W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Jan 25, 2023 – 06:45 AM EST

PDB ID	:	3QNK
Title	:	Crystal structure of a SusD-like protein (BF3747) from Bacteroides fragilis NCTC 9343 at 2.70 A resolution
Authors	:	Joint Center for Structural Genomics (JCSG)
Deposited on	:	2011-02-08
Resolution	:	2.70 Å(reported)

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

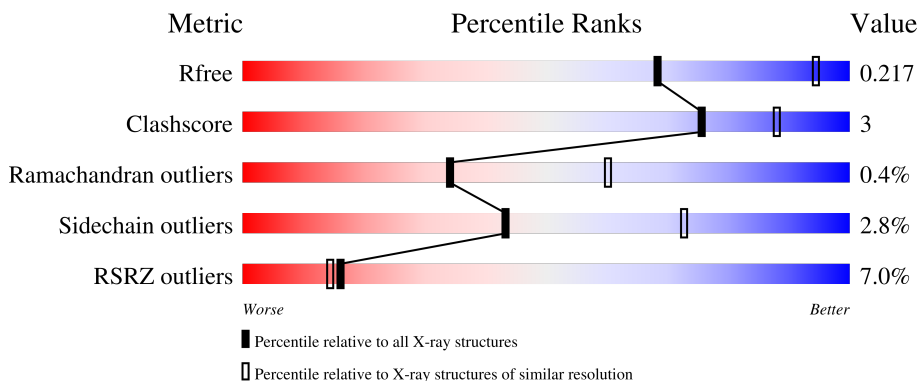|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.70 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	Percentile Ranks	Value
Rfree		0.217
Clashscore		3
Ramachandran outliers		0.4%
Sidechain outliers		2.8%
RSRZ outliers		7.0%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. 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 <=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	517	7% / 89% / 9% •
1	B	517	4% / 90% / 9% •
1	C	517	2% / 88% / 10% ••
1	D	517	14% / 89% / 10% •

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	PO4	B	601	-	X	-	-
3	PO4	B	602	-	X	-	-
3	PO4	C	603	-	X	-	-
3	PO4	C	605	-	X	-	-
4	ACT	B	608	-	X	-	-
4	ACT	C	609	-	X	X	-
4	ACT	C	610	-	X	-	-
4	ACT	C	611	-	X	-	-
4	ACT	D	613	-	X	-	-

2 Entity composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 16607 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 Putative lipoprotein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	512	Total	C	N	O	S	Se	0	0	0
			4090	2600	680	794	5	11			
1	B	513	Total	C	N	O	S	Se	0	1	0
			4157	2642	697	802	5	11			
1	C	511	Total	C	N	O	S	Se	0	2	0
			4156	2642	696	802	5	11			
1	D	514	Total	C	N	O	S	Se	0	0	0
			4073	2585	676	796	5	11			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q5L904
B	0	GLY	-	expression tag	UNP Q5L904
C	0	GLY	-	expression tag	UNP Q5L904
D	0	GLY	-	expression tag	UNP Q5L904

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Cl	0	0
			1	1		

- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).

PO4

O3

O⁻

O2 O⁻ — P = O O1

P

O⁻

O4

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total 5	O 4	P 1	0	0
3	B	1	Total 5	O 4	P 1	0	0
3	B	1	Total 5	O 4	P 1	0	0
3	C	1	Total 5	O 4	P 1	0	0
3	C	1	Total 5	O 4	P 1	0	0
3	C	1	Total 5	O 4	P 1	0	0

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).

ACT

OXT

O⁻

C

CH3

O

O

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total 4	C 2	O 2	0	0
4	B	1	Total 4	C 2	O 2	0	0
4	C	1	Total 4	C 2	O 2	0	0
4	C	1	Total 4	C 2	O 2	0	0
4	C	1	Total 4	C 2	O 2	0	0
4	C	1	Total 4	C 2	O 2	0	0
4	D	1	Total 4	C 2	O 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	11	Total 11	O 11	0	0
5	B	25	Total 25	O 25	0	0
5	C	28	Total 28	O 28	0	0
5	D	8	Total 8	O 8	0	0

WORLD WIDE
PDB
PROTEIN DATA BANK
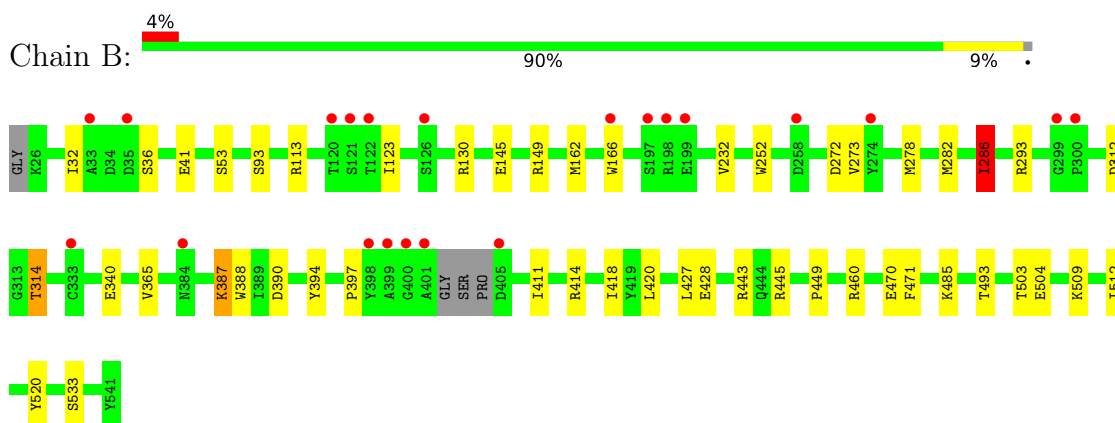
3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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.
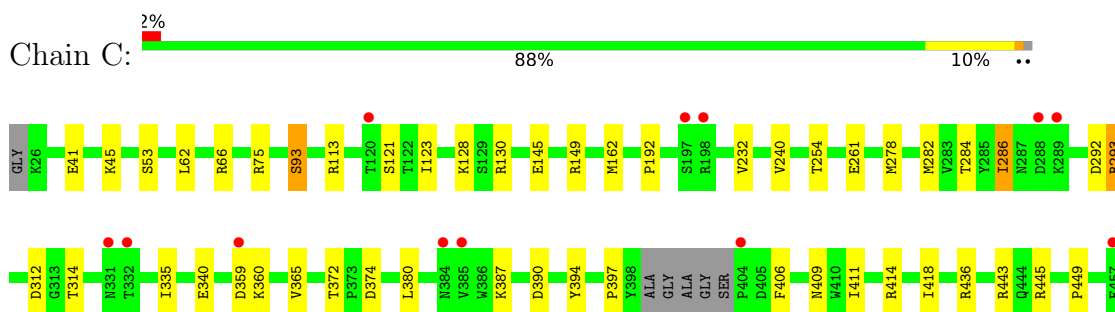
- Molecule 1: Putative lipoprotein

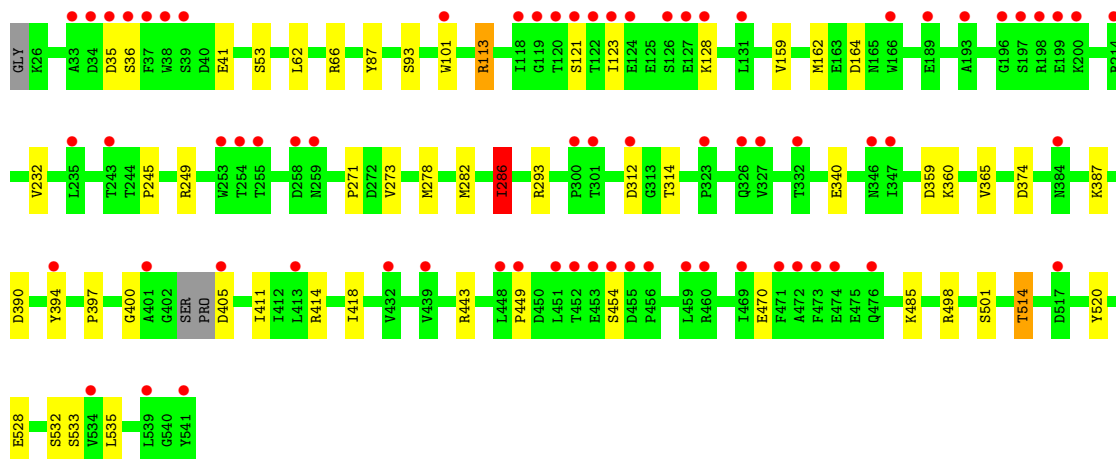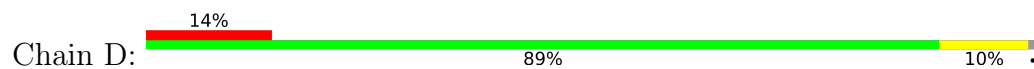Chain A:

- Molecule 1: Putative lipoprotein

Chain B:

- Molecule 1: Putative lipoprotein

Chain C:

E470 K485 T514 Y520 S533 L539 G540 Y541

• Molecule 1: Putative lipoprotein

Chain D:

14%

89%   10%

GLY K26 A33 D34 D35 S36 F37 W38 S39 D40 E41 S53 L62 R66 Y87 S93 W101 R113 I118 G119 T120 S121 T122 I123 E124 E125 S126 E127 K128 L131 V159 M162 E163 D164 N165 W166 E189 A193 G196 S197 R198 E199 K200 R214

V232 L235 T243 T244 P245 R249 W253 T254 T255 D258 N259 P271 D272 V273 M278 M282 T286 R293 P300 T301 D312 G313 T314 P323 Q326 V327 T332 E340 N346 I347 D359 K360 V365 D374 N384 K387

D390 Y394 P397 G400 A401 G402 SER PRO D405 I411 I412 L413 R414 I418 Y432 V439 R443 L448 P449 D450 L451 T452 E453 S454 D455 P456 L459 R460 I469 E470 F471 A472 F473 E474 E475 Q476 K485 R498 S501 T514 D517 Y520

E528 S532 S533 V534 L535 L539 G540 Y541

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	146.65Å 146.65Å 226.18Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	29.90 – 2.70 29.90 – 2.70	Depositor EDS
% Data completeness (in resolution range)	(Not available) (29.90-2.70) 99.8 (29.90-2.70)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 2.68Å)	Xtriage
Refinement program	BUSTER	Depositor
R, R_{free}	0.183 , 0.211 0.190 , 0.217	Depositor DCC
R_{free} test set	3910 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	61.5	Xtriage
Anisotropy	0.402	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 58.5	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.018 for -h,-k,l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	16607	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.25% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5   Model quality ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: PO4, CL, ACT

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	0.50	0/4180	0.67	1/5669 (0.0%)
1	B	0.53	0/4250	0.69	1/5749 (0.0%)
1	C	0.54	0/4252	0.70	0/5749
1	D	0.48	0/4163	0.65	1/5650 (0.0%)
All	All	0.51	0/16845	0.68	3/22817 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	286	ILE	CA-CB-CG2	5.28	121.46	110.90
1	D	286	ILE	CA-CB-CG2	5.25	121.39	110.90
1	A	286	ILE	CA-CB-CG2	5.18	121.26	110.90

There are no chirality outliers.

There are no planarity outliers.

5.2   Too-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 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	4090	0	3779	27	0
1	B	4157	0	3906	27	0
1	C	4156	0	3917	33	0
1	D	4073	0	3725	26	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1	0	0	0	0
3	B	15	0	0	0	0
3	C	15	0	0	1	0
4	B	8	0	6	0	0
4	C	16	0	12	2	0
4	D	4	0	3	0	0
5	A	11	0	0	0	0
5	B	25	0	0	0	0
5	C	28	0	0	0	0
5	D	8	0	0	0	0
All	All	16607	0	15348	100	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 3.

The worst 5 of 100 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:145:GLU:HG2	4:C:609:ACT:H3	1.58	0.84
1:C:162:MSE:HE1	1:D:273:VAL:HA	1.59	0.84
1:C:162:MSE:HE2	1:D:271:PRO:O	1.91	0.71
1:C:145:GLU:CG	4:C:609:ACT:H3	2.23	0.68
1:C:278:MSE:HE2	1:C:282:MSE:HE2	1.76	0.68

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	508/517 (98%)	490 (96%)	16 (3%)	2 (0%)	34	60
1	B	510/517 (99%)	495 (97%)	13 (2%)	2 (0%)	34	60

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	509/517 (98%)	497 (98%)	10 (2%)	2 (0%)	34	60
1	D	510/517 (99%)	492 (96%)	15 (3%)	3 (1%)	25	50
All	All	2037/2068 (98%)	1974 (97%)	54 (3%)	9 (0%)	34	60

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	36	SER
1	D	387	LYS
1	A	36	SER
1	A	387	LYS
1	B	387	LYS

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	420/445 (94%)	407 (97%)	13 (3%)	40	69
1	B	434/445 (98%)	423 (98%)	11 (2%)	47	76
1	C	436/445 (98%)	424 (97%)	12 (3%)	43	73
1	D	415/445 (93%)	403 (97%)	12 (3%)	42	71
All	All	1705/1780 (96%)	1657 (97%)	48 (3%)	43	73

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	286	ILE
1	C	539	LEU
1	C	293	ARG
1	C	374	ASP
1	D	53	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

There are no RNA molecules 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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 14 ligands modelled in this entry, 1 is monoatomic - leaving 13 for Mogul analysis.

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	#$|Z| > 2$
4	ACT	D	613	-	3,3,3	2.01	1 (33%)	3,3,3	3.21	2 (66%)
4	ACT	C	610	-	3,3,3	1.47	1 (33%)	3,3,3	2.69	2 (66%)
4	ACT	B	608	-	3,3,3	2.31	2 (66%)	3,3,3	1.77	1 (33%)
3	PO4	C	605	-	4,4,4	2.56	4 (100%)	6,6,6	2.04	3 (50%)
4	ACT	C	611	-	3,3,3	1.84	1 (33%)	3,3,3	2.35	2 (66%)
4	ACT	C	612	-	3,3,3	1.20	0	3,3,3	0.79	0
3	PO4	B	604	-	4,4,4	2.01	2 (50%)	6,6,6	0.81	0
4	ACT	C	609	-	3,3,3	2.00	1 (33%)	3,3,3	1.84	2 (66%)
3	PO4	B	601	-	4,4,4	3.22	3 (75%)	6,6,6	1.60	1 (16%)
3	PO4	C	606	-	4,4,4	2.33	2 (50%)	6,6,6	1.47	1 (16%)
3	PO4	C	603	-	4,4,4	4.81	4 (100%)	6,6,6	2.63	3 (50%)
4	ACT	B	607	-	3,3,3	1.03	0	3,3,3	2.58	2 (66%)
3	PO4	B	602	-	4,4,4	5.70	3 (75%)	6,6,6	2.28	4 (66%)

The worst 5 of 24 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	602	PO4	P-O1	8.28	1.70	1.50
3	B	602	PO4	P-O2	6.94	1.75	1.54
3	C	603	PO4	P-O4	6.29	1.73	1.54
3	C	603	PO4	P-O3	4.76	1.68	1.54
3	B	601	PO4	P-O1	4.67	1.61	1.50

The worst 5 of 23 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	613	ACT	O-C-CH3	-4.47	104.94	122.33
3	C	603	PO4	O3-P-O1	-4.34	95.03	110.89
4	C	610	ACT	OXT-C-CH3	3.68	130.37	115.18
3	C	605	PO4	O3-P-O1	-3.66	97.48	110.89
4	B	607	ACT	O-C-CH3	-3.30	109.48	122.33

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
4	C	609	ACT	2	0
3	C	606	PO4	1	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, 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	501/517 (96%)	0.20	37 (7%)	14	12	45, 75, 109, 145	0
1	B	502/517 (97%)	0.06	21 (4%)	36	35	39, 58, 94, 121	0
1	C	500/517 (96%)	-0.09	12 (2%)	59	60	39, 57, 86, 116	0
1	D	503/517 (97%)	0.64	71 (14%)	2	1	57, 89, 120, 147	0
All	All	2006/2068 (97%)	0.20	141 (7%)	16	14	39, 69, 109, 147	0

The worst 5 of 141 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	197	SER	7.6
1	B	122	THR	6.3
1	B	399	ALA	5.5
1	D	471	PHE	5.4
1	D	120	THR	5.1

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 monosaccharides 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. 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	B-factors(Å^2)	Q<0.9
4	ACT	C	612	4/4	0.69	0.33	86,87,88,88	0
2	CL	B	600	1/1	0.74	0.23	88,88,88,88	0
3	PO4	C	605	5/5	0.76	0.32	128,131,133,135	0
4	ACT	D	613	4/4	0.78	0.20	67,70,70,70	0
3	PO4	B	602	5/5	0.84	0.29	81,83,89,92	0
4	ACT	C	611	4/4	0.84	0.24	57,61,62,63	0
3	PO4	B	601	5/5	0.88	0.23	107,111,113,115	0
3	PO4	C	603	5/5	0.89	0.24	81,81,85,88	0
4	ACT	C	610	4/4	0.90	0.35	54,59,61,64	0
4	ACT	B	607	4/4	0.91	0.21	62,65,66,66	0
4	ACT	C	609	4/4	0.91	0.26	55,56,58,60	0
4	ACT	B	608	4/4	0.94	0.42	58,59,62,63	0
3	PO4	C	606	5/5	0.96	0.13	78,81,86,86	0
3	PO4	B	604	5/5	0.98	0.13	70,75,78,79	0

6.5 Other polymers ⓘ

There are no such residues in this entry.