



Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 01:32 AM GMT

PDB ID : 2DRJ
Title : Xray structure of alpha-2,3/8-sialyltransferaseCstII F91Y mutant
Authors : Chiu, C.P.C.; Strynadka, N.C.J.
Deposited on : 2006-06-09
Resolution : 2.25 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

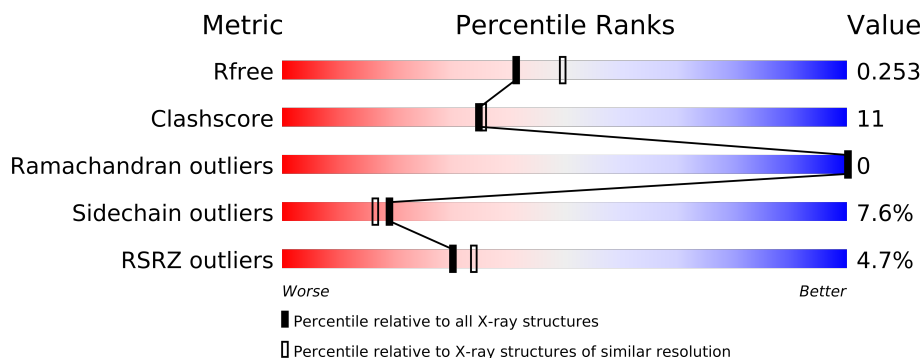
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.25 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	1108 (2.28-2.24)
Clashscore	79885	1326 (2.28-2.24)
Ramachandran outliers	78287	1291 (2.28-2.24)
Sidechain outliers	78261	1291 (2.28-2.24)
RSRZ outliers	66119	1110 (2.28-2.24)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	262	

2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 2091 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

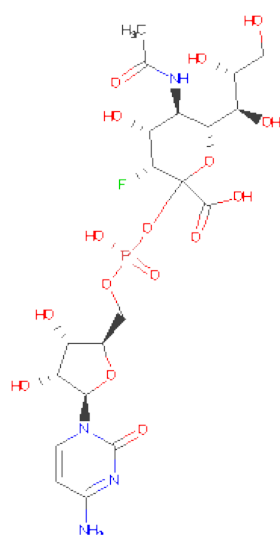
- Molecule 1 is a protein called Alpha-2,3/8-sialyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	233	1949	1282	304	355	8	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

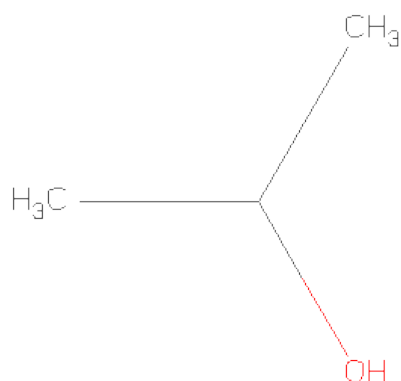
Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q9LAK3
A	-1	SER	-	EXPRESSION TAG	UNP Q9LAK3
A	0	HIS	-	EXPRESSION TAG	UNP Q9LAK3
A	53	SER	ILE	ENGINEERED	UNP Q9LAK3
A	91	TYR	PHE	ENGINEERED	UNP Q9LAK3
A	222	GLY	GLU	ENGINEERED	UNP Q9LAK3

- Molecule 2 is CYTIDINE-5'-MONOPHOSPHATE-3-FLUORO-N-ACETYL-NEURAMINI CACID (three-letter code: CSF) (formula: C₂₀H₃₀FN₄O₁₆P).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	F	N	O	P	0	0
			42	20	1	4	16	1		

- Molecule 3 is ISOPROPYL ALCOHOL (three-letter code: IPA) (formula: C₃H₈O).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	3	1		
3	A	1	Total	C	O	0	0
			4	3	1		

- Molecule 4 is water.

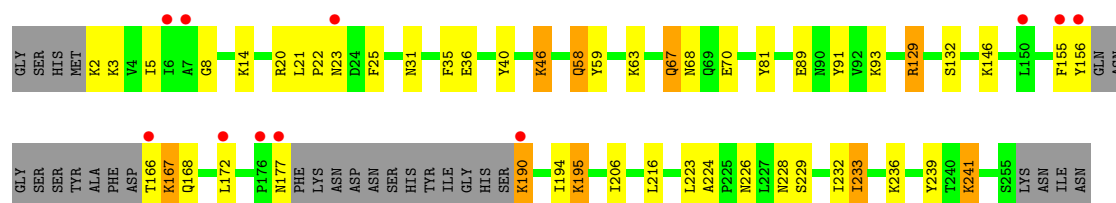
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	92	Total	O	0	0
			92	92		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Alpha-2,3/8-sialyltransferase

Chain A: 



4 Data and refinement statistics

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α , β , γ	116.65Å 116.65Å 44.09Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	23.81 – 2.25 23.81 – 2.22	Depositor EDS
% Data completeness (in resolution range)	100.0 (23.81-2.25) 92.6 (23.81-2.22)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.19 (at 2.22Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.194 , 0.261 0.190 , 0.253	Depositor DCC
R_{free} test set	701 reflections (5.36%)	DCC
Wilson B-factor (Å ²)	27.6	Xtriage
Anisotropy	0.893	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 32.3	EDS
Estimated twinning fraction	0.025 for -k,-h,-l	Xtriage
L-test for twinning	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 13791 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	2091	wwPDB-VP
Average B, all atoms (Å ²)	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.95% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CSF, IPA

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	1.02	1/2001 (0.0%)	0.82	0/2698

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	81	TYR	CD1-CE1	5.25	1.47	1.39

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	23	ASN	Peptide

5.2 Close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the

Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1949	0	1901	41	0
2	A	42	0	29	5	0
3	A	8	0	16	4	0
4	A	92	0	0	1	0
All	All	2091	0	1946	41	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 11.

All (41) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:129:ARG:HG2	1:A:129:ARG:HH21	0.97	1.07
1:A:129:ARG:HG2	1:A:129:ARG:NH2	1.76	0.97
1:A:129:ARG:CG	1:A:129:ARG:HH21	1.86	0.89
1:A:2:LYS:HG2	1:A:3:LYS:H	1.47	0.80
1:A:31:ASN:HD22	2:A:1001:CSF:H5	1.38	0.64
1:A:166:THR:HG23	1:A:167:LYS:CE	2.31	0.60
1:A:166:THR:HG23	1:A:167:LYS:HE3	1.83	0.60
1:A:129:ARG:NH2	1:A:129:ARG:CG	2.54	0.59
1:A:236:LYS:HE2	1:A:239:TYR:CD2	2.39	0.58
1:A:5:ILE:HD13	1:A:25:PHE:CD1	2.39	0.57
1:A:190:LYS:O	1:A:194:ILE:HD12	2.06	0.56
1:A:233:ILE:HD12	1:A:233:ILE:N	2.20	0.56
1:A:5:ILE:HD13	1:A:25:PHE:CG	2.40	0.55
1:A:91:TYR:CB	3:A:2001:IPA:H13	2.37	0.55
1:A:14:LYS:HG2	1:A:40:TYR:CZ	2.47	0.49
1:A:91:TYR:CD1	3:A:2001:IPA:H13	2.48	0.48
1:A:91:TYR:HB3	3:A:2001:IPA:H13	1.95	0.48
1:A:2:LYS:HG2	1:A:3:LYS:N	2.22	0.47
1:A:226:ASN:OD1	1:A:228:ASN:N	2.27	0.47
1:A:206:ILE:C	1:A:206:ILE:HD12	2.36	0.46
1:A:146:LYS:HA	1:A:206:ILE:HG22	1.96	0.46
1:A:177:ASN:HB3	2:A:1001:CSF:H111	1.98	0.46
1:A:59:TYR:CZ	1:A:63:LYS:HE3	2.50	0.46
1:A:91:TYR:HD1	3:A:2001:IPA:H13	1.80	0.45
1:A:68:ASN:ND2	1:A:70:GLU:OE2	2.41	0.45
1:A:233:ILE:CD1	1:A:233:ILE:N	2.80	0.45
1:A:35:PHE:CD1	1:A:35:PHE:N	2.85	0.45
1:A:155:PHE:N	1:A:156:TYR:HA	2.32	0.44
1:A:3:LYS:HD3	1:A:22:PRO:HB3	2.00	0.44
1:A:8:GLY:HA3	2:A:1001:CSF:H4'	1.99	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:58:GLN:HE21	1:A:58:GLN:HA	1.81	0.44
1:A:195:LYS:HD2	1:A:195:LYS:HA	1.71	0.43
1:A:46:LYS:HE2	4:A:2089:HOH:O	2.18	0.43
1:A:5:ILE:HD11	1:A:22:PRO:HG2	2.00	0.43
1:A:132:SER:OG	2:A:1001:CSF:OAA	2.34	0.42
1:A:35:PHE:O	1:A:36:GLU:C	2.58	0.42
1:A:168:GLN:HG3	1:A:241:LYS:O	2.20	0.41
1:A:21:LEU:HA	1:A:22:PRO:HD2	1.82	0.41
1:A:67:GLN:HE21	1:A:67:GLN:HB2	1.66	0.40
1:A:156:TYR:CE1	2:A:1001:CSF:H2'	2.57	0.40
1:A:20:ARG:HD3	1:A:224:ALA:O	2.21	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	227/262 (87%)	217 (96%)	10 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	211/236 (89%)	195 (92%)	16 (8%)	19	16

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	46	LYS
1	A	58	GLN
1	A	67	GLN
1	A	89	GLU
1	A	93	LYS
1	A	129	ARG
1	A	167	LYS
1	A	172	LEU
1	A	190	LYS
1	A	195	LYS
1	A	216	LEU
1	A	223	LEU
1	A	229	SER
1	A	232	ILE
1	A	233	ILE
1	A	241	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (6) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	31	ASN
1	A	58	GLN
1	A	67	GLN
1	A	103	HIS
1	A	128	GLN
1	A	238	ASN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	CSF	A	1001	1	44,44,44	2.24	11 (25%)	63,67,67	2.23	15 (23%)
3	IPA	A	2001	-	3,3,3	0.60	0	3,3,3	0.42	0
3	IPA	A	2002	-	3,3,3	0.72	0	3,3,3	0.27	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	CSF	A	1001	1	-	0/30/75/75	0/3/3/3
3	IPA	A	2001	-	-	0/0/0/0	0/0/0/0
3	IPA	A	2002	-	-	0/0/0/0	0/0/0/0

All (11) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1001	CSF	O4'-C1'	7.27	1.52	1.41
2	A	1001	CSF	C2-N1	6.21	1.45	1.38
2	A	1001	CSF	C3A-C4A	5.62	1.56	1.51
2	A	1001	CSF	C6-N1	4.06	1.42	1.35
2	A	1001	CSF	O6A-C2A	3.99	1.47	1.41
2	A	1001	CSF	C4-N3	3.33	1.42	1.35
2	A	1001	CSF	O1A-C2A	2.77	1.50	1.43
2	A	1001	CSF	C2-N3	2.61	1.42	1.35
2	A	1001	CSF	PA-O1A	2.07	1.67	1.59
2	A	1001	CSF	C2A-C3A	2.03	1.55	1.53
2	A	1001	CSF	C2A-C1A	-2.02	1.52	1.55

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1001	CSF	C6-C5-C4	8.06	120.82	117.47
2	A	1001	CSF	F3A-C3A-C4A	5.60	113.23	108.58
2	A	1001	CSF	C6A-C5A-N5A	-4.64	102.89	110.99
2	A	1001	CSF	C7A-C6A-C5A	-4.45	107.68	114.24
2	A	1001	CSF	O4'-C1'-N1	4.32	117.17	108.06
2	A	1001	CSF	C4A-C5A-N5A	3.95	118.92	110.56
2	A	1001	CSF	C2-N3-C4	3.54	120.69	115.57
2	A	1001	CSF	O4A-C4A-C5A	3.29	116.36	109.61
2	A	1001	CSF	O6A-C6A-C7A	3.24	110.52	106.34
2	A	1001	CSF	C3A-C2A-C1A	-3.02	106.15	112.06
2	A	1001	CSF	C3'-C2'-C1'	2.86	105.38	100.91
2	A	1001	CSF	OAA-C1A-C2A	-2.61	116.05	123.16
2	A	1001	CSF	C11-C10-N5A	-2.60	111.02	116.11
2	A	1001	CSF	PA-O1A-C2A	-2.36	121.04	125.56
2	A	1001	CSF	O10-C10-N5A	2.32	126.75	121.90

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	233/262 (88%)	-0.07	11 (4%) 30 34	16, 30, 51, 70	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	177	ASN	4.3
1	A	156	TYR	4.3
1	A	23	ASN	4.0
1	A	166	THR	3.1
1	A	150	LEU	2.7
1	A	176	PRO	2.4
1	A	155	PHE	2.2
1	A	7	ALA	2.2
1	A	190	LYS	2.2
1	A	6	ILE	2.1
1	A	172	LEU	2.0

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains.

The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å ²)	Q<0.9
3	IPA	A	2002	4/4	0.14	0.50	43,43,43,44	0
3	IPA	A	2001	4/4	0.13	0.41	27,28,28,29	0
2	CSF	A	1001	42/42	0.18	0.37	44,53,63,64	0

6.5 Other polymers ⓘ

There are no such residues in this entry.