



# Full wwPDB X-ray Structure Validation Report ⓘ

Sep 19, 2017 – 01:38 PM EDT

PDB ID : 5W0P  
Title : Crystal structure of rhodopsin bound to visual arrestin determined by X-ray free electron laser  
Authors : Zhou, X.E.; He, Y.; de Waal, P.W.; Gao, X.; Kang, Y.; Van Eps, N.; Yin, Y.; Pal, K.; Goswami, D.; White, T.A.; Barty, A.; Latorraca, N.R.; Chapman, H.N.; Hubbell, W.L.; Dror, R.O.; Stevens, R.C.; Cherezov, V.; Gurevich, V.V.; Griffin, P.R.; Ernst, O.P.; Melcher, K.; Xu, H.E.  
Deposited on : unknown  
Resolution : 3.01 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](mailto: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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20029824  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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-20029824

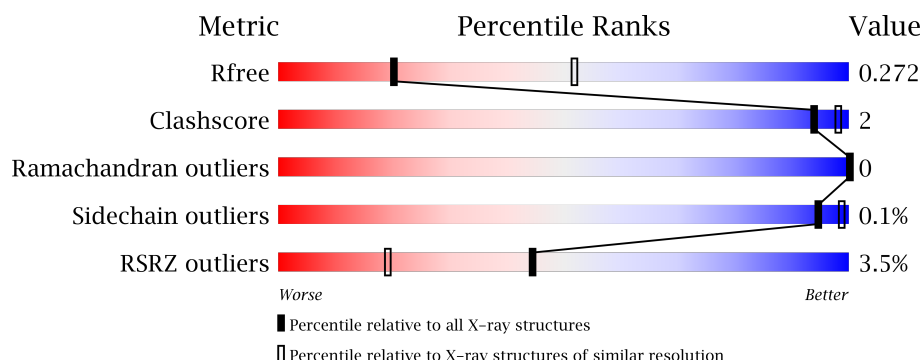
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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))
$R_{free}$	100719	1924 (3.04-3.00)
Clashscore	112137	2279 (3.04-3.00)
Ramachandran outliers	110173	2207 (3.04-3.00)
Sidechain outliers	110143	2210 (3.04-3.00)
RSRZ outliers	101464	1948 (3.04-3.00)

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  $\geq 3$ , 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	906	<div> <div>3%</div> <div>89%</div> <div>5%</div> <div>6%</div> </div>
1	B	906	<div> <div>3%</div> <div>85%</div> <div>•</div> <div>11%</div> </div>
2	C	906	<div> <div>3%</div> <div>91%</div> <div>•</div> <div>6%</div> </div>
2	D	906	<div> <div>5%</div> <div>82%</div> <div>•</div> <div>15%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 25976 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 Endolysin,Rhodopsin,S-arrestin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	849	Total	C	N	O	P	S	0	0	0
			6702	4337	1096	1227	2	40			
1	B	804	Total	C	N	O	P	S	0	0	0
			6351	4116	1035	1159	2	39			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-107	THR	CYS	engineered mutation	UNP A0A097J792
A	-64	ALA	CYS	engineered mutation	UNP A0A097J792
A	2	CYS	ASN	engineered mutation	UNP P08100
A	113	GLN	GLU	engineered mutation	UNP P08100
A	257	TYR	MET	engineered mutation	UNP P08100
A	282	CYS	ASN	engineered mutation	UNP P08100
A	1995	GLY	-	linker	UNP P08100
A	1996	SER	-	linker	UNP P08100
A	1997	ALA	-	linker	UNP P08100
A	1998	GLY	-	linker	UNP P08100
A	1999	SER	-	linker	UNP P08100
A	2000	ALA	-	linker	UNP P08100
A	2001	GLY	-	linker	UNP P08100
A	2002	SER	-	linker	UNP P08100
A	2003	ALA	-	linker	UNP P08100
A	2004	GLY	-	linker	UNP P08100
A	2005	SER	-	linker	UNP P08100
A	2006	ALA	-	linker	UNP P08100
A	2007	GLY	-	linker	UNP P08100
A	2008	SER	-	linker	UNP P08100
A	2009	ALA	-	linker	UNP P08100
A	2374	ALA	LEU	engineered mutation	UNP P20443
A	2375	ALA	VAL	engineered mutation	UNP P20443
A	2376	ALA	PHE	engineered mutation	UNP P20443
B	-107	THR	CYS	engineered mutation	UNP A0A097J792

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-64	ALA	CYS	engineered mutation	UNP A0A097J792
B	2	CYS	ASN	engineered mutation	UNP P08100
B	113	GLN	GLU	engineered mutation	UNP P08100
B	257	TYR	MET	engineered mutation	UNP P08100
B	282	CYS	ASN	engineered mutation	UNP P08100
B	1995	GLY	-	linker	UNP P08100
B	1996	SER	-	linker	UNP P08100
B	1997	ALA	-	linker	UNP P08100
B	1998	GLY	-	linker	UNP P08100
B	1999	SER	-	linker	UNP P08100
B	2000	ALA	-	linker	UNP P08100
B	2001	GLY	-	linker	UNP P08100
B	2002	SER	-	linker	UNP P08100
B	2003	ALA	-	linker	UNP P08100
B	2004	GLY	-	linker	UNP P08100
B	2005	SER	-	linker	UNP P08100
B	2006	ALA	-	linker	UNP P08100
B	2007	GLY	-	linker	UNP P08100
B	2008	SER	-	linker	UNP P08100
B	2009	ALA	-	linker	UNP P08100
B	2374	ALA	LEU	engineered mutation	UNP P20443
B	2375	ALA	VAL	engineered mutation	UNP P20443
B	2376	ALA	PHE	engineered mutation	UNP P20443

- Molecule 2 is a protein called Endolysin,Rhodopsin,S-arrestin.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	850	Total	C	N	O	P	S	0	0	0
			6703	4340	1099	1223	1	40			
2	D	774	Total	C	N	O	P	S	0	0	0
			6108	3966	997	1106	1	38			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-107	THR	CYS	engineered mutation	UNP A0A097J792
C	-64	ALA	CYS	engineered mutation	UNP A0A097J792
C	2	CYS	ASN	engineered mutation	UNP P08100
C	113	GLN	GLU	engineered mutation	UNP P08100
C	257	TYR	MET	engineered mutation	UNP P08100
C	282	CYS	ASN	engineered mutation	UNP P08100
C	1995	GLY	-	linker	UNP P08100

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	1996	SER	-	linker	UNP P08100
C	1997	ALA	-	linker	UNP P08100
C	1998	GLY	-	linker	UNP P08100
C	1999	SER	-	linker	UNP P08100
C	2000	ALA	-	linker	UNP P08100
C	2001	GLY	-	linker	UNP P08100
C	2002	SER	-	linker	UNP P08100
C	2003	ALA	-	linker	UNP P08100
C	2004	GLY	-	linker	UNP P08100
C	2005	SER	-	linker	UNP P08100
C	2006	ALA	-	linker	UNP P08100
C	2007	GLY	-	linker	UNP P08100
C	2008	SER	-	linker	UNP P08100
C	2009	ALA	-	linker	UNP P08100
C	2374	ALA	LEU	engineered mutation	UNP P20443
C	2375	ALA	VAL	engineered mutation	UNP P20443
C	2376	ALA	PHE	engineered mutation	UNP P20443
D	-107	THR	CYS	engineered mutation	UNP A0A097J792
D	-64	ALA	CYS	engineered mutation	UNP A0A097J792
D	2	CYS	ASN	engineered mutation	UNP P08100
D	113	GLN	GLU	engineered mutation	UNP P08100
D	257	TYR	MET	engineered mutation	UNP P08100
D	282	CYS	ASN	engineered mutation	UNP P08100
D	1995	GLY	-	linker	UNP P08100
D	1996	SER	-	linker	UNP P08100
D	1997	ALA	-	linker	UNP P08100
D	1998	GLY	-	linker	UNP P08100
D	1999	SER	-	linker	UNP P08100
D	2000	ALA	-	linker	UNP P08100
D	2001	GLY	-	linker	UNP P08100
D	2002	SER	-	linker	UNP P08100
D	2003	ALA	-	linker	UNP P08100
D	2004	GLY	-	linker	UNP P08100
D	2005	SER	-	linker	UNP P08100
D	2006	ALA	-	linker	UNP P08100
D	2007	GLY	-	linker	UNP P08100
D	2008	SER	-	linker	UNP P08100
D	2009	ALA	-	linker	UNP P08100
D	2374	ALA	LEU	engineered mutation	UNP P20443
D	2375	ALA	VAL	engineered mutation	UNP P20443
D	2376	ALA	PHE	engineered mutation	UNP P20443

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



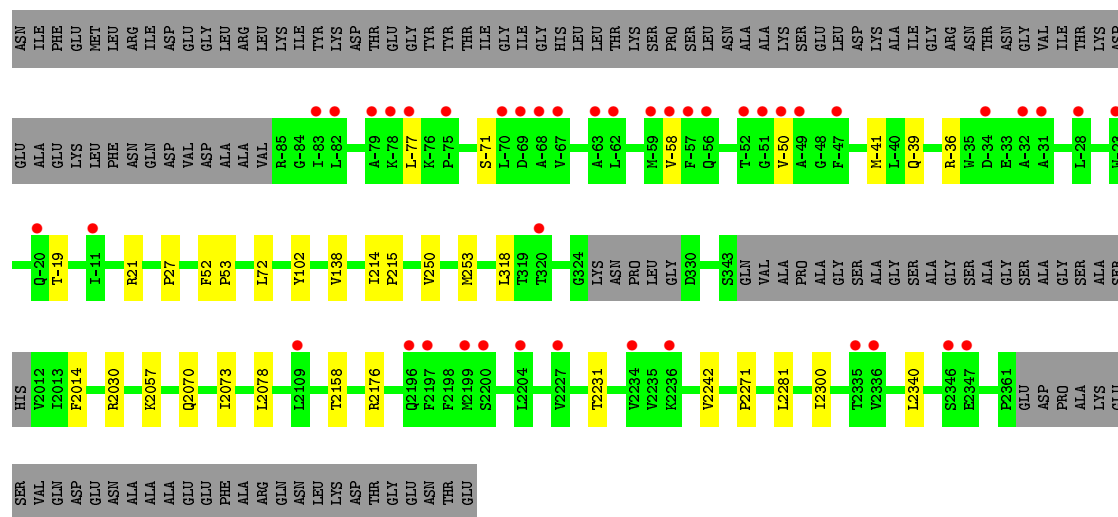
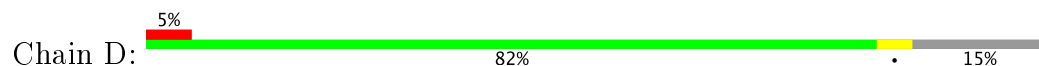
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	D	1	Total	C	N	O	0	0
			14	8	1	5		
3	D	1	Total	C	N	O	0	0
			14	8	1	5		



- Molecule 1: Endolysin,Rhodopsin,S-arrestin



- Molecule 2: Endolysin,Rhodopsin,S-arrestin





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	109.24Å 109.24Å 452.64Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.03 – 3.01 30.03 – 3.01	Depositor EDS
% Data completeness (in resolution range)	70.7 (30.03-3.01) 70.7 (30.03-3.01)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.91 (at 3.00Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.234 , 0.272 0.234 , 0.272	Depositor DCC
$R_{free}$ test set	3764 reflections (4.93%)	DCC
Wilson B-factor (Å <sup>2</sup> )	113.4	Xtriage
Anisotropy	0.566	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 138.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.35$ , $\langle L^2 \rangle = 0.18$	Xtriage
Estimated twinning fraction	0.499 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	25976	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	201.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.97% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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: TPO, NAG, SEP

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	0.26	0/6837	0.43	0/9285
1	B	0.29	0/6480	0.47	0/8804
2	C	0.27	0/6851	0.45	0/9306
2	D	0.28	0/6248	0.46	0/8495
All	All	0.28	0/26416	0.45	0/35890

There are no bond length outliers.

There are no bond angle outliers.

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	6702	0	6749	27	0
1	B	6351	0	6385	19	0
2	C	6703	0	6755	15	0
2	D	6108	0	6152	21	0
3	A	28	0	25	0	0
3	B	28	0	25	0	0
3	C	28	0	25	1	0
3	D	28	0	25	2	0
All	All	25976	0	26141	83	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 2.

All (83) 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:B:2301:LYS:HG3	1:B:2302:HIS:H	1.51	0.74
2:D:2176:ARG:HG3	2:D:2300:ILE:HD11	1.70	0.72
1:B:72:LEU:HD12	1:B:2073:ILE:HD12	1.72	0.72
1:A:72:LEU:HB2	1:A:2073:ILE:HD11	1.74	0.68
2:C:-74:VAL:HA	2:C:-39:GLN:OE1	1.93	0.68
1:B:2027:LEU:HD12	1:B:2173:LEU:HD23	1.77	0.65
1:A:2306:ASN:HD21	1:A:2358:HIS:CE1	2.15	0.64
1:B:-150:GLU:HA	1:B:-16:ARG:NH2	2.14	0.61
2:D:2070:GLN:HB3	2:D:2073:ILE:HG22	1.83	0.61
1:A:312:GLN:HB3	1:A:2076:MET:SD	2.41	0.60
2:D:72:LEU:HB2	2:D:2073:ILE:HD11	1.82	0.60
1:B:-74:VAL:HA	1:B:-39:GLN:OE1	2.05	0.56
2:C:146:PHE:HZ	2:C:2135:PRO:HA	1.70	0.55
1:A:2176:ARG:HG3	1:A:2300:ILE:HD11	1.89	0.55
2:C:34:PRO:HB2	2:D:318:LEU:HD21	1.90	0.54
2:D:21:ARG:HA	3:D:2401:NAG:H83	1.88	0.53
1:B:70:THR:HG21	1:B:2073:ILE:HD13	1.89	0.53
2:C:2282:VAL:O	2:C:2282:VAL:HG13	2.09	0.52
1:A:2070:GLN:HB2	1:A:2073:ILE:HG22	1.92	0.52
2:D:2030:ARG:HD3	2:D:2300:ILE:HG21	1.92	0.51
2:D:253:MET:HB2	2:D:2078:LEU:HD21	1.92	0.51
1:B:72:LEU:HD21	1:B:138:VAL:HG21	1.93	0.51
1:A:27:PRO:HA	1:A:102:TYR:HB3	1.93	0.50
2:D:-58:VAL:HG22	2:D:-50:VAL:HG11	1.93	0.50
1:A:-71:SER:HB2	1:A:-39:GLN:HE21	1.76	0.50
1:B:2302:HIS:O	1:B:2302:HIS:CG	2.64	0.49
2:C:146:PHE:CZ	2:C:2135:PRO:HA	2.47	0.49
1:A:332:GLU:HG3	1:A:333:ALA:N	2.27	0.49
2:C:257:TYR:HE2	2:C:2078:LEU:HD21	1.78	0.48
1:B:145:ASN:ND2	1:B:2069:GLY:HA3	2.29	0.48
2:D:27:PRO:HA	2:D:102:TYR:HB3	1.95	0.47
1:A:-19:THR:HG23	1:A:-19:THR:O	2.14	0.47
1:A:-150:GLU:HA	1:A:-16:ARG:NH2	2.30	0.47
2:D:-19:THR:O	2:D:-19:THR:HG23	2.15	0.47
2:D:-71:SER:HB2	2:D:-39:GLN:OE1	2.15	0.47
2:D:-77:LEU:HD21	2:D:-50:VAL:HG13	1.97	0.46
1:A:-158:ILE:HG12	1:A:-64:ALA:HB1	1.98	0.46

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:52:PHE:HB3	1:A:53:PRO:HD3	1.98	0.46
2:C:2053:VAL:HG12	2:C:2053:VAL:O	2.15	0.46
1:A:-45:ASN:HB3	1:A:-29:ASN:HD21	1.81	0.46
1:B:70:THR:HG21	1:B:2073:ILE:HG21	1.98	0.46
1:A:253:MET:HB2	1:A:2078:LEU:HD11	1.99	0.45
2:C:2301:LYS:HG3	2:C:2302:HIS:H	1.82	0.45
1:A:2014:PHE:CZ	1:A:2030:ARG:HG3	2.52	0.45
2:D:52:PHE:HB3	2:D:53:PRO:HD3	1.98	0.45
1:B:10:TYR:CE1	1:B:192:TYR:HE1	2.34	0.45
1:B:2053:VAL:O	1:B:2053:VAL:HG12	2.16	0.45
1:B:52:PHE:HB3	1:B:53:PRO:HD3	1.99	0.44
1:A:70:THR:HG21	1:A:2073:ILE:HD13	1.98	0.44
1:A:271:VAL:HG21	1:A:291:PRO:HG3	1.99	0.44
2:D:2014:PHE:CZ	2:D:2030:ARG:HG3	2.52	0.44
1:A:-158:ILE:HG12	1:A:-64:ALA:CB	2.48	0.44
2:D:250:VAL:HG22	2:D:2078:LEU:HD13	1.99	0.44
1:B:2014:PHE:HE1	1:B:2029:LYS:HA	1.83	0.43
1:B:2301:LYS:HG3	1:B:2302:HIS:N	2.26	0.43
2:D:2242:VAL:HG23	2:D:2281:LEU:HD22	2.01	0.43
1:A:209:VAL:O	1:A:214:ILE:HG12	2.17	0.43
1:A:72:LEU:HD21	1:A:138:VAL:HG21	2.00	0.43
1:B:2082:ARG:HB3	1:B:2250:LEU:HD11	2.00	0.43
2:C:2027:LEU:HD12	2:C:2173:LEU:HD23	2.01	0.43
1:A:33:GLU:HA	1:A:34:PRO:HD3	1.92	0.43
2:C:52:PHE:HB3	2:C:53:PRO:HD3	2.00	0.42
1:A:2285:LEU:HD13	1:A:2296:LEU:HD21	2.00	0.42
1:A:2206:LEU:HD13	1:A:2334:LEU:HD11	2.02	0.42
2:C:2077:GLY:O	2:C:2078:LEU:HB2	2.19	0.42
2:D:72:LEU:HD21	2:D:138:VAL:HG21	2.01	0.42
2:C:2014:PHE:HE1	2:C:2029:LYS:HA	1.84	0.42
2:D:-41:MET:HB3	2:D:-36:ARG:HB2	2.01	0.42
1:A:2206:LEU:CD1	1:A:2334:LEU:HD11	2.50	0.42
2:C:209:VAL:O	2:C:214:ILE:HG12	2.20	0.42
2:D:214:ILE:HB	2:D:215:PRO:HD3	2.02	0.41
3:C:2402:NAG:HO3	3:C:2402:NAG:C7	2.32	0.41
2:C:72:LEU:HD21	2:C:138:VAL:HG21	2.03	0.41
1:A:2077:GLY:O	1:A:2078:LEU:HB3	2.20	0.41
3:D:2401:NAG:O3	3:D:2401:NAG:C7	2.69	0.41
1:A:2242:VAL:HG23	1:A:2281:LEU:HD22	2.03	0.41
1:B:126:TRP:HB2	1:B:160:THR:HG23	2.03	0.41
1:B:2188:GLN:HA	1:B:2189:PRO:HD3	1.87	0.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:2057:LYS:HE3	2:D:2158:THR:HA	2.03	0.40
1:B:33:GLU:HA	1:B:34:PRO:HD3	1.92	0.40
2:D:2231:THR:O	2:D:2271:PRO:HB3	2.20	0.40
1:A:-155:MET:HG2	1:A:0:TYR:CD1	2.56	0.40
2:C:310:ASN:HB3	2:C:2076:MET:HG3	2.03	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

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	841/906 (93%)	827 (98%)	14 (2%)	0	100	100
1	B	794/906 (88%)	784 (99%)	10 (1%)	0	100	100
2	C	843/906 (93%)	831 (99%)	12 (1%)	0	100	100
2	D	767/906 (85%)	759 (99%)	8 (1%)	0	100	100
All	All	3245/3624 (90%)	3201 (99%)	44 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

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	736/776 (95%)	736 (100%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	698/776 (90%)	697 (100%)	1 (0%)	94	99
2	C	737/777 (95%)	737 (100%)	0	100	100
2	D	673/777 (87%)	672 (100%)	1 (0%)	94	99
All	All	2844/3106 (92%)	2842 (100%)	2 (0%)	94	99

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	2343	LEU
2	D	2340	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	-60	ASN
1	A	-39	GLN
1	A	-29	ASN
1	A	199	ASN
1	A	200	ASN
1	A	2244	GLN
1	A	2358	HIS
1	B	278	HIS
1	B	2244	GLN
2	C	145	ASN
2	C	2244	GLN
2	D	2244	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

6 non-standard protein/DNA/RNA residues 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	$\# Z  > 2$
1	TPO	A	336	1	9,10,11	1.35	1 (11%)	10,14,16	0.83	0
1	SEP	A	338	1	9,9,10	1.42	1 (11%)	9,12,14	0.92	0
1	TPO	B	336	1	9,10,11	1.84	2 (22%)	10,14,16	0.77	0
1	SEP	B	338	1	9,9,10	1.40	1 (11%)	9,12,14	0.87	0
2	SEP	C	338	2	9,9,10	1.34	1 (11%)	9,12,14	1.06	0
2	SEP	D	338	2	9,9,10	1.37	1 (11%)	9,12,14	0.85	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
1	TPO	A	336	1	-	1/8/11/13	0/0/0/0
1	SEP	A	338	1	-	0/5/8/10	0/0/0/0
1	TPO	B	336	1	-	1/8/11/13	0/0/0/0
1	SEP	B	338	1	-	0/5/8/10	0/0/0/0
2	SEP	C	338	2	-	0/5/8/10	0/0/0/0
2	SEP	D	338	2	-	0/5/8/10	0/0/0/0

All (7) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	336	TPO	CA-C	2.62	1.53	1.50
2	C	338	SEP	P-O1P	2.92	1.60	1.50
1	B	338	SEP	P-O1P	2.94	1.60	1.50
2	D	338	SEP	P-O1P	2.95	1.60	1.50
1	A	338	SEP	P-O1P	2.98	1.60	1.50
1	B	336	TPO	P-O1P	3.03	1.61	1.50
1	B	336	TPO	CA-C	3.21	1.54	1.50

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	336	TPO	OG1-CB-CA-N
1	B	336	TPO	OG1-CB-CA-N

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](#)

8 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$
3	NAG	A	2401	1,3	14,14,15	0.87	1 (7%)	15,19,21	0.99	1 (6%)
3	NAG	A	2402	3	14,14,15	0.55	0	15,19,21	0.55	0
3	NAG	B	2401	1,3	14,14,15	0.72	0	15,19,21	0.89	0
3	NAG	B	2402	3	14,14,15	0.98	1 (7%)	15,19,21	1.64	2 (13%)
3	NAG	C	2401	3,2	14,14,15	0.90	0	15,19,21	0.97	1 (6%)
3	NAG	C	2402	3	14,14,15	0.91	1 (7%)	15,19,21	0.87	1 (6%)
3	NAG	D	2401	3,2	14,14,15	0.96	1 (7%)	15,19,21	1.28	1 (6%)
3	NAG	D	2402	3	14,14,15	0.80	1 (7%)	15,19,21	0.73	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
3	NAG	A	2401	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	2402	3	-	0/6/23/26	0/1/1/1

*Continued on next page...*



*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	B	2401	1,3	-	0/6/23/26	0/1/1/1
3	NAG	B	2402	3	-	0/6/23/26	0/1/1/1
3	NAG	C	2401	3,2	-	0/6/23/26	0/1/1/1
3	NAG	C	2402	3	-	0/6/23/26	0/1/1/1
3	NAG	D	2401	3,2	-	0/6/23/26	0/1/1/1
3	NAG	D	2402	3	-	0/6/23/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	2402	NAG	O5-C1	-2.98	1.38	1.43
3	D	2401	NAG	C2-N2	-2.36	1.42	1.46
3	A	2401	NAG	C1-C2	2.03	1.55	1.52
3	D	2402	NAG	C1-C2	2.18	1.55	1.52
3	C	2402	NAG	C1-C2	2.19	1.55	1.52

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	2402	NAG	C1-C2-N2	-5.07	101.83	110.49
3	A	2401	NAG	C4-C3-C2	2.06	114.04	111.02
3	C	2402	NAG	C4-C3-C2	2.15	114.17	111.02
3	C	2401	NAG	C4-C3-C2	2.16	114.18	111.02
3	B	2402	NAG	C2-N2-C7	2.60	126.74	122.94
3	D	2401	NAG	C2-N2-C7	3.27	127.72	122.94

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	C	2402	NAG	1	0
3	D	2401	NAG	2	0

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	847/906 (93%)	-0.02	26 (3%)	49	22	115, 194, 307, 335	0
1	B	802/906 (88%)	-0.08	25 (3%)	49	22	116, 196, 297, 321	0
2	C	849/906 (93%)	-0.11	23 (2%)	55	26	119, 190, 323, 347	0
2	D	773/906 (85%)	0.07	42 (5%)	26	10	83, 171, 341, 370	0
All	All	3271/3624 (90%)	-0.04	116 (3%)	44	19	83, 188, 311, 370	0

All (116) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	-124	PRO	8.9
2	C	-148	LEU	8.2
2	D	2336	VAL	7.8
2	D	-34	ASP	7.6
2	D	-82	LEU	6.6
2	D	-69	ASP	6.5
2	D	2197	PHE	6.3
1	A	-34	ASP	6.2
1	A	-109	ARG	5.6
1	A	-144	ILE	5.5
1	A	-126	LYS	5.5
2	D	-11	ILE	5.1
2	D	-70	LEU	4.9
2	C	-109	ARG	4.8
2	D	2346	SER	4.8
2	D	-62	LEU	4.7
1	A	-67	VAL	4.7
2	D	-78	LYS	4.6
1	A	-148	LEU	4.6
1	A	-125	SER	4.4
1	A	-68	ALA	4.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	-63	ALA	4.3
2	D	-68	ALA	4.3
2	D	2234	VAL	4.2
2	D	-79	ALA	4.1
2	C	-47	PHE	4.1
2	C	2162	GLU	4.1
2	D	-50	VAL	3.9
2	C	-32	ALA	3.8
2	D	2200	SER	3.8
2	D	-49	ALA	3.8
1	B	-28	LEU	3.8
2	C	-126	LYS	3.7
2	D	2204	LEU	3.6
2	D	2199	MET	3.5
1	A	-136	TYR	3.5
1	A	-131	GLY	3.4
1	B	-62	LEU	3.4
1	B	-11	ILE	3.4
2	D	-57	PHE	3.3
1	A	-139	GLU	3.3
2	C	-111	ILE	3.2
1	B	183	LEU	3.2
2	C	-73	TYR	3.1
1	A	-157	PHE	3.0
2	C	-149	GLY	3.0
2	D	-56	GLN	3.0
1	A	2198	PHE	3.0
2	C	-146	LEU	3.0
2	C	-153	ARG	3.0
2	C	-157	PHE	3.0
2	D	-67	VAL	2.9
2	D	2227	VAL	2.9
1	B	322	CYS	2.9
2	C	-147	ARG	2.9
1	B	-70	LEU	2.9
2	D	-52	THR	2.9
1	B	-23	TRP	2.9
2	D	2347	GLU	2.9
1	A	-110	GLY	2.8
1	A	2197	PHE	2.8
1	B	192	TYR	2.8
1	A	-146	LEU	2.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	-20	GLN	2.7
1	B	31	LEU	2.7
2	C	-108	ASN	2.7
2	D	320	THR	2.7
2	C	-124	PRO	2.7
1	B	2100	VAL	2.7
2	C	-125	SER	2.6
2	D	-77	LEU	2.6
1	A	-55	MET	2.6
2	D	2196	GLN	2.6
2	D	-75	PRO	2.6
2	D	2335	THR	2.5
2	C	-133	GLY	2.5
1	B	-61	ILE	2.5
1	B	-63	ALA	2.5
2	D	-32	ALA	2.5
1	B	2200	SER	2.4
1	B	2336	VAL	2.4
1	B	-40	LEU	2.4
1	B	-55	MET	2.4
1	B	2274	THR	2.4
2	D	-58	VAL	2.4
2	D	-59	MET	2.4
2	D	-47	PHE	2.3
2	D	-83	ILE	2.3
1	A	-90	VAL	2.3
2	C	-40	LEU	2.3
1	B	172	LEU	2.3
2	C	-55	MET	2.3
2	D	-31	ALA	2.3
2	C	2198	PHE	2.3
2	D	-51	GLY	2.2
2	D	-23	TRP	2.2
1	B	323	CYS	2.2
1	B	-158	ILE	2.2
2	D	-28	LEU	2.2
1	A	-23	TRP	2.2
1	A	-115	LEU	2.2
2	C	-95	LEU	2.2
2	D	2109	LEU	2.2
1	A	-94	PHE	2.2
2	C	-62	LEU	2.1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	D	2236	LYS	2.1
1	B	-51	GLY	2.1
2	C	68	LEU	2.1
1	A	-21	ASN	2.1
1	B	-50	VAL	2.1
1	A	189	ILE	2.0
1	A	37	PHE	2.0
1	B	-96	LYS	2.0
1	A	24	PHE	2.0
1	B	343	SER	2.0
1	B	229	THR	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
1	SEP	B	338	10/11	0.66	0.15	-	244,270,302,307	0
2	SEP	C	338	10/11	0.74	0.14	-	209,224,261,294	0
1	SEP	A	338	10/11	0.90	0.12	-	161,177,223,236	0
1	TPO	B	336	11/12	0.70	0.19	-	290,319,323,323	0
1	TPO	A	336	11/12	0.73	0.12	-	213,238,258,261	0
2	SEP	D	338	10/11	0.93	0.13	-	171,190,205,216	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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
3	NAG	C	2401	14/15	0.92	0.16	0.41	261,289,303,307	0
3	NAG	A	2401	14/15	0.93	0.20	0.23	217,244,266,290	0
3	NAG	B	2401	14/15	0.95	0.18	0.12	264,287,315,325	0
3	NAG	D	2401	14/15	0.95	0.17	0.04	123,171,181,188	0
3	NAG	C	2402	14/15	0.93	0.14	-	221,271,296,319	0
3	NAG	D	2402	14/15	0.95	0.17	-	153,186,215,232	0
3	NAG	B	2402	14/15	0.86	0.31	-	320,341,347,347	0
3	NAG	A	2402	14/15	0.92	0.16	-	237,270,293,298	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.