
<PDBo:has_pdbx_entity_branch_descriptor>
  <PDBo:pdbx_entity_branch_descriptor rdf:about="https://rdf.wwpdb.org/pdb/1B5F/pdbx_entity_branch_descriptor/2">
    <PDBo:of_datablock rdf:resource="https://rdf.wwpdb.org/pdb/1B5F"/>
    <PDBo:reference_to_entity>
      <rdf:Description rdf:about="https://rdf.wwpdb.org/pdb/1B5F/entity/3">
        <PDBo:referenced_by_pdbx_entity_branch_descriptor rdf:resource="https://rdf.wwpdb.org/pdb/1B5F/pdbx_entity
      </rdf:Description>
    </PDBo:reference_to_entity>
    <PDBo:pdbx_entity_branch_descriptor.ordinal>2</PDBo:pdbx_entity_branch_descriptor.ordinal>
    <PDBo:pdbx_entity_branch_descriptor.descriptor>WURCS=2.0/4,5,4/[a2122h-1b_1-5_2*NCC/3=0][a1221m-1a_1-5][a1122h-1
or.descriptor>
    <PDBo:pdbx_entity_branch_descriptor.entity_id>3</PDBo:pdbx_entity_branch_descriptor.entity_id>
    <PDBo:pdbx_entity_branch_descriptor.program>PDB2Glycan</PDBo:pdbx_entity_branch_descriptor.program>
    <PDBo:pdbx_entity_branch_descriptor.program_version>1.1.0</PDBo:pdbx_entity_branch_descriptor.program_version>
    <PDBo:pdbx_entity_branch_descriptor.type>WURCS</PDBo:pdbx_entity_branch_descriptor.type>
    <PDBo:link_to_glycoinfo rdf:resource="http://rdf.glycoinfo.org/glycan/G17689EW"
      rdfs:label="glytoucan:G17689EW"/>
    <rdfs:seeAlso rdf:resource="http://identifiers.org/glytoucan/G17689EW"
      rdfs:label="glytoucan:G17689EW"/>
  </PDBo:pdbx_entity_branch_descriptor>
</PDBo:has_pdbx_entity_branch_descriptor>
<PDBo:has_pdbx_entity_branch_descriptor>
```

PDB/RDFアーカイブに糖鎖を記述し、糖鎖のWURCS記述子（PDB ID: 1b5f, Entity ID: 3, WURCS=2.0から始まる値）とGlyTouCanの登録ID（G17689EW）を関連づけて検索できるようになりました。

NMR統合データファイルによるPDBへのNMR構造の登録（１）

OneDepシステムを介したPDBへNMR構造のデータ登録について、化学シフト、距離制限、およびピークリストのような各種NMR実験データを個別にアップロードする必要がありました。本年5月より、これに加えて、NMR-STAR¹またはNEF²形式の単一ファイルによるNMR実験データの受付けを開始しました。

NMR-STARは、NMRのデータを記述する公式なwwPDBのフォーマットです。一方、NEFは軽量なフォーマット及びディクショナリーで定義され、NMR構造決定において主流のソフトウェアがサポートしています。これら2つの相互変換可能な標準フォーマットを一つのデータファイルとして利用することにより、データの保管や配布が簡単になるだけでなく、登録過程が容易になりました。

1. Ulrich, E. L. *et al.* (2019) NMR-STAR: comprehensive ontology for representing, archiving and exchanging data from nuclear magnetic resonance spectroscopic experiments *Journal of Biomolecular NMR*, 73: 5–9. doi: [10.1007/s10858-018-0220-3](https://doi.org/10.1007/s10858-018-0220-3)

2. Gutmanas *et al.* (2015) NMR Exchange Format: a unified and open standard for representation of NMR restraint data *Nature Structural & Molecular Biology* 22: 433–434 doi: [10.1038/nsmb.3041](https://doi.org/10.1038/nsmb.3041)

NMR統合データファイルによるPDBへのNMR構造の登録（2）

OneDep wwPDB Deposition: D_800327 -- Requested IDs: PDB, BMRB

FAQ Tutorial

WORLDWIDE PDB PROTEIN DATA BANK

List requirements

All items

Mandatory items

Navigation

- Instructions
- Communication
- Re-upload files
- Upload summary
- Admin
 - Contact information
 - Grant information
 - Release status
 - Entry title & author
 - Citation information
- Macromolecules
 - 1) Uncharacterized protein
- NMR experimental
 - NMR samples
 - NMR data collection
- NMR software
 - NMR software
- NMR data and refinement
 - Peak lists
 - Chemical shift references
 - Chemical shift connection

Found total 5 warnings in NMR data.

Warning: Anomalous chemical shift, Total: 2

Row/Category/Saveframe	Description
Row of atom_name: C, residue_name: ALA, sequence_code: 54, chain_code: A, Category: nef_chemical_shift, Saveframe: nef_chemical_shift_list_bmr15702.str	Verify chemical shift value for A:54:ALA:C (167.4 ppm, 5.01 sigma) is outside of expected range (188.21 ~ 167.43 ppm, avg 177.817, std 2.078, min 164.477, max 187.2). Neither aromatic ring nor paramagnetic atom were found in the vicinity.
Saveframe: nef_chemical_shift_list_bmr15702.str	trans-peptide bond of A:37:PRO can not be verified with the assigned chemical shift values (CB 33.6 ppm, cis_trans_pred cis 81.8 (%), trans 18.2 (%)).

Warning: Encouragement, Total: 1

Description
The wwPDB NMR Validation Task Force strongly encourages the submission of spectral peak lists, in particular those generated from NOESY spectra.

Warning: Remarkable data, Total: 2

Row/Category/Saveframe	Description
Row of atom_name: HG, residue_name: SER, sequence_code: 55, chain_code: A, Category: nef_chemical_shift, Saveframe: nef_chemical_shift_list_bmr15702.str	value 5.66 is remarkable assignment. Occurrence of HG in SER is 1.1 % according to BMRB.
Row of atom_name: HG1, residue_name: THR, sequence_code: 123, chain_code: A,	value 5.94 is remarkable assignment. Occurrence of HG1 in THR is 2.2 %

Version: V4.5

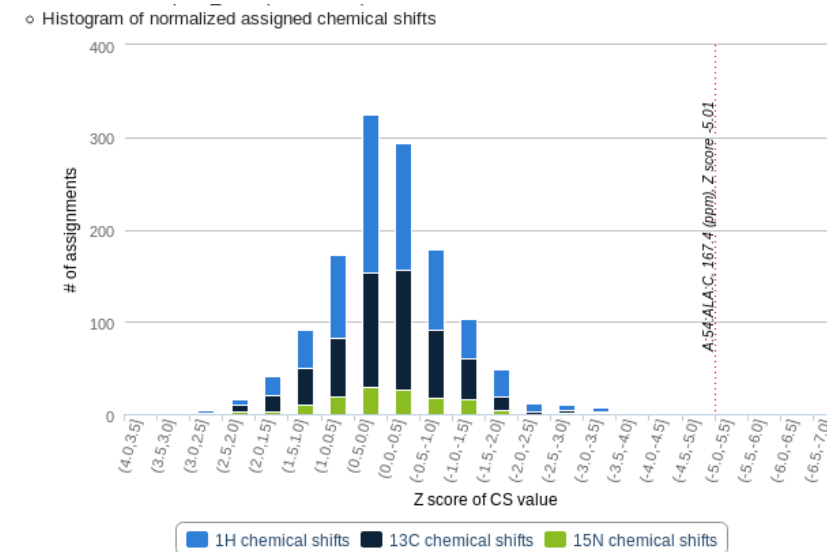
NMR統合データファイルは、エンティティー構成に関わる情報（配列、修飾、共有結合）、帰属された化学シフト、実験的に決定された各種制限情報（距離制限、二面角制限、残余双極子結合）、ピークリストを含みます。

OneDepへデータアップロードする際に、フォーマット、それぞれの辞書によるデータ検証、同時にアップロードされた構造モデルとの交差検証が行われ、ユーザーに結果が返されます。

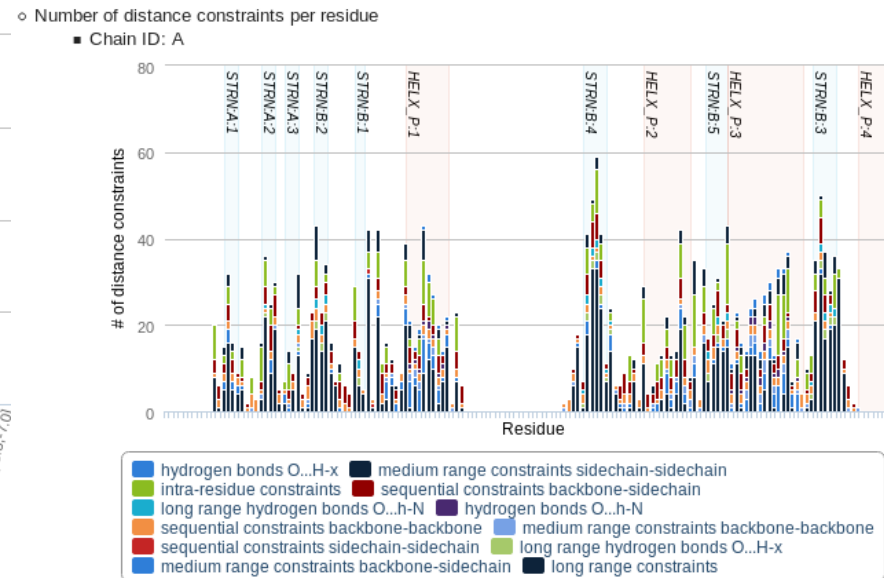
NMR統合データファイルによるPDBへのNMR構造の登録（3）

アップロードされたNMR実験データは、OneDepのページ上で可視化されます。一部の項目は自動的に入力され、登録作業の負担を軽減します。また、公開済みのエントリーのNMR統合データは、PDBの所定のFTPエリアからダウンロード可能になります。

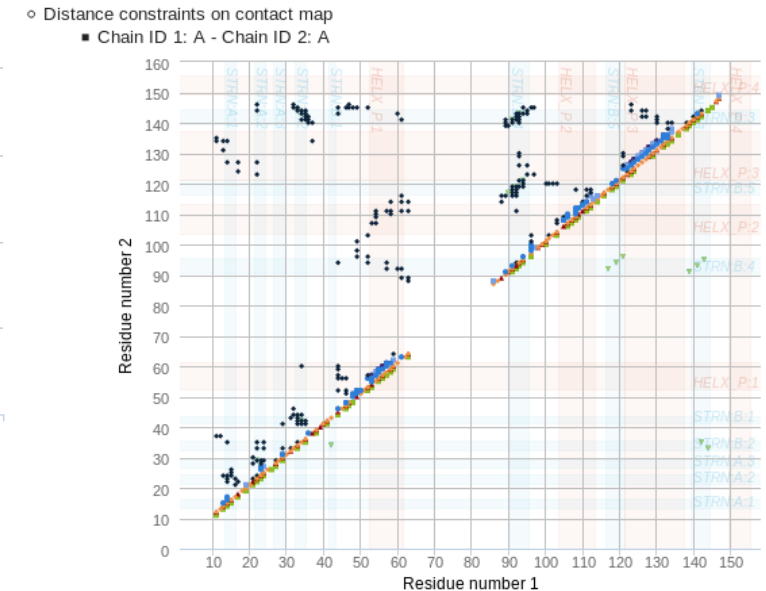
異常な化学シフトの検出



距離制限情報（配列上の度数分布）

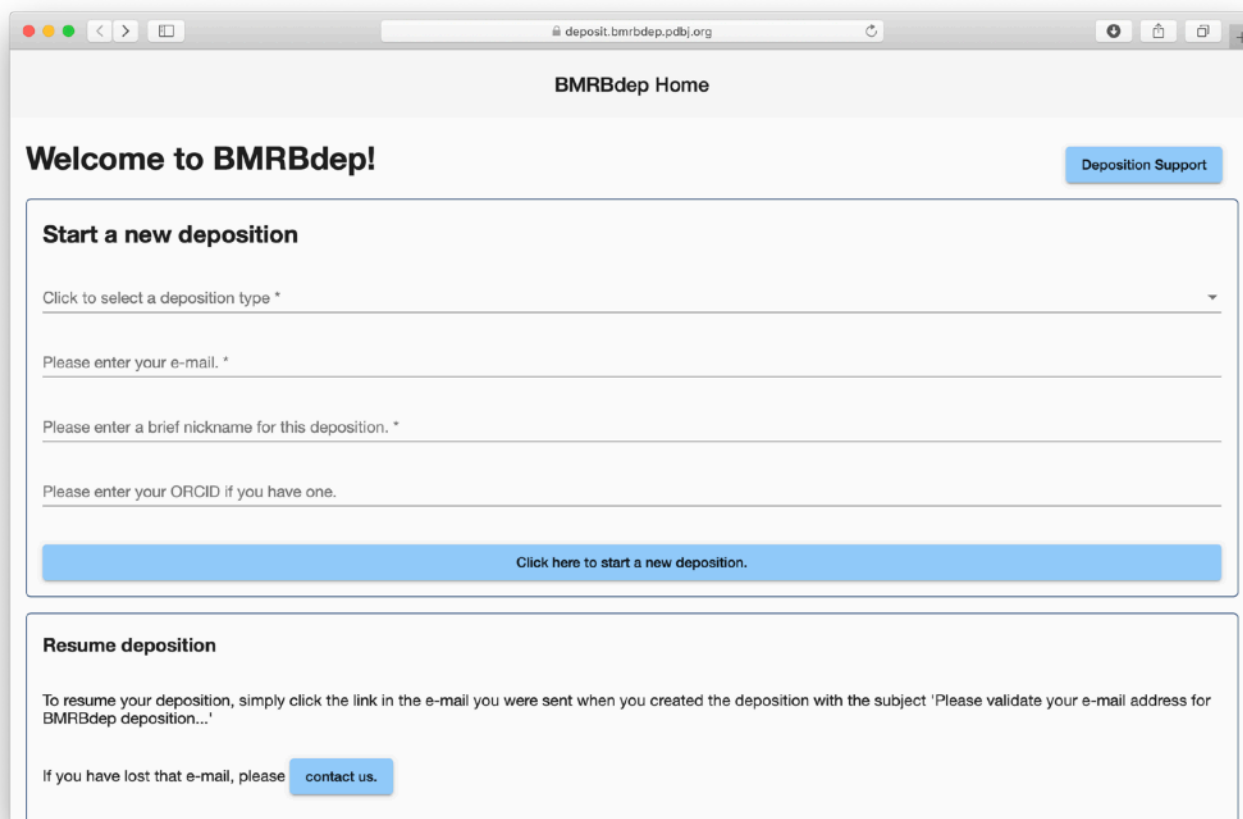


距離制限情報（コンタクトマップ形式）



BMRBdep: (構造モデルを伴わない) NMR実験データの登録システム

構造モデルを伴わないNMR実験データのBMRB登録には、従来ADIT-NMRシステムが利用されてきましたが、新しい登録システムBMRBdepに移行しました。



The screenshot shows the BMRBdep Home page in a web browser. The page has a header with the BMRBdep logo and a 'Deposition Support' button. The main content area is titled 'Welcome to BMRBdep!' and contains two sections: 'Start a new deposition' and 'Resume deposition'. The 'Start a new deposition' section includes a dropdown menu for 'Click to select a deposition type *', an email input field 'Please enter your e-mail. *', a nickname input field 'Please enter a brief nickname for this deposition. *', and an ORCID input field 'Please enter your ORCID if you have one.'. Below these fields is a large blue button labeled 'Click here to start a new deposition.'. The 'Resume deposition' section includes instructions on how to resume a deposition and a 'contact us' button.

登録作業の負担軽減：

- 辞書ベースの入力値のチェック、自動補完
- セッションの逐次保存、結果の反映
- NMR-STARファイルを用いた自動入力
- クローンエントリーの生成（連続登録時の入力支援）

<https://deposit.bmrbddep.pdbj.org>