



Full wwPDB X-ray Structure Validation Report ⓘ

Feb 19, 2016 – 07:48 PM GMT

PDB ID : 4U61
Title : Trichodysplasia spinulosa-associated polyomavirus (TSPyV) VP1 in complex with 6'-sialyllactose
Authors : Stroh, L.J.; Stehle, T.
Deposited on : 2014-07-26
Resolution : 1.65 Å(reported)

This is a Full wwPDB X-ray Structure 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/validation/2016/XrayValidationReportHelp>
with specific help available everywhere you see the ⓘ symbol.

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

MolProbity : 4.02b-467
Mogul : 1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026982
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20026982

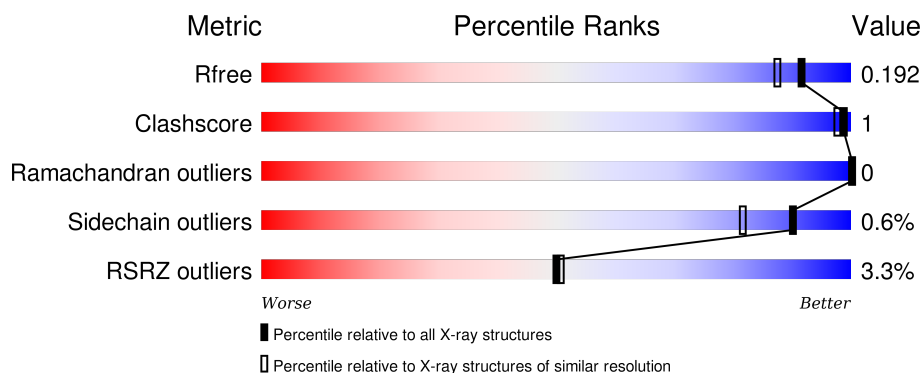
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.65 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}	91344	1226 (1.66-1.66)
Clashscore	102246	1323 (1.66-1.66)
Ramachandran outliers	100387	1295 (1.66-1.66)
Sidechain outliers	100360	1295 (1.66-1.66)
RSRZ outliers	91569	1227 (1.66-1.66)

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. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	280	<div> <div>3%</div> <div>93%</div> <div>• •</div> </div>
1	B	280	<div> <div>2%</div> <div>91%</div> <div>• 6%</div> </div>
1	C	280	<div> <div>2%</div> <div>94%</div> <div>• •</div> </div>
1	D	280	<div> <div>2%</div> <div>92%</div> <div>• 6%</div> </div>
1	E	280	<div> <div>5%</div> <div>93%</div> <div>• •</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	280	
1	G	280	
1	H	280	
1	I	280	
1	J	280	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	SIA	B	401	-	-	-	X
3	SIA	C	402	-	-	-	X
3	SIA	E	401	-	-	-	X
3	SIA	H	402	-	-	-	X
3	SIA	I	401	-	-	-	X
3	SIA	J	401	-	-	-	X
4	GOL	E	402	-	-	-	X
4	GOL	E	403	-	-	-	X

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 23184 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Structural protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	271	Total	C	N	O	S	0	3	0
			2119	1337	353	414	15			
1	B	264	Total	C	N	O	S	0	2	0
			2057	1298	343	403	13			
1	C	270	Total	C	N	O	S	0	4	0
			2100	1324	347	414	15			
1	D	264	Total	C	N	O	S	0	1	0
			2043	1291	342	397	13			
1	E	272	Total	C	N	O	S	0	2	0
			2114	1330	353	416	15			
1	F	264	Total	C	N	O	S	0	1	0
			2035	1285	337	400	13			
1	G	270	Total	C	N	O	S	0	0	0
			2084	1310	349	411	14			
1	H	264	Total	C	N	O	S	0	1	0
			2052	1295	342	402	13			
1	I	271	Total	C	N	O	S	0	1	0
			2100	1323	352	412	13			
1	J	264	Total	C	N	O	S	0	0	0
			2031	1284	337	398	12			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	24	GLY	-	expression tag	UNP E2ESL7
A	25	SER	-	expression tag	UNP E2ESL7
A	26	HIS	-	expression tag	UNP E2ESL7
A	27	MET	-	expression tag	UNP E2ESL7
A	28	ALA	-	expression tag	UNP E2ESL7
A	29	SER	-	expression tag	UNP E2ESL7
B	24	GLY	-	expression tag	UNP E2ESL7
B	25	SER	-	expression tag	UNP E2ESL7
B	26	HIS	-	expression tag	UNP E2ESL7

Continued on next page...

Continued from previous page...

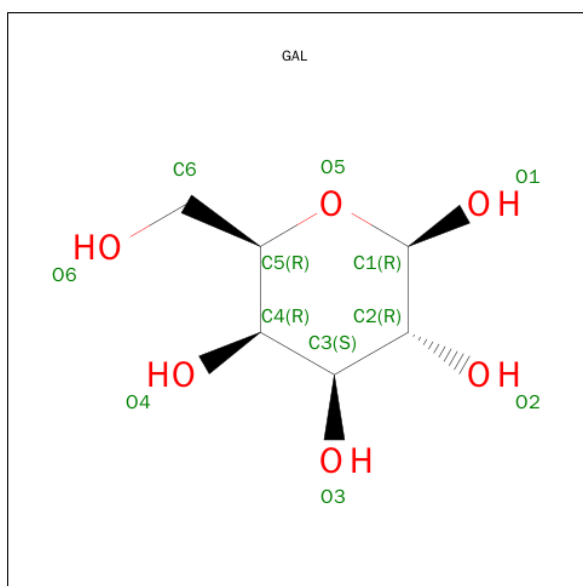
Chain	Residue	Modelled	Actual	Comment	Reference
B	27	MET	-	expression tag	UNP E2ESL7
B	28	ALA	-	expression tag	UNP E2ESL7
B	29	SER	-	expression tag	UNP E2ESL7
C	24	GLY	-	expression tag	UNP E2ESL7
C	25	SER	-	expression tag	UNP E2ESL7
C	26	HIS	-	expression tag	UNP E2ESL7
C	27	MET	-	expression tag	UNP E2ESL7
C	28	ALA	-	expression tag	UNP E2ESL7
C	29	SER	-	expression tag	UNP E2ESL7
D	24	GLY	-	expression tag	UNP E2ESL7
D	25	SER	-	expression tag	UNP E2ESL7
D	26	HIS	-	expression tag	UNP E2ESL7
D	27	MET	-	expression tag	UNP E2ESL7
D	28	ALA	-	expression tag	UNP E2ESL7
D	29	SER	-	expression tag	UNP E2ESL7
E	24	GLY	-	expression tag	UNP E2ESL7
E	25	SER	-	expression tag	UNP E2ESL7
E	26	HIS	-	expression tag	UNP E2ESL7
E	27	MET	-	expression tag	UNP E2ESL7
E	28	ALA	-	expression tag	UNP E2ESL7
E	29	SER	-	expression tag	UNP E2ESL7
F	24	GLY	-	expression tag	UNP E2ESL7
F	25	SER	-	expression tag	UNP E2ESL7
F	26	HIS	-	expression tag	UNP E2ESL7
F	27	MET	-	expression tag	UNP E2ESL7
F	28	ALA	-	expression tag	UNP E2ESL7
F	29	SER	-	expression tag	UNP E2ESL7
G	24	GLY	-	expression tag	UNP E2ESL7
G	25	SER	-	expression tag	UNP E2ESL7
G	26	HIS	-	expression tag	UNP E2ESL7
G	27	MET	-	expression tag	UNP E2ESL7
G	28	ALA	-	expression tag	UNP E2ESL7
G	29	SER	-	expression tag	UNP E2ESL7
H	24	GLY	-	expression tag	UNP E2ESL7
H	25	SER	-	expression tag	UNP E2ESL7
H	26	HIS	-	expression tag	UNP E2ESL7
H	27	MET	-	expression tag	UNP E2ESL7
H	28	ALA	-	expression tag	UNP E2ESL7
H	29	SER	-	expression tag	UNP E2ESL7
I	24	GLY	-	expression tag	UNP E2ESL7
I	25	SER	-	expression tag	UNP E2ESL7
I	26	HIS	-	expression tag	UNP E2ESL7

Continued on next page...

Continued from previous page...

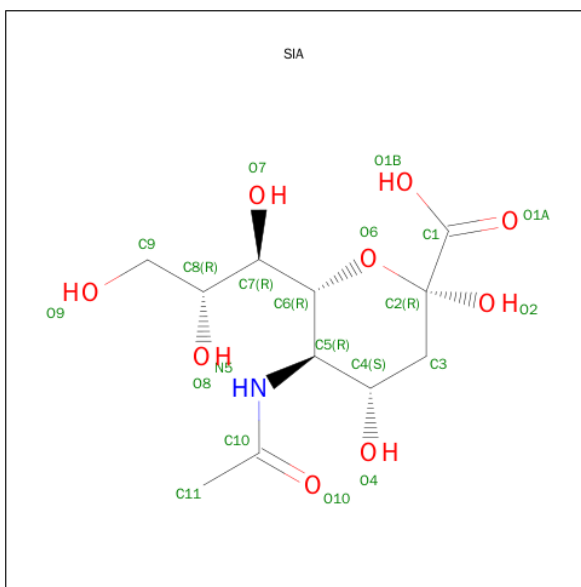
Chain	Residue	Modelled	Actual	Comment	Reference
I	27	MET	-	expression tag	UNP E2ESL7
I	28	ALA	-	expression tag	UNP E2ESL7
I	29	SER	-	expression tag	UNP E2ESL7
J	24	GLY	-	expression tag	UNP E2ESL7
J	25	SER	-	expression tag	UNP E2ESL7
J	26	HIS	-	expression tag	UNP E2ESL7
J	27	MET	-	expression tag	UNP E2ESL7
J	28	ALA	-	expression tag	UNP E2ESL7
J	29	SER	-	expression tag	UNP E2ESL7

- Molecule 2 is BETA-D-GALACTOSE (three-letter code: GAL) (formula: $C_6H_{12}O_6$).



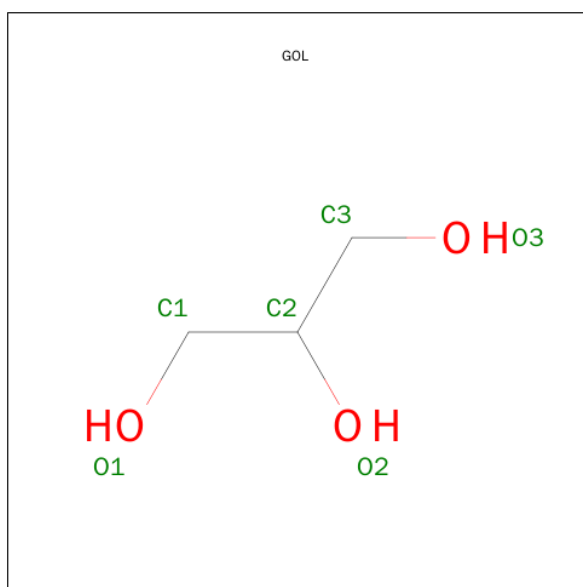
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		

- Molecule 3 is O-SIALIC ACID (three-letter code: SIA) (formula: $C_{11}H_{19}NO_9$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			20	11	1	8		
3	B	1	Total	C	N	O	0	0
			21	11	1	9		
3	C	1	Total	C	N	O	0	0
			21	11	1	9		
3	E	1	Total	C	N	O	0	0
			21	11	1	9		
3	H	1	Total	C	N	O	0	0
			21	11	1	9		
3	I	1	Total	C	N	O	0	0
			21	11	1	9		
3	J	1	Total	C	N	O	0	0
			21	11	1	9		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	C	1	Total	C	O	0	0
			6	3	3		
4	E	1	Total	C	O	0	0
			6	3	3		
4	E	1	Total	C	O	0	0
			6	3	3		
4	H	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	263	Total	O	0	0
			263	263		
5	B	223	Total	O	0	0
			223	223		
5	C	238	Total	O	0	0
			238	238		
5	D	245	Total	O	0	0
			245	245		
5	E	233	Total	O	0	0
			233	233		
5	F	219	Total	O	0	0
			219	219		
5	G	203	Total	O	0	0
			203	203		
5	H	207	Total	O	0	0
			207	207		

Continued on next page...

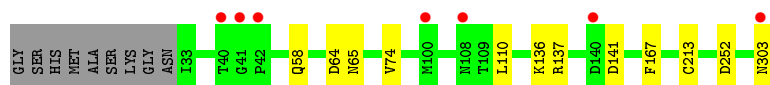
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	I	216	Total 216	O 216	0	0
5	J	220	Total 220	O 220	0	0

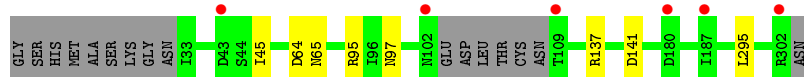
3 Residue-property plots [i](#)

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.

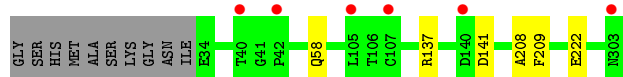
- Molecule 1: Structural protein VP1



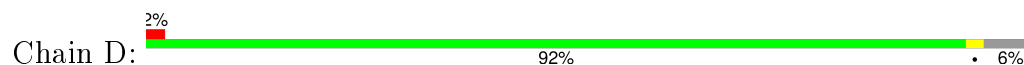
- Molecule 1: Structural protein VP1



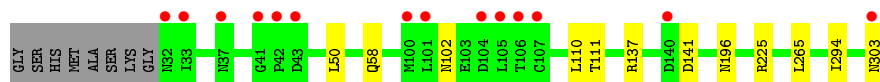
- Molecule 1: Structural protein VP1



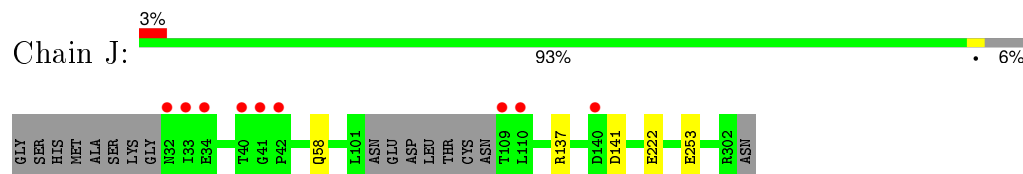
- Molecule 1: Structural protein VP1



- Molecule 1: Structural protein VP1



- Molecule 1: Structural protein VP1



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	139.56Å 146.35Å 151.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 1.65 29.92 – 1.65	Depositor EDS
% Data completeness (in resolution range)	99.7 (30.00-1.65) 99.8 (29.92-1.65)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.62 (at 1.65Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.165 , 0.184 0.174 , 0.192	Depositor DCC
R_{free} test set	18570 reflections (5.30%)	DCC
Wilson B-factor (Å ²)	19.0	Xtriage
Anisotropy	0.418	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 39.2	EDS
Estimated twinning fraction	0.016 for -h,l,k	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 368927 reflections	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	23184	wwPDB-VP
Average B, all atoms (Å ²)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.09% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, SIA, GAL

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	0.42	0/2173	0.67	1/2952 (0.0%)
1	B	0.44	0/2107	0.67	0/2865
1	C	0.43	0/2156	0.67	0/2929
1	D	0.42	0/2087	0.66	0/2838
1	E	0.42	0/2165	0.66	2/2943 (0.1%)
1	F	0.38	0/2082	0.64	0/2834
1	G	0.39	0/2129	0.66	0/2897
1	H	0.38	0/2099	0.62	0/2854
1	I	0.42	0/2148	0.66	1/2922 (0.0%)
1	J	0.41	0/2075	0.66	0/2824
All	All	0.41	0/21221	0.66	4/28858 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	225	ARG	NE-CZ-NH2	-5.80	117.40	120.30
1	I	225	ARG	NE-CZ-NH2	-5.33	117.63	120.30
1	E	225	ARG	NE-CZ-NH1	5.29	122.95	120.30
1	A	252	ASP	CB-CG-OD1	5.03	122.83	118.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

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 hydrogen

atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2119	0	2083	6	0
1	B	2057	0	2004	4	0
1	C	2100	0	2044	3	0
1	D	2043	0	1992	2	0
1	E	2114	0	2053	6	0
1	F	2035	0	1960	1	0
1	G	2084	0	2015	1	0
1	H	2052	0	1998	3	0
1	I	2100	0	2048	3	0
1	J	2031	0	1966	3	0
2	A	12	0	11	0	0
3	A	20	0	17	1	0
3	B	21	0	18	0	0
3	C	21	0	18	0	0
3	E	21	0	18	0	0
3	H	21	0	18	0	0
3	I	21	0	18	0	0
3	J	21	0	18	0	0
4	C	6	0	8	0	0
4	E	12	0	16	2	0
4	H	6	0	8	0	0
5	A	263	0	0	1	0
5	B	223	0	0	0	0
5	C	238	0	0	0	0
5	D	245	0	0	0	0
5	E	233	0	0	0	0
5	F	219	0	0	0	0
5	G	203	0	0	0	0
5	H	207	0	0	1	0
5	I	216	0	0	0	0
5	J	220	0	0	0	0
All	All	23184	0	20331	32	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 1.

All (32) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:208:ALA:C	1:C:209:PHE:CA	2.46	0.84
1:I:50:LEU:HD11	1:I:267[B]:VAL:HG21	1.80	0.62
1:E:102:ASN:HD21	1:E:111:THR:H	1.47	0.61
1:J:222:GLU:H	1:J:222:GLU:CD	2.06	0.59
1:E:196:ASN:HD21	4:E:403:GOL:C3	2.17	0.57
1:J:137:ARG:HG3	1:J:141:ASP:HA	1.88	0.56
1:A:136:LYS:NZ	5:A:501:HOH:O	2.35	0.55
1:E:196:ASN:HD21	4:E:403:GOL:H32	1.72	0.54
1:A:137:ARG:HG3	1:A:141:ASP:HA	1.93	0.51
1:C:222:GLU:CD	1:C:222:GLU:H	2.14	0.50
1:E:137:ARG:HG3	1:E:141:ASP:HA	1.93	0.49
1:E:110:LEU:HG	1:E:303:ASN:HB2	1.95	0.49
1:I:137:ARG:HG3	1:I:141:ASP:HA	1.94	0.48
1:B:95:ARG:NH1	1:B:97:ASN:OD1	2.45	0.48
1:J:253:GLU:CD	1:J:253:GLU:H	2.18	0.47
1:A:64:ASP:O	1:A:65:ASN:HB2	2.17	0.45
1:B:137:ARG:HG3	1:B:141:ASP:HA	1.99	0.44
1:B:64:ASP:O	1:B:65:ASN:HB2	2.18	0.44
1:G:137:ARG:HG3	1:G:141:ASP:HA	2.00	0.43
1:A:74:VAL:O	3:A:402:SIA:H91	2.19	0.43
1:A:110:LEU:HG	1:A:303:ASN:HB2	2.01	0.43
1:A:167:PHE:CE1	1:A:213:CYS:HB2	2.54	0.42
1:F:167:PHE:CE1	1:F:213:CYS:HB2	2.55	0.42
1:E:265:LEU:HD21	1:E:294:ILE:HD13	2.01	0.41
1:H:167:PHE:CE1	1:H:213:CYS:HB2	2.55	0.41
1:D:90:THR:OG1	1:D:207:GLY:HA2	2.21	0.41
1:D:167:PHE:CE1	1:D:213:CYS:HB2	2.56	0.41
1:H:137:ARG:HG3	1:H:141:ASP:HA	2.01	0.41
1:C:137:ARG:HG3	1:C:141:ASP:HA	2.02	0.41
1:H:279:LYS:NZ	5:H:504:HOH:O	2.52	0.41
1:I:45:ILE:CG2	1:I:295:LEU:HD11	2.51	0.40
1:B:45:ILE:CG2	1:B:295:LEU:HD11	2.52	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	272/280 (97%)	265 (97%)	7 (3%)	0	100	100
1	B	262/280 (94%)	255 (97%)	7 (3%)	0	100	100
1	C	270/280 (96%)	263 (97%)	7 (3%)	0	100	100
1	D	261/280 (93%)	252 (97%)	9 (3%)	0	100	100
1	E	272/280 (97%)	265 (97%)	7 (3%)	0	100	100
1	F	261/280 (93%)	252 (97%)	9 (3%)	0	100	100
1	G	268/280 (96%)	259 (97%)	9 (3%)	0	100	100
1	H	261/280 (93%)	253 (97%)	8 (3%)	0	100	100
1	I	270/280 (96%)	261 (97%)	9 (3%)	0	100	100
1	J	260/280 (93%)	253 (97%)	7 (3%)	0	100	100
All	All	2657/2800 (95%)	2578 (97%)	79 (3%)	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	240/246 (98%)	239 (100%)	1 (0%)	93	89
1	B	231/246 (94%)	231 (100%)	0	100	100
1	C	236/246 (96%)	235 (100%)	1 (0%)	93	89
1	D	228/246 (93%)	226 (99%)	2 (1%)	84	72
1	E	237/246 (96%)	235 (99%)	2 (1%)	86	76
1	F	225/246 (92%)	224 (100%)	1 (0%)	93	89
1	G	233/246 (95%)	231 (99%)	2 (1%)	84	72
1	H	230/246 (94%)	229 (100%)	1 (0%)	93	89
1	I	236/246 (96%)	234 (99%)	2 (1%)	86	76

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	J	225/246 (92%)	224 (100%)	1 (0%)	93	89
All	All	2321/2460 (94%)	2308 (99%)	13 (1%)	90	81

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	58	GLN
1	C	58	GLN
1	D	50	LEU
1	D	58	GLN
1	E	50	LEU
1	E	58	GLN
1	F	58	GLN
1	G	50	LEU
1	G	58	GLN
1	H	58	GLN
1	I	50	LEU
1	I	58	GLN
1	J	58	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	102	ASN
1	F	65	ASN
1	G	152	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

1 non-standard protein/DNA/RNA residue is 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	GAL	A	401	3	12,12,12	0.48	0	17,17,17	0.92	1 (5%)

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	GAL	A	401	3	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	GAL	O6-C6-C5	-2.03	104.52	111.30

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

12 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	GAL	A	401	3	12,12,12	0.48	0	17,17,17	0.92	1 (5%)
3	SIA	A	402	2	17,20,21	0.44	0	18,28,31	1.13	2 (11%)
3	SIA	B	401	-	18,21,21	1.11	1 (5%)	18,31,31	0.73	0
4	GOL	C	401	-	5,5,5	0.33	0	5,5,5	0.36	0
3	SIA	C	402	-	18,21,21	1.01	1 (5%)	18,31,31	0.71	0
3	SIA	E	401	-	18,21,21	1.21	1 (5%)	18,31,31	0.79	0
4	GOL	E	402	-	5,5,5	0.43	0	5,5,5	0.51	0
4	GOL	E	403	-	5,5,5	0.38	0	5,5,5	0.31	0
4	GOL	H	401	-	5,5,5	0.30	0	5,5,5	0.52	0
3	SIA	H	402	-	18,21,21	1.06	1 (5%)	18,31,31	0.72	0
3	SIA	I	401	-	18,21,21	1.08	1 (5%)	18,31,31	0.88	0
3	SIA	J	401	-	18,21,21	1.20	1 (5%)	18,31,31	0.70	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	GAL	A	401	3	-	0/2/22/22	0/1/1/1
3	SIA	A	402	2	-	0/14/34/38	0/1/1/1
3	SIA	B	401	-	-	0/14/38/38	0/1/1/1
4	GOL	C	401	-	-	0/4/4/4	0/0/0/0
3	SIA	C	402	-	-	0/14/38/38	0/1/1/1
3	SIA	E	401	-	-	0/14/38/38	0/1/1/1
4	GOL	E	402	-	-	0/4/4/4	0/0/0/0
4	GOL	E	403	-	-	0/4/4/4	0/0/0/0
4	GOL	H	401	-	-	0/4/4/4	0/0/0/0
3	SIA	H	402	-	-	0/14/38/38	0/1/1/1
3	SIA	I	401	-	-	0/14/38/38	0/1/1/1
3	SIA	J	401	-	-	0/14/38/38	0/1/1/1

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	402	SIA	O2-C2	3.62	1.44	1.39
3	H	402	SIA	O2-C2	3.75	1.44	1.39
3	I	401	SIA	O2-C2	3.76	1.44	1.39
3	B	401	SIA	O2-C2	3.89	1.44	1.39
3	J	401	SIA	O2-C2	3.97	1.44	1.39
3	E	401	SIA	O2-C2	4.04	1.44	1.39

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	402	SIA	O8-C8-C9	-2.30	103.75	109.23
2	A	401	GAL	O6-C6-C5	-2.03	104.52	111.30
3	A	402	SIA	C3-C4-C5	2.11	113.83	111.47

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	402	SIA	1	0
4	E	403	GOL	2	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

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	271/280 (96%)	-0.22	7 (2%) 59 60	13, 20, 38, 55	0
1	B	264/280 (94%)	-0.12	6 (2%) 64 65	13, 20, 34, 46	0
1	C	270/280 (96%)	-0.14	6 (2%) 65 67	12, 19, 44, 55	0
1	D	264/280 (94%)	-0.12	6 (2%) 64 65	13, 20, 43, 55	0
1	E	272/280 (97%)	-0.03	14 (5%) 32 29	14, 21, 47, 64	0
1	F	264/280 (94%)	-0.02	14 (5%) 30 27	15, 22, 41, 61	0
1	G	270/280 (96%)	0.03	10 (3%) 45 46	16, 24, 44, 59	0
1	H	264/280 (94%)	-0.03	10 (3%) 44 44	16, 23, 43, 67	0
1	I	271/280 (96%)	-0.09	7 (2%) 59 60	15, 21, 36, 68	0
1	J	264/280 (94%)	-0.11	9 (3%) 49 49	15, 21, 38, 58	0
All	All	2674/2800 (95%)	-0.08	89 (3%) 50 51	12, 21, 42, 68	0

All (89) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	106	THR	6.4
1	I	107	CYS	6.2
1	E	32	ASN	6.2
1	I	104	ASP	5.3
1	F	42	PRO	4.9
1	F	110	LEU	4.9
1	I	105	LEU	4.7
1	G	41	GLY	4.5
1	J	42	PRO	4.2
1	D	109	THR	4.2
1	F	33	ILE	4.2
1	J	40	THR	4.1
1	D	42	PRO	4.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	109	THR	4.0
1	A	42	PRO	3.9
1	C	105	LEU	3.9
1	G	42	PRO	3.9
1	E	100	MET	3.8
1	G	40	THR	3.8
1	G	303	ASN	3.7
1	F	41	GLY	3.6
1	F	140	ASP	3.5
1	J	109	THR	3.5
1	D	110	LEU	3.4
1	J	41	GLY	3.4
1	F	109	THR	3.4
1	E	107	CYS	3.3
1	D	33	ILE	3.3
1	E	105	LEU	3.3
1	G	108	ASN	3.3
1	J	110	LEU	3.3
1	E	42	PRO	3.2
1	F	34	GLU	3.2
1	F	35	VAL	3.2
1	A	303	ASN	3.2
1	A	41	GLY	3.1
1	C	303	ASN	3.1
1	H	42	PRO	3.1
1	J	32	ASN	3.1
1	D	41	GLY	3.0
1	E	303	ASN	3.0
1	I	103	GLU	2.9
1	D	40	THR	2.9
1	E	106	THR	2.9
1	G	107	CYS	2.8
1	A	140	ASP	2.7
1	H	110	LEU	2.7
1	G	109	THR	2.7
1	H	140	ASP	2.7
1	F	36	LEU	2.7
1	E	104	ASP	2.7
1	H	187	ILE	2.6
1	J	34	GLU	2.6
1	A	108	ASN	2.6
1	E	33	ILE	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	G	302	ARG	2.6
1	H	302	ARG	2.6
1	G	180	ASP	2.6
1	F	302	ARG	2.6
1	B	43	ASP	2.6
1	E	37	ASN	2.5
1	J	140	ASP	2.5
1	I	140	ASP	2.5
1	F	108	ASN	2.5
1	C	107	CYS	2.5
1	H	108	ASN	2.4
1	I	303	ASN	2.4
1	B	102	ASN	2.4
1	B	187	ILE	2.3
1	J	33	ILE	2.3
1	H	40	THR	2.3
1	F	43	ASP	2.3
1	H	47	THR	2.3
1	A	100	MET	2.3
1	E	43	ASP	2.3
1	H	43	ASP	2.3
1	B	180	ASP	2.2
1	B	302	ARG	2.2
1	C	42	PRO	2.2
1	G	260	CYS	2.2
1	A	40	THR	2.2
1	C	40	THR	2.2
1	E	41	GLY	2.2
1	E	140	ASP	2.2
1	F	280	ASP	2.2
1	C	140	ASP	2.1
1	F	180	ASP	2.1
1	E	101	LEU	2.1
1	H	109	THR	2.0

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

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	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
2	GAL	A	401	12/12	0.84	0.20	-	28,40,43,45	0

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

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	RSCC	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	SIA	E	401	21/21	0.73	0.30	5.48	29,35,43,46	0
4	GOL	E	403	6/6	0.64	0.25	5.11	36,41,41,47	0
3	SIA	I	401	21/21	0.73	0.24	4.88	29,38,48,53	0
4	GOL	E	402	6/6	0.80	0.17	4.57	30,34,35,35	0
3	SIA	B	401	21/21	0.77	0.24	4.47	29,37,45,48	0
3	SIA	C	402	21/21	0.82	0.19	4.38	22,30,36,41	0
3	SIA	J	401	21/21	0.81	0.23	4.37	28,36,40,45	0
3	SIA	H	402	21/21	0.78	0.24	2.10	31,39,43,46	0
4	GOL	H	401	6/6	0.87	0.14	1.37	28,32,34,38	0
3	SIA	A	402	20/21	0.90	0.13	1.24	19,25,31,33	0
4	GOL	C	401	6/6	0.92	0.09	-0.19	23,27,28,30	0
2	GAL	A	401	12/12	0.84	0.20	-	28,40,43,45	0

6.5 Other polymers [i](#)

There are no such residues in this entry.